

Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation

Cigdem Altintas,[‡] Omer Faruk Altundal,[‡] Seda Keskin,^{*} and Ramazan Yildirim^{*}

 Cite This: *J. Chem. Inf. Model.* 2021, 61, 2131–2146

 Read Online

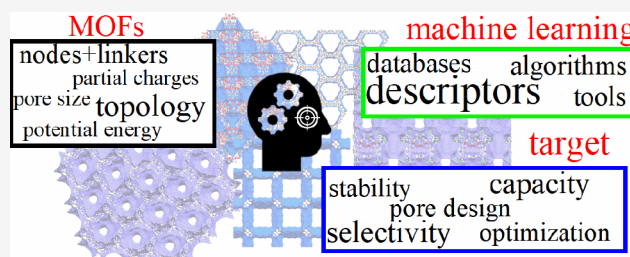
ACCESS |

 Metrics & More

 Article Recommendations

ABSTRACT: The acceleration in design of new metal organic frameworks (MOFs) has led scientists to focus on high-throughput computational screening (HTCS) methods to quickly assess the promises of these fascinating materials in various applications. HTCS studies provide a massive amount of structural property and performance data for MOFs, which need to be further analyzed. Recent implementation of machine learning (ML), which is another growing field in research, to HTCS of MOFs has been very fruitful not only for revealing the hidden structure–performance relationships of materials but also for understanding their performance trends in different applications, specifically for gas storage and separation. In this review, we highlight the current state of the art in ML-assisted computational screening of MOFs for gas storage and separation and address both the opportunities and challenges that are emerging in this new field by emphasizing how merging of ML and MOF simulations can be useful.

KEYWORDS: Metal–organic frameworks, Machine learning, High-throughput computational screening, Gas storage, Gas separation, Structure–performance relationships, Modeling, Material design



INTRODUCTION

Metal organic frameworks (MOFs) have been named as one of “the top ten emerging technologies in chemistry” by the International Union of Pure and Applied Chemistry (IUPAC).¹ The large number of possible metal nodes and organic ligand variations for synthesis of MOFs leads to attractive physical and chemical features such as high thermal stabilities (as high as 500 °C), various porosities (0.3–0.9), large surface areas (>8000 m² g⁻¹), low densities (0.2 g cm⁻³), and wide range of pore sizes (3–100 Å).^{2,3} The ability to tune physical and chemical properties of MOFs during or after synthesis by metal cation exchange,⁴ attachment or insertion of functional groups,^{5,6} allowed the researchers to generate various types of MOFs with desired properties for a specific application. MOFs have been studied for many different areas such as gas storage and separation,^{7–10} bioimaging,¹¹ catalysis,^{12,13} batteries,¹⁴ supercapacitors,¹⁵ and drug delivery.¹⁶ Currently, there are 103 951 experimentally synthesized MOFs deposited into the Cambridge Structural Database (CSD),¹⁷ and this number is continuously increasing due to the existence of theoretically infinite number of possible MOF structures. This very large number of MOFs is a great advantage for the potential applications of materials but, on the other hand, it is not practical to experimentally test the performances of all these MOFs even for a single application at the lab scale. A high-throughput computational screening (HTCS) approach enabled researchers to evaluate the

performance of thousands of MOFs for a desired application in a time- and cost-efficient manner using molecular simulations.^{18,19} Simulation results of large numbers of MOFs help to identify the most promising material candidates for desired application purposes. Results of these HTCS studies reveal an enormous amount of data about both the physical (e.g., pore size, surface area) and performance properties (e.g., selectivity, working capacity) of MOFs. The challenge here is that it is nearly impossible to see all the trends and relations in this large MOF data without additional computational resources.

Machine learning (ML) aims to develop computer programs capable of learning from large data sets using statistics and some algorithms; it may help to identify the hidden patterns and relations in data or to construct models to correlate some input variables (descriptors) with the output (performance) of the system.²⁰ ML has been applied in various fields such as chemistry,²¹ chemical engineering,²² catalysis,²³ and energy related materials.^{24–27} In addition to the availability of new and more effective ML algorithms, developments in computational

Received: February 18, 2021

Published: April 29, 2021



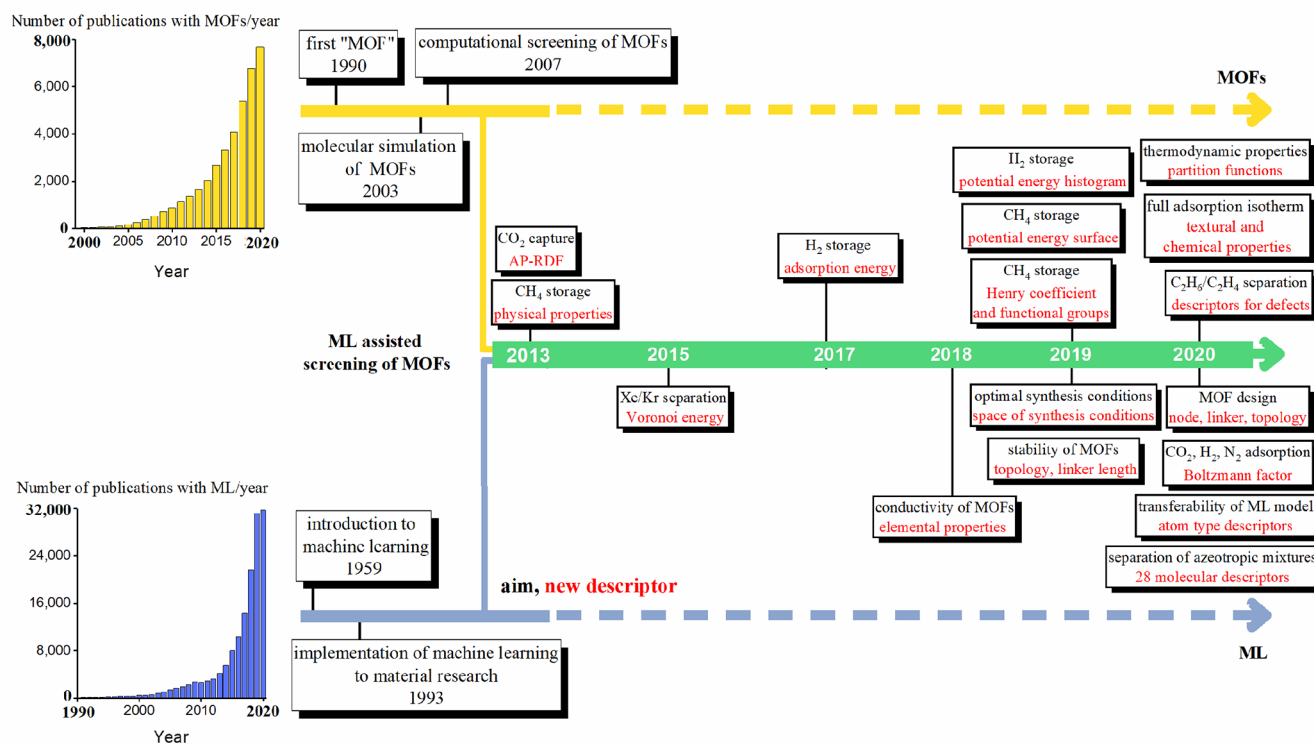


Figure 1. Timeline for developments in the areas of MOFs, ML, and ML-assisted screening of MOFs. The number of publications in MOF and ML fields per year is given in the inset figures (retrieved from the Web of Science⁸⁰ in February, 2021).

technologies including better data storage, management, and retrieval capabilities have contributed to the rise of ML. New experimental (like high-throughput experimentation tools and high-resolution spectroscopy)^{28,29} and computational tools (classical and quantum mechanical methods)³⁰ provide a large amount of accurate data, which is essential for ML applications, and in the meantime, advances in supercomputing capabilities, high-throughput computational workflow managers,^{31,32} and open source accessible software packages dedicated to ML^{33–38} make the field much more accessible to newcomers. ML has been applied in materials science^{39–41} for different purposes including the prediction of polymer properties,⁴² classification of zeolite structures,⁴³ discovery of drugs,^{44,45} identification of peptides as antibacterial agents,⁴⁶ design of homogeneous catalysts from different ligands,⁴⁷ or detection of biological effects of nanoparticles.⁴⁸ MOFs are a relatively new class of materials compared to polymers or zeolites, and ML has been recently applied to the field of MOFs to determine the best materials among the existing ones for a specific use, to discover new and better materials from practically endless possible structures, and to unravel the correlations in the data obtained from molecular simulations of MOFs. These data were used to build quantitative structure–property relations (QSPRs),⁴⁹ to predict the mechanical stability of structures,⁵⁰ to estimate gas storage capacity of MOFs,^{51,52} and to design novel MOFs by ML.⁵³ Analysis of MOFs with ML has accelerated in recent years^{25,54–73} for a variety of fields such as identifying electronic structure properties of MOFs,^{63,64} predicting colors of MOFs,⁶¹ defining the oxidation states of metals in MOFs,⁶² assigning partial charges to MOF atoms,^{59,73} optimizing the swing adsorption process conditions with MOFs,^{58,60} and for predicting the

performances of MOFs as sensors,^{55,56} heat pumps,⁵⁷ and gas storage and separation materials.^{66,74–79}

In this review, we first provide a brief introduction and background to the MOFs and ML methods and then address the current state of the art in ML applications of MOFs for gas storage and separation, which are the most studied application areas of MOFs to date, while the focus of most of the ML studies on MOFs is also on these two areas. We then address both the opportunities and the challenges in the intersection of two fast-growing fields and present a detailed perspective for the future of ML-assisted computational MOF studies.

■ BACKGROUND

As the future is shaped by the past, focusing on the past trends in the fields of ML and MOFs may guide us to make projections for the possible future developments on the subject. The timeline for studies on ML, MOFs, and ML applications on MOFs is given in Figure 1. It is hard to describe the detailed timeline for ML studies because most of the popular and relatively old ML techniques such as linear regression (LR), artificial neural networks (ANN), decision tree (DT), and support vector machine (SVM) have been used in the literature without directly referring to the ML concept in early applications. Although the first appearance of the ML concept in the Web of Science database⁸⁰ goes back to the 1950s, the frequency of appearance is limited to the order of tens per decade until the 1990s; the acceleration started in the 1990s. The revolutionary development in computational technologies coincided with the spread of the Internet which revolutionized data sharing capabilities has not only increased the use of ML but also shifted the focus to large data sets in recent years. If we exclude LR, which has wider (and older) application areas, ANN seem to be the most commonly

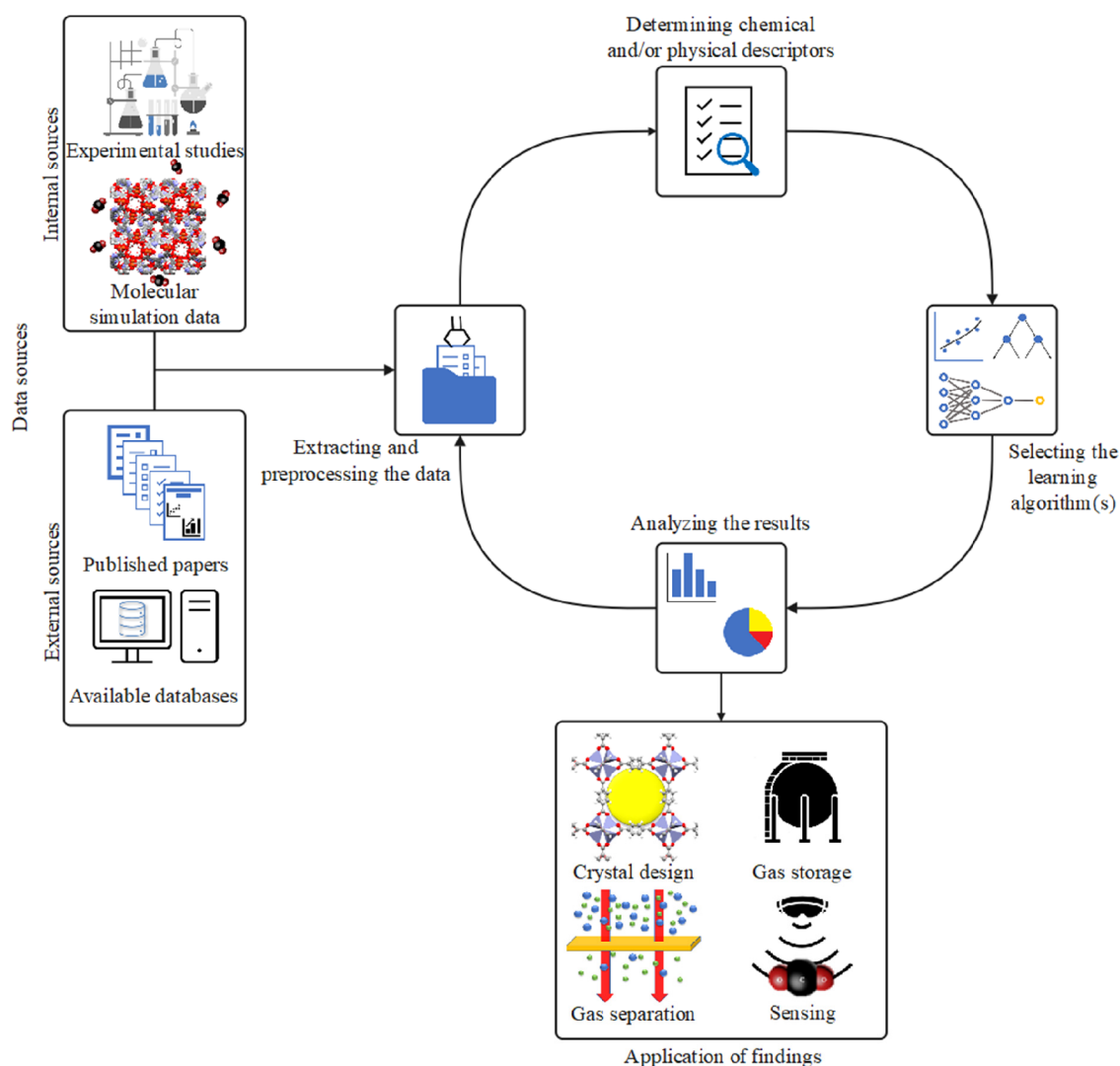


Figure 2. Generalized implementation of ML on MOF studies. The data source is selected, preprocessed, and the descriptors to correlate the data are determined. The data is then fed into selected ML algorithm(s) to predict properties of MOFs. The results obtained from ML models are utilized for various applications, and analysis of the results help to determine new and better descriptors for discovery of more accurate models.

implemented technique in chemistry since the 1980s; there are even review articles on the subject published in the 1990s. A paper, dated 1991,⁸¹ introduced the basic concept and implementation of ANN and reviewed the chemistry related applications such as analysis of spectroscopic data, determining amino acid sequences and protein structures, and classifying atomic energy levels. Burns and Whitesides⁸² (in 1993) reviewed the early application of ANN in chemistry. The first review covering the chemical engineering applications of ANN was published in 2000 focusing on the area of fault detection, prediction of polymer quality, data rectification, modeling and control.⁸³ Early applications of ML had been mostly based on small data sets generated in a single laboratory; although such works are still valuable, the attention started to shift to large data sets in the late 2000s, and this trend intensified in the 2010s with the increasing availability of large data sets.

The pioneering works on MOFs in the 1990s^{84–86} attracted the interest of many scientists from various fields to the potential of these materials. Due to this interest, the publishing rate on topics including “metal organic framework” has increased dramatically, reaching up to 7000 articles per year in 2019, and the number of experimentally synthesized MOFs

exceeded 100 000.¹⁷ Many of the early publications focused on the synthesis of new MOFs whereas studies after the 2000s were more directed to the applications.^{87–89} With the increasing demand to evaluate the performances of MOFs in a time-efficient and accurate manner to guide the experimental efforts, the first molecular simulations were performed in 2003.⁹⁰ Grand canonical Monte Carlo (GCMC) simulations have been very widely used to compute gas adsorption in MOFs. Computational screening of MOFs considering ten or more structures started in 2007 with the aim of evaluating the accessible surface area of materials.⁹¹ Establishment of the first hypothetical MOF database by Wilmer et al.⁸⁹ was the real kick-off for the screening of MOFs. Since then, many HTCS of MOFs were accomplished using that hypothetical MOF database⁹² or in-house databases of experimental MOFs.⁹³ The computation-ready experimental MOF database⁹⁴ was the real catalyst for HTCS of thousands of experimental MOFs using GCMC simulations. Generation of different databases involving various MOFs allowed the use of ML in conjunction with HTCS for structure prediction and estimating properties of MOFs as well as their potentials for gas storage and gas

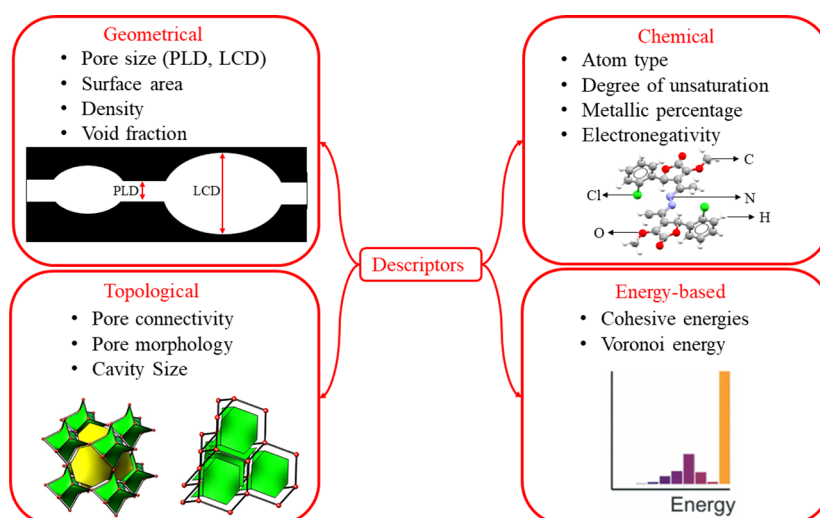


Figure 3. Classification of descriptors used in MOF related ML studies. Examples of descriptors are given for each class. The topological representations are taken from the Reticular Chemistry Structure Resource (RCSR).¹¹⁴

separation.^{31,33} We represented the generalized implementation of ML to MOF studies in Figure 2.

■ ML BASICS

ML is usually implemented through a framework involving *constructing a database, preprocessing of data, selecting descriptors, deciding on ML techniques, performing analysis, and interpretation of results* steps even though the number and wording of steps may change from source to source. One can find the details of such frameworks in various publications.^{20,95–97} Here, we briefly discuss three key ingredients (*techniques, descriptors, and data*) from the MOF perspective, and we summarize the major issues to be considered in *ML implementation*.

■ ML TASKS AND TECHNIQUES

ML techniques are developed to perform some specific tasks such as *regression, clustering, classification, or association*,⁹⁶ and these tasks are also used to categorize the ML techniques even though some techniques can be used to perform more than one task with some modifications. *Regression* involves the development of models correlating the input and output variables so that the output of a new input set can be determined. The predictive models, like *linear (or nonlinear) regression*, have been around since long before the ML concept itself, and many additional techniques such as least absolute shrinkage and selection operator (LASSO) regression, ridge regression, ANN, random forest (RF), and support vector regression have been developed through years.^{20,96–99} There are also variations of some techniques such as a recently popularized set of new ANN algorithms called *deep learning*.^{100,101}

Clustering refers to the task of grouping the data based on the similarities in input variables; it is usually used as a preanalysis tool before other ML techniques. Various techniques based on the centroid (like *k-mean clustering*), connectivity (like *hierarchical clustering*), distribution, and density are used for clustering the data.¹⁰² *Classification*, on the other hand, categorize the data set into subsets based on the output variable; this way one can identify the values (or ranges) of input variables leading to those subsets, and determine the possible category of a new input set. The

common classification techniques are decision trees, *k*-nearest neighbor algorithms, logistic regression, Bayesian classification, ANN, and SVM.⁹⁶ Finally, *association* aims to extract hidden relations among the variables (including output variables) in large data sets; it is used to identify the features (variables) that appear *together* in the same subsets.¹⁰³

The most important issue in selecting the ML technique is the task to be achieved. The characteristics of the data set (like size, homogeneity, type of variables, and so on) should be also considered.

■ DESCRIPTORS FOR MOFS

The descriptors, which are also called *input variables, factors, features, or fingerprints*, should be able to describe the systems in a way that the differences in output variables due to the change in descriptors could be explained in desired accuracy. In other words, the selected descriptors should be strongly correlated to the output variable. In material research, including MOFs, the selection of the descriptor set can be quite complicated due to the presence of a large number and diversity of potential candidates. Ward et al.¹⁰⁴ identified 145 potential materials descriptors and classified them into four groups, *stoichiometric attributes, elemental property statistics, electronic structure attributes, and ionic compound attributes*, while researchers also categorize the descriptors as *atomic/molecular* and *structural* properties.^{105,106} While the atomic/molecular descriptors are used for the intrinsic properties such as molecular weight, elemental composition, atomic/molecular radius, charge, and so on, the structural descriptors refer to bulk properties such as crystal structure, pore size, and surface properties.

The final descriptor set will vary from model to model even for the same data set depending on the knowledge to be extracted. However, the smaller ML models (i.e., less descriptors) are more robust and informative;²⁰ hence, the number of descriptors is usually reduced with an action called *dimensionality reduction*. This may be done by either eliminating some less significant descriptors (*feature selection*) or creating a smaller number of new descriptors from the original list (*feature extraction*)²⁰ using a method like principal component analysis. Some new approaches are also provided such as the ML neural-network representation of density

functional theory (DFT) potential-energy surfaces to describe the energy and forces as a function of atomic positions,¹⁰⁷ smooth overlap of atomic positions.¹⁰⁸ The atomistic structure-learning algorithm¹⁰⁹ and many-body tensor representation¹¹⁰ were also suggested to create the descriptors representing the atomic/molecular systems better.

Various descriptors from the synthesis conditions (synthesis techniques and parameters, solvent types, microwave power, organic linkers, and metal precursors) to properties obtained from ab initio calculations (cohesive energies and force components) for the accurate molecular simulations¹¹¹ have been also used for MOF research. The classification of descriptors such as elemental versus structural⁵⁹ or chemical versus structural^{50,112} was also used while Chong et al.²⁶ grouped them as geometrical, chemical, topological, and energy-based as illustrated in Figure 3. We classified the descriptors as user defined (such as synthesis method, linker, metal, and so on) versus structural (like void fraction, pore diameters, and so on) in one of our works to draw attention to the idea that the structural properties are the results of user defined descriptors; hence, strong correlations, which should be avoided in some ML techniques, may exist between two groups.¹¹³ There are also newly created descriptors like atomic property weighted radial distribution functions (AP-RDF)⁵¹ to predict the gas uptake more accurately.

■ SOURCES OF DATA

The most common approach in ML analysis of MOFs is to acquire the MOF related information from the experimental sources such as the computation-ready experimental (CoRE) MOF database^{94,115,116} or computationally produced data such as hypothetical MOFs,^{89,117–119} while the applications (such as gas uptake) are usually computed in house (such as with GCMC simulations).

The most critical issue related to the MOF data is that HTCS^{120–122} and ML^{51,123} on MOFs generally focus on computation-ready databases; however, these databases consist of different MOFs with different properties, and the range of properties in one database may have huge impacts on the transferability of results or models to other MOF databases. Moosavi et al.¹²⁴ used ML to quantify the structural and chemical similarities of MOFs in different databases, to clarify if synthesizing a new MOF will be able to increase the variety of MOFs. They used CoRE MOF¹¹⁶ and hypothetical MOF databases of different groups^{89,117,125,126} for this purpose. The Henry's constants for CO₂ and CH₄ and gas uptakes at various pressures were computed using GCMC simulations and results revealed that each MOF database has distinctly different distributions of properties. Furthermore, the importance analysis resulted in different ranking of the descriptors for gas adsorption if the ML models were constructed using different MOF databases. These results indicate the possibility of biases in ML analysis and question the transferability of models and results among the databases while it may also indicate the necessity of agreed upon standard protocols in generating, reporting, and storing the MOF data.

■ CRITICAL ISSUES IN ML IMPLEMENTATION

The most critical step in implementation of ML is descriptor selection, which may involve feature selection or feature extraction for choosing proper descriptors. If the selected descriptors are not predictive of the target variable and do not

capture the right information, the accuracy of results will be poor regardless of the model. Model selection is another crucial step in implementation of ML, which involves the construction of various candidates, with various values of model hyperparameters (like number of hidden layers, number of nodes, and learning rate in ANN) and selecting the one that represents the data best. This is usually performed by constructing (training) candidate models using some portion of data with a screening strategy (like grid search) for the values of model hyperparameters within plausible ranges, and testing the models with the remaining (unused during model building) data; the performance of the candidate models is evaluated using some measures for fitness (like root-mean-square error) and the one that represents the data best is selected as the final model. The model selection step should be executed without data leakage, which refers to the transferring information from testing (i.e., using the same information in training and testing). The simplest procedure, which is called holdout testing, divides the data into two subsets (for example 80% and 20%) randomly; the large subset is used to construct (train) the model while the small subset is used for testing. Two important mistakes are often made during this procedure. First, the training and testing sets may contain the same information; this usually happens when pairs of data points with very similar information are divided into training and testing resulting in overtraining and, therefore, poor generalization ability of the model. The second problem is that the researcher may not like the testing performance of the model and repeat the procedure by changing model parameters until the desired testing results are obtained; this also represents leakage because the training and testing are not independent anymore. *K-Fold cross validation* may allow building of a feedback loop during model construction without leakage. This procedure also requires the separation of testing set (let say 20%) before model construction. The remaining data is further divided into *k* subsets; *k* – 1 subsets are used for training while the remaining one is used for validation, in rotation, to tune the model hyperparameters. The model (i.e., the set of model hyperparameters) that resulted in the best average performance is selected as the final and tested using the testing data separated in the first step. Unfortunately, the validation is sometimes treated as the testing and used for both model selection (determining the model hyperparameters) and testing the model fitness, again, causing data leakage. Another possible cause of information leakage is the normalization/standardization of data before splitting into training and testing (carrying information related to the testing data to the model construction); to prevent this, the normalization/standardization should be applied to the training data set, and then the testing set should be normalized using the mean and variance of training.

Preprocessing of data before the ML modeling can be also crucial for the success of analysis. To begin with, the missing values, duplications, or inconsistencies in the data, which are especially common in the data sets constructed from multiple sources, should be eliminated (inconsistencies among the MOF data, as reported by Moosavi et al.,¹²⁴ were briefly discussed in the previous section). Various transformations on the descriptors (like normalization, standardization, or encoding) may also be needed while the dimensionality of the data can be reduced for smaller and more robust models as discussed above. Additionally, some techniques may require more specific actions; for example, the unbalanced data is a

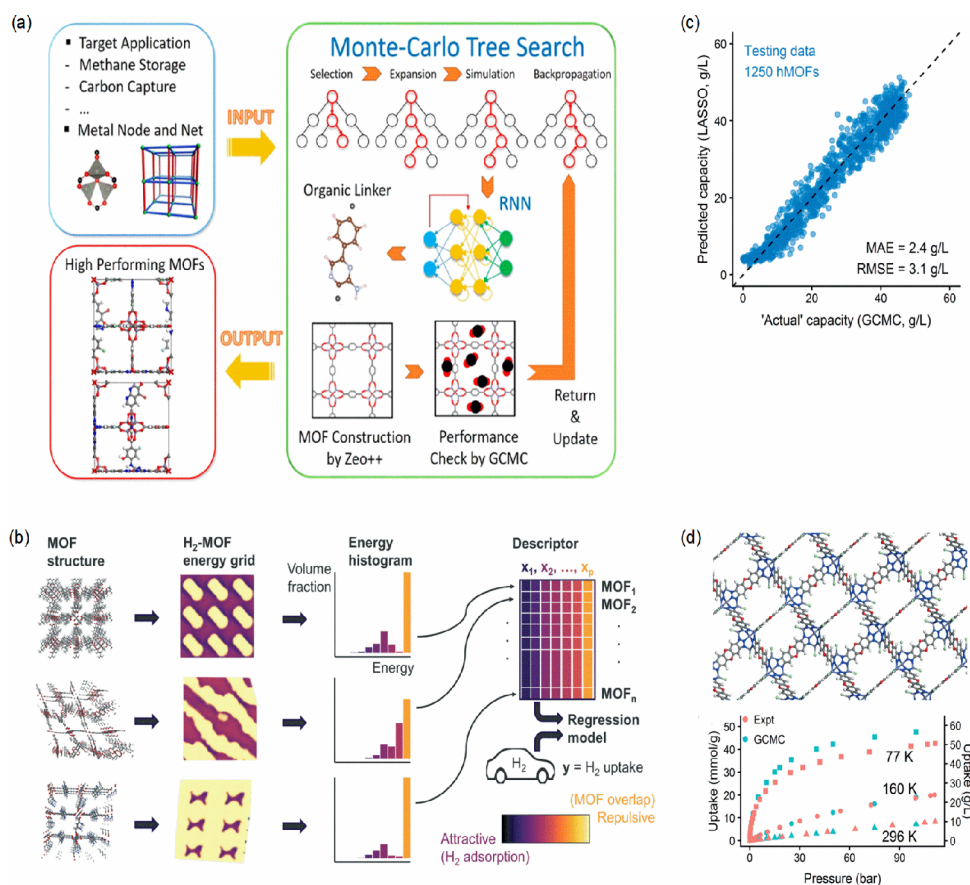


Figure 4. Design and discovery of new MOFs for gas storage. (a) Flowchart of tailor-made design of MOFs with machine learning for methane storage and carbon capture (Reproduced with permission from the work of Zhang et al.⁵³ Copyright 2020 American Chemical Society). (b–d) Workflow of the ML algorithm of Bucior et al.⁵² using H₂-MOF energy histograms as descriptors (Adapted with permission from ref 52. Copyright 2019 Royal Society of Chemistry).

serious problem in DT classification; briefly, the number of data points in classes should be approximately the same. Otherwise, the incorrectly classified instances from large classes are placed into the neighboring small classes, and this decreases their prediction accuracy because, even if the fraction of misplaced data is small in its source (large class), it may constitute a significant portion of data in the target (small class). If the use of equal size classes is not practical, the problem can be fixed by random sampling (duplicating some data points) to increase the size of the small classes approximately to the size of the largest.¹²⁷ Unbalanced data (i.e., a heavily tailed data set) may be challenging for the predictive models as well.

The error should be also analyzed and reported properly to assess the true potential of ML models. Various measures such as root-mean-square error (RMSE), percent accuracy, correlation coefficient (R), coefficient of determination (R^2), recall, and precision are used depending on the ML technique employed; the most common mistake on this issue is to report the validation error as testing error as the result of the confusion discussed above. Additionally, the distinction between two types of uncertainty should be also made. One type (also called as *aleatoric uncertainty*) is associated with the data set; all experimental measurements (and most of the computational tools) have certain levels of uncertainty. This type of uncertainty cannot be reduced during ML application. The second type of uncertainty (also called as *epistemic uncertainty*) arises from the lack of knowledge in practice even

though the knowledge exists in theory; for example, the failure to identify an important descriptor or lack of sufficient representation of certain effects in the data set may be considered in this category. This type of uncertainty may be reduced by designing better descriptors which can be strongly related to target variables, and it should be done as much as possible.^{128,129}

■ IMPLEMENTATION OF ML ALGORITHMS FOR MOFS

ML has generally been used to predict the stability of MOFs, to unlock the gas storage and gas separation potentials of MOFs, and to design novel MOFs. Among various applications, the most mature application field of MOFs is gas storage and separation. ML studies on gas storage and separation performances of MOFs generally focus on structure–performance relationships to select the best descriptors and/or introduce new ones that can accurately predict the gas uptakes and selectivities of MOFs in a time-efficient manner. We presented some recent, representative studies with a concise and forward-looking overview.

■ DESIGN AND DISCOVERY OF NEW MOFS

ML has been used as a tool for predicting the structural properties of MOFs and generating well performing MOFs. For example, the SVM algorithm was used with molecular descriptors to define the most important properties that can

lead to porous structures, and 481 porous, mechanically stable structures were identified among 156 333 CSD-derived experimental crystal structures.¹³⁰ The surface area descriptor provided an important insight for determining porous and stable structures prior to experimental studies. Also, 3385 hypothetical, computer-generated MOFs (containing 14 types of organic ligands, 28 different types of organic or metal-based nodes, and 41 topologies) were used to predict bulk and shear moduli from chemical–structural properties via combining HTCS with the ANN algorithm.⁵⁰ Topology and coordination number were identified as the most important factors for the mechanical stability of MOFs as they determined the energy cost for the changes in bond lengths and bond angles. Following the comprehensive investigation of bulk and shear moduli with respect to structural properties, a web-based tool was provided for researchers to examine structure–mechanical stability relations of MOFs.¹³¹

The failed experiments, which are basically the synthesis conditions that did not lead to porous, crystalline MOFs, have been also considered as the source of data as the parameters and conditions leading to failure should be also known.¹³² In such a work, robotic MOF synthesis was combined with genetic algorithm (GA) and chemical intuition obtained from failed experiments was quantified for the first time, to extend the knowledge on how the building blocks self-assemble into one of the most widely studied MOFs, HKUST-1, at different synthesis conditions.¹³³ The most optimal conditions, such as the type of solvent used, reactants ratio, microwave power, and reaction time as well as their relative importance for the synthesis of HKUST-1 with the largest surface area were determined using the RF algorithm. The model predicted the synthesis outcome depending on the synthesis conditions with a testing error of 14%. This approach highlighted the importance of reporting the information obtained from failed experiments, and it may lead the way to the development of transferable synthesis methodologies that can be used to synthesize different MOFs with desired porosity and surface area without tedious repetition of experiments at many different conditions.

Industrial application of MOFs mandates stability under humid and/or aqueous conditions, which is not straightforward to determine with experimental and computational methods. SVM, RF, and GB algorithms were utilized to screen MOFs for their water stability using three categories of descriptors: metal node, organic linker, and molar ratio of the number of organic linkers, oxygen, hydroxyl, and water groups to the number of metals in MOFs.¹³⁴ RF and SVM models were trained with the experimentally determined water stability data of 207 MOFs,¹³⁵ and accurately predicted the water stability of 10 MOFs when atomic radius and ionization potential of the metal ion and number of cyclic divalent nodes and six-membered rings were used as the descriptors. That work is a good example of the utilization of ML models to prioritize the experimental synthesis of stable MOFs.

Recently popularized deep learning algorithms have been also utilized in MOF research. Combining Monte Carlo tree search with the recurrent neural network (RNN) algorithm, as shown in Figure 4a, is a useful approach to design new well-performing MOFs with high density of adsorption sites for CH₄ and CO₂ gases.⁵³ Performances of the designed MOFs were tested for CO₂ and CH₄ storage after they were hypothetically produced. Due to the design of pore space and organic linkers, hypothetical MOFs were found to have

higher deliverable CH₄ capacity, the amount of gas that can be stored between adsorption and desorption pressures, and higher CO₂ uptake than the existing MOFs having the same metal node and topology.

■ GAS STORAGE PERFORMANCES OF MOFS

CH₄, H₂, and CO₂ storage capacities of MOFs are in scope since the development of an appropriate adsorbent with optimum storage and delivery capacities for these gases requires knowledge of the factors defining the materials' performance. Consequently, the works published mainly aimed to identify the significant chemical or structural properties or select the most suitable MOF structure. While the performances of various ML techniques are compared in some works, there are also reports that the ML techniques were used together in a complementary manner to enhance the knowledge extracted.

Most ML studies have focused on the CH₄ storage capacity of MOFs. For instance, CH₄ adsorption data of 137 953 hypothetically designed MOFs were computed using GCMC simulations.⁸⁹ These MOFs were later used to develop QSPR models to reveal relations between CH₄ storage capacities of MOFs and simple geometric descriptors such as pore size, surface area, and void fraction.⁴⁹ Using the SVM models, new hypotheses about combinations of material properties that might lead to very high CH₄ storage capacities were proposed. Simple multivariable linear models were also successfully used to investigate the structural factors that determine CH₄ uptake capacity of a small number of experimentally reported MOFs at various pressures.¹¹² According to these models, high gravimetric surface area, ~6000 m² g⁻¹, and high porosity, ~0.9, were required to reach the Department of Energy (DOE) target for CH₄ uptake capacity (0.5 g g⁻¹) at 35 bar. This analysis was extended to CH₄ uptake data of 2224 experimentally synthesized MOFs using DT and ANN algorithms to determine the effect of user-defined descriptors (such as metal type, crystal structure, metallic percentage) and structural descriptors (such as pore volume, maximum pore diameter) on deliverable CH₄ capacity.¹¹³ An ML algorithm on accurately predicting CH₄ uptake capacities of 4763 CoRE MOFs and 69 840 hypothetical covalent organic frameworks (COFs) at 5.8 bar and 298 K was proposed.¹³⁶ In the proposed approach only 100 materials are used in the training set to predict the top materials with the highest CH₄ uptake capacities. The predicted top materials are then added to the training set, and a new prediction is made. This procedure was repeated until the training set included the last predicted 100 top materials. The model was able to identify most of the top 100 materials in both CoRE MOF and hypothetical COF databases using only a small fraction of materials in the subsets.

Selection of the most suitable ML tools may depend on the size and structure of data; hence a significant number of researchers have compared the performances of various techniques and selected the one that has the highest fitness. In such a work, a topological barcode system,¹³⁷ including information about the pore geometry of materials, was combined with different ML algorithms, including RF, DT, kernel ridge regression (KRR), and SVM, to predict structural properties and deliverable CH₄ capacities of zeolites and MOFs, which were previously computed with GCMC simulations.^{138,139} Structural descriptors included void fraction, surface area, density, pore diameter, and information about interpenetration while chemical descriptors included atom

type, degree of unsaturation, metallic percentage, oxygen to metal ratio, nitrogen to oxygen ratio, and electronegativity. The RF model using both chemical and structural descriptors led to the most accurate CH₄ uptake predictions. The use of the RF algorithm was extended to 69 839 hypothetical COFs to predict their deliverable CH₄ capacities utilizing structural and chemical descriptors.⁷⁵ The use of chemical and structural descriptors increased the accuracy of ML predictions for COFs, especially at low pressure. Those works emphasized the need for the consideration of chemical descriptors for accurate prediction of gas uptakes of nanoporous materials at low pressures.

Similar works have been performed for other gases as well. For instance, the RF algorithm was utilized to predict the CO₂, H₂, and H₂S adsorption capacities of 2932 CoRE MOFs using structural properties of MOFs along with a new descriptor, the probability of a set of different probe atoms to be adsorbed by the material.¹⁴⁰ Three different probe atoms were modeled for adsorption probability calculations: Vprobes (neutral, nonpolar probes), Dprobes (neutral, small dipole moment), and Qprobes (small charge in the center). The number of MOFs used in the training set was varied from 50 to 1000 while the rest of the MOFs were used for the validation set. The model was built considering structural properties, and the adsorption probability of Vprobes and Dprobes had the best R² value for the adsorption capacity of each gas (0.92 for CO₂, 0.94 for H₂S, and 0.97 for H₂). This work showed that the adsorption capacities of MOFs for different gases can be predicted by using adsorption probability of pseudo atoms without the need for expensive simulations. Similarly, a large variety of ML techniques such as multilinear regression (MLR), DT, kNN, SVMs, ANNs, and RF was used to develop QSPR models based on binary classifiers that are built from void fraction, surface area, and pore size of MOFs to predict CO₂ uptake capacities of hypothetical MOFs.¹⁴¹ The model built with the RF algorithm had the highest accuracy (over 94%) for uptakes of both gases. Moreover, the model identified over 60% of high-performing materials for CO₂ and N₂ adsorption in a large and diverse set of ~65 000 MOFs, setting a great example that ML techniques can lead to a significant time efficiency in screening large databases. In another work, H₂ adsorption isotherms obtained from GCMC simulations were coupled with the NN algorithm to analyze the limits of H₂ storage in >850 000 nanoporous materials including MOFs, COFs, ZIFs, PPNs, and zeolites.¹⁴² Results revealed that H₂ storage capacities of hypothetical materials could not exceed the performances of the experimentally synthesized materials reported to date. This was a motivating result for the research community working on adsorbent materials to change the common approach for designing better materials.

With the aim of finding materials with optimum binding energy to reach high deliverable H₂ capacities, the LASSO algorithm was used to develop a model which uses the guest–host energy histograms as descriptors instead of structural properties to predict deliverable H₂ capacities between 100 and 2 bar, at 77 K.⁵² The overall procedure of the analysis is given in Figure 4b while the H₂ adsorption capacities obtained from GCMC versus ML for testing is given in Figure 4c indicating that the model is quite successful. GCMC simulations were performed for H₂ uptake of 137 953 hypothetical MOFs. The adsorbate–MOF potential energy landscape was sampled, and the interaction energy between the framework and an H₂ probe was computed to construct the energy histogram

creating the feature matrix for the hypothetical MOFs. The LASSO model was used to predict the adsorption energy from these features. The model was then used to predict the H₂ uptake of 54 776 experimentally synthesized MOF structures in the CSD. Finally, the model was verified experimentally using the H₂ capacity of MFU-4l, one of the top identified MOFs with the ML model for H₂ storage capacity, and a good agreement was obtained as shown in Figure 4d.

To define the extent of volumetric H₂ storage capacity depending on adsorbent and operating conditions, Anderson et al.¹⁴³ used an ANN algorithm to predict H₂ loadings of 105 hypothetical MOFs with 17 different topologies at multiple temperature and pressure conditions. Under nonisothermal conditions, changing the storage pressure from 100 to 35 bar did not significantly affect the deliverable H₂ capacity of a top MOF while changing pressure under isothermal conditions led to a more significant drop in deliverable H₂ capacity. That work demonstrated that it is possible to arrange the set of conditions for a material to acquire the target performance, and ML can provide a very fast means of arranging conditions.

With the experience gained on application of ML algorithms for gas storage in MOFs, ML applications have been extended to the different classes of nanoporous materials, different gases, and more effective descriptors and procedures. For instance, the ANN algorithm was used to develop a model to predict adsorption isotherms of 2400 topologically and chemically diverse MOFs for small, near-spherical, nonpolar, and mono- and diatomic adsorbates (alchemical species) at different pressures.⁷⁹ The model efficiently predicted Ar, CH₄, Kr, Xe, N₂, and C₂H₆ adsorption in MOFs. So that, for the first time via combining the descriptors of both adsorbate and adsorbents in the same ML model, aiming to make predictions for any new guest/host system without any prior knowledge of the components, the ML predictions for various gases were made without the need of providing any specific descriptors during the training of the model. The use of new molecular descriptors, atomic property weighted radial distribution functions (AP-RDFs), captures geometric features of materials while they utilize tabulated atomic properties such as electronegativity. To build QSPR models for CO₂, CH₄, and N₂ uptake of MOFs, the MLR and SVM algorithms were used with descriptors based on AP-RDF.⁵¹ Approximately 58 000 unique hypothetical MOFs⁸⁹ out of ~83 000 were used to train and calibrate SVM and MLR algorithms while the remaining MOFs were used for training. QSPR models developed with the SVM algorithm generally predicted simulated gas uptake with a higher accuracy than MLR. A comprehensive QSPR model that can predict high-performing (CO₂ uptake > 1 mmol/g at 0.15 bar and > 4 mmol/g at 1 bar) and low-performing MOFs, depending on the CO₂ uptake at conditions relevant to postcombustion (0.15 bar) and landfill gas purification (1 bar), was developed for 324 500 hypothetical MOFs.¹⁴⁴ CO₂ uptake data calculated with GCMC simulations were used to train and validate the ML models. 10% of the database was used to train the SVM algorithm again with AP-RDF descriptors while the results of the developed QSPR model was validated using the rest of the database. The QSPR model successfully predicted 945 high-performing MOFs (CO₂ uptake capacity > 1 mol/kg) out of 1000 top MOFs identified with GCMC at 0.15 bar. The work by Guda et al.,¹⁴⁵ which involves the use of ML to investigate the structural parameters of CO₂ adsorption on a CPO-27-Ni MOF using its X-ray absorption near-edge structure (XANES)

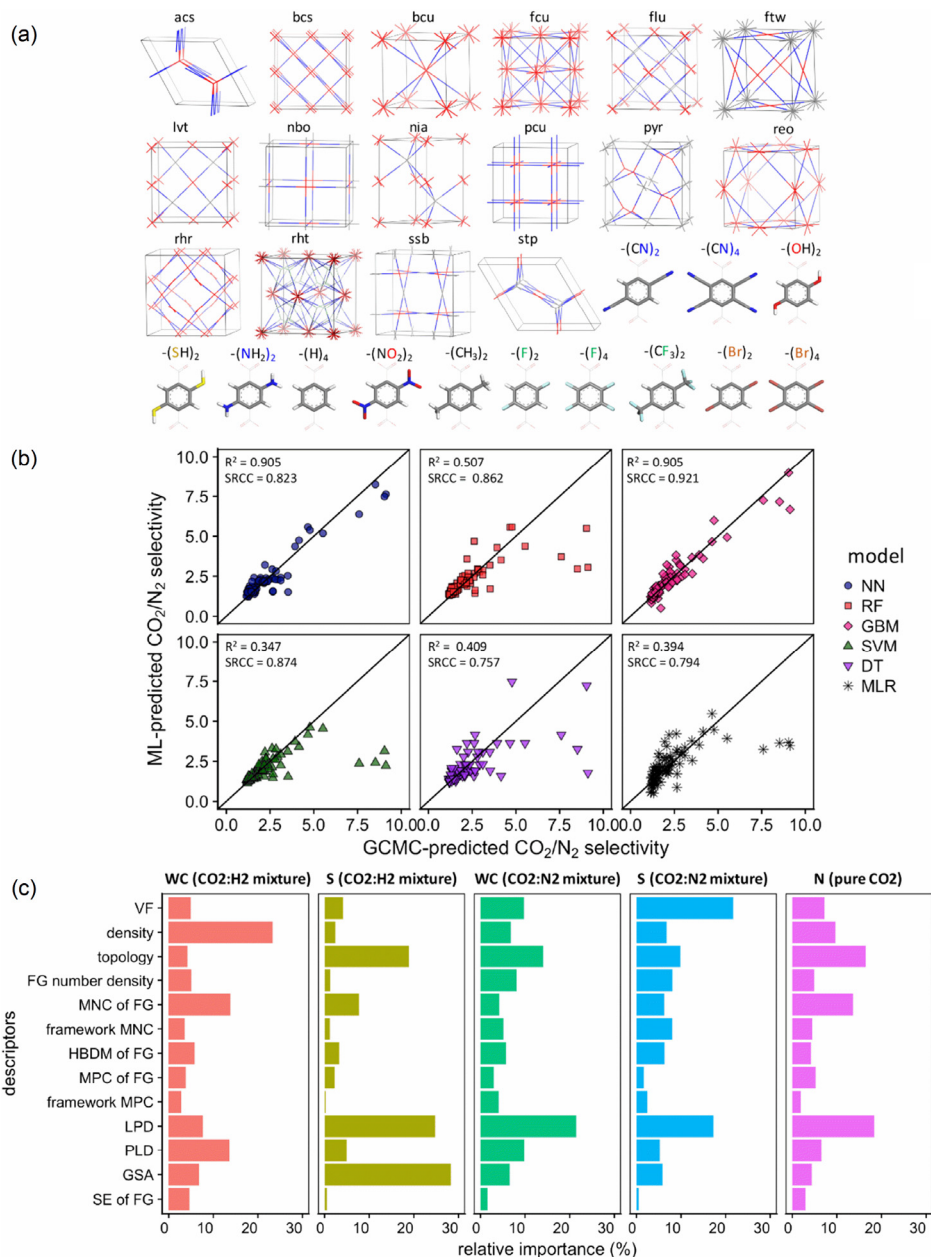


Figure 5. Effect of pore chemistry and topology on CO_2 capture performances of MOFs. (a) Topological nets and molecular building blocks used to construct MOFs. (b) Comparison of different model predictions with GCMC results for CO_2/N_2 :15/85 mixture selectivity of MOFs.¹²⁶ (c) Relative importance of descriptors obtained from GBM training for CO_2/H_2 and CO_2/N_2 separation performances of MOFs (Reproduced with permission from the work of Anderson et al.¹²⁶ Copyright 2018 American Chemical Society).

spectra, indicates that the applications may extend to more diverse areas of MOF research in the future.

■ GAS SEPARATION PERFORMANCES OF MOFS

MOFs have been widely investigated for gas separation because of their high porosities, chemical tunability, and diversity. However, screening large databases of MOFs to find well-performing materials is very time-consuming; hence ML techniques were used in recent years to predict the gas separation performances of large numbers of MOFs and to identify the structural parameters for designing high-performing MOFs.

One of the common ML applications in this area is the analysis of CO_2 capture capabilities of MOFs. For example, the

effect of pore chemistry and topology on CO_2 capture performances of 400 hypothetical MOFs was studied by combining molecular simulations with ML models.¹²⁶ The computational data for building the ML models was generated by GCMC simulations for pure CO_2 adsorption, CO_2/H_2 , and CO_2/N_2 mixture adsorption by mimicking industrial conditions. To be able to focus on a particular region of the MOF structure-space, a set of MOFs encompassing all possible combinations of 16 topologies and 13 functionalized molecular building blocks was considered as shown in Figure 5a. Various ML algorithms such as MLR, SVM, DT, RF, and NN were used to build predictive models for CO_2 capture metrics of 31 parent MOFs and their derivatives while the DT algorithm was used to predict the improvement or deterioration of CO_2 capture performances of MOFs after functionalization.

GCMC-computed versus ML-predicted CO₂/N₂ selectivity for testing is given in Figure 5b. The relative importance of descriptors changes significantly with the change of gas mixture composition as shown in Figure 5c. Results also revealed that functionalization of MOFs with thiol, cyano, amino, and nitro groups often improved CO₂ capture performances of MOFs depending on their topology. That work elucidated the importance of pore chemistry for determining the gas separation potential of MOFs with simulations and ML models. In another work, the gradient boosted trees regression (GBTR) method was used to predict CO₂ working capacity and CO₂/H₂ mixture selectivity of a topologically diverse set of hypothetical MOFs.¹⁴⁶ The data required to build the models was generated by GCMC simulations on the hypothetical MOFs for separation of CO₂/H₂ mixtures. Six different geometric descriptors and three chemical descriptors, which were constructed using AP-RDF, were used. The chemical descriptors were found to be more important for accurately predicting adsorption performance of MOFs for precombustion CO₂ capture.

Separation performances of MOFs can be improved wisely using different ML algorithms even before experimental synthesis. A deep generative model, supramolecular variational autoencoder (SmVAE), was used to automate the design of MOFs in a way that increases their performances for CO₂/CH₄ and CO₂/N₂ separation.¹⁴⁷ First, MOFs were deconstructed into their building blocks to obtain the collections of edges, vertices, and topologies following the methodology provided by Bucior et al.,¹⁴⁸ and then these were analyzed using descriptive statistical tools. Using the edges, vertices, and topologies, two million MOFs were generated, and the top MOF candidates, which were identified by GCMC simulations, had higher/similar working capacity and selectivity than the best-performing MOFs reported in the literature.

Capture of dilute CO₂ from air could be a useful process to prevent climate change. For that purpose, the CO₂ capture potential of 6013 CoRE MOFs from air was predicted using four different ML algorithms (back-propagation neural network (BPNN), DT, RF, and SVM) with structural descriptors such as pore sizes, volumetric surface area, and heat of adsorption (Q_{st}).⁷² To train the models, a maximum of 1000 MOFs were used while the remaining MOFs were used for the validation of the ML models. The ML model built with RF algorithm was reported to have the best prediction performance of CO₂ adsorption selectivity ($R^2 = 0.98$) for the validation set and the relative importance analysis revealed that Q_{st} is the most important parameter to predict adsorption selectivity for CO₂. The excellent agreement of ML results with simulation results in that work highlights the potential of using ML models to achieve rapid and accurate screening of MOFs for CO₂ capture even from dilute streams like air.

ML techniques have been also applied to predict the gas separation performances of MOFs for other relevant gas mixtures for industry such as Xe/Kr separation.¹⁴⁹ The RF algorithm was used to build the models considering six common structural descriptors as well as a newly developed descriptor, Voronoi energy of Xe, which is the average energy of an Xe atom at the Voronoi nodes of the accessible pore space. The data for training the models were generated by GCMC simulations for 15 000 randomly selected structures among 670 000 porous materials (MOFs, PPNs, ZIFs, COFs, and zeolites) acquired from various sources, and Xe/Kr selectivities of the remaining structures were predicted by the

models. Molecular simulations were then performed for 20 000 structures which the ML model deemed to be high performing, Xe/Kr selectivity >11, to obtain more accurate results from GCMC simulations. Many materials were predicted to have better separation performance for Xe/Kr mixture than a leading material, CC3. However, there was not a simple recipe for the geometric descriptors to guarantee that a material would be good for Xe/Kr separation.

Defects in MOFs can provide different adsorption sites for gas molecules, or facilitate gas diffusion via increasing the pore size, and can change the separation performances of MOFs. The significance of defects for ethane/ethylene separation performances of 425 hypothetically created MOF (UiO-66) structures with missing linker defects were also studied by GCMC simulations and ML algorithms.¹⁵⁰ Among six different ML models trained with structural and chemical descriptors, LR model accurately predicted the working capacity, selectivity, and shear modulus ($R^2 > 0.98$) while RF model successfully predicted the bulk modulus ($R^2 = 0.92$). Gravimetric surface area and pore volume were found as the two most important descriptors to predict the defect concentration of UiO-66 MOFs. LR model generally led to good predictions while kNN and SVM models provided overall good fittings.

■ OUTLOOK

In the light of astonishing developments in experimental synthesis techniques, computational power, computation-ready material databases, and data sharing capabilities observed in the last two decades, we can expect that the use of ML techniques in MOF research, as any field of science, will likely increase simply because we will have much more effective algorithms and computational tools in the future. The availability and quality of data will grow as a result of continuously improved experimental and computational research tools. The recent trends such as *open-access* publications as well as increasing data sharing options such as *data repositories* and *new databases* will also contribute to data availability while the existing databases can be expected to transform into more ML friendly formats (i.e., more suitable for automated data extraction). Additionally, more established databases, similar to those in material research such as Material Project,⁴¹ OQMD (Open Quantum Materials Database),¹⁵¹ AFLOWLIB (Automatic Flow),¹⁵² Computational Material Repository,¹⁵³ and AiiDA (Automated Interactive Infrastructure and Database for Computational Science)¹⁵⁴ are likely to be developed in MOF specific research as some examples are emerging: the MOFDB (MOF database of Snurr Research group),¹⁵⁵ Quantum MOF database,⁵⁸ and Nanoporous Materials Adsorption Energy Database.¹⁵⁶

The ML techniques usually require a large data set to reduce the uncertainty in models developed or knowledge extracted, and creating such large data sets are difficult and costly; hence, the new ML algorithms and approaches that can be used with small data sets are also searched,^{76,157,158} and this trend may be expected to grow in the future. Another approach that may be expected to spread more in the future is automated machine learning (AutoML), which aims to combine and automate the machine learning steps discussed above.^{159,160} It was addressed by one of the earliest studies on large scale screening of MOFs by ML methods that automated predictive analytics pipelines can learn predictive models that generalize to new, unseen MOFs.¹⁰⁰ As more people, including nonexperts, are

anticipated to involve in ML with the increasing popularity of subject and availability of suitable data, automated ML tools are also likely to be more preferred.

It is of the utmost importance to use the most updated, accurate, consistent, and comprehensive data set in ML studies. We expect future experimental and computational studies on MOFs to share their methodology and results in complete detail in a ML-friendly format (available to use with different scripts or software) in addition to graphs, figures, and tables for reproducibility and comparability of data utilized in ML studies. It is also important to report the results of the failed experiments and performances of bad-performing materials to achieve a comprehensive range of structure–property relationships using ML. Increasing the number of MOFs will bring a diversity in MOF properties such as chemical composition, topology, and pore sizes, and this will pave the way for new computation-ready experimental MOF databases. These MOF databases will enlarge the scope of HTCS studies to a wider range of MOF properties. However, MOFs reported in different computation-ready databases may not have the same range of structural properties,¹²⁴ and particularly, computational databases derived from experimental crystal structures do not contain all perfectly clean structures.^{161–163} Therefore, it is important to consider the chemical diversity and consistency of MOFs in different libraries when assessing the results of ML studies. It can be useful if future ML studies combine and examine MOFs from different databases which would help to build generalized ML models for MOFs.

The choice of informative, representative, and uncorrelated descriptors is important for ML studies on MOFs. We expect that future ML studies will consider and utilize various chemical descriptors widely. In terms of physical properties, topological descriptors provide valuable insights and can effectively be used for gas storage and separation studies. It is highly expected that ML studies for MOFs will continue to mostly use the results of HTCS studies in the future. Therefore, the assumptions and approaches used in the HTCS studies, such as selection of the charge assignment method and force field parameters, significantly affect the results. It would be very useful to utilize and compare different assumptions/approaches of HTCS to analyze their effects on the overall results of ML-assisted computational MOF studies. Defining force field parameters using ML, even for flexible MOFs, can also be a new research direction. ML-based methods can be used to predict molecular energies for small molecules with lower costs.¹⁶⁴ In the future, this might be possible for MOFs and increase the accuracy of their screening without requiring significant computational time. Even workflows can be designed to process large data sets of MOFs and identify the most promising MOFs for which further computational and experimental analysis can be made.

Combination of MOFs with other materials such as polymers or ionic liquids also brings an additional area of research which requires investigation with computational modeling and ML to gain valuable insight about distinct properties of these new MOF-based composite systems. Results of ML studies depend on the accuracy of various factors. It is only possible with experimental testing to determine the validity of ML results in identifying the top performing MOFs. Thus, stronger communication between experimental studies and theoretical studies is needed. Future studies examining the synthesizability of hypothetical MOFs

can significantly contribute into the experimental efforts and potentially accelerate the commercialization of MOFs if more studies focus on predicting mechanical properties of MOFs with descriptors such as organic linker and metal type, temperature, and pH of solvent pool before experimental synthesis. Boyd et al.¹⁶⁵ recently identified two MOFs among 325 000 hypothetical MOFs by utilizing the concept “adsorbaphore”, which describes the pore shape and the chemistry of the optimal binding site in a MOF and proved their high CO₂ selectivities from dry and wet flue gas with experimental methods. These topics can be further investigated by a collaboration of chemists, computer engineers, chemical engineers, and material scientists in many interdisciplinary studies while ML studies on extracting scientific knowledge from published articles also provide new insights for the synthesis and characterization of nanomaterials.^{166,167}

To conclude, ML-assisted HTCS of MOFs has great prospects for the future. MOFs have proven their potential for various applications including gas storage and separation, sensing, drug delivery, and catalysis, while the incorporation of ML to the research about MOFs is still a fresh topic with many unexplored areas. In this review, we considered ML studies which focus on the design and discovery of MOFs with high gas storage and separation performances. However, incorporation of ML into MOF research is not limited to only gas storage and separation applications. Nowadays, ML is utilized with acceleration to provide insights also for the use of MOFs in other applications. The most promising working fluid–MOF combinations are investigated with ML and structure–performance relationships revealed the best combination of properties required for the design of an efficient adsorption driven heat pump.⁵⁷ ML algorithms trained with MOF electronic noses to detect volatile organic oils can pave the way for the use of MOFs as sensors.⁵⁶ With the control of morphology and crystal growth of MOFs, ML can assist the synthesis of nanoscale MOFs in the desired morphology and properties.^{168,169} Specification of the electrochemical properties of MOFs according to the synthesis conditions can be determined using ML which facilitates the design of MOFs for Li-ion batteries.¹⁷⁰ In addition to these, drug delivery performance, conductivity, catalytic performance, toxicity, and biocompatibility of MOFs are yet to be thoroughly researched, and utilization of ML can provide useful insights for these topics. Thus, we are excited for the future advancements in this field which will guide many material scientists to discover new and fascinating aspects of MOFs.

Data and Software Availability. Since this is a review article, the authors do not report any data or software.

■ AUTHOR INFORMATION

Corresponding Authors

Seda Keskin – Department of Chemical and Biological Engineering, Koc University, 34450 Istanbul, Turkey;
orcid.org/0000-0001-5968-0336; Email: skeskin@ku.edu.tr

Ramazan Yildirim – Department of Chemical Engineering, Boğaziçi University, Bebek, 34342 Istanbul, Turkey;
Email: yildirra@boun.edu.tr

Authors

Cigdem Altintas – Department of Chemical and Biological Engineering, Koc University, 34450 Istanbul, Turkey

Omer Faruk Altundal – Department of Chemical and Biological Engineering, Koc University, 34450 Istanbul, Turkey

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.1c00191>

Author Contributions

[‡]C.A. and O.F.A. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

S.K. acknowledges an ERC-2017-Starting Grant. This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2017-Starting Grant, grant agreement no. 756489-COSMOS).

REFERENCES

- (1) Gomollón-Bel, F. Ten Chemical Innovations That Will Change Our World: Iupac Identifies Emerging Technologies in Chemistry with Potential to Make Our Planet More Sustainable. *Chem. Int.* **2019**, *41* (2), 12–17.
- (2) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gándara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; et al. Large-Pore Apertures in a Series of Metal-Organic Frameworks. *Science* **2012**, *336* (6084), 1018–1023.
- (3) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341* (6149), 1230444.
- (4) Caskey, S. R.; Wong-Foy, A. G.; Matzger, A. J. Dramatic Tuning of Carbon Dioxide Uptake via Metal Substitution in a Coordination Polymer with Cylindrical Pores. *J. Am. Chem. Soc.* **2008**, *130* (33), 10870–10871.
- (5) Banerjee, R.; Furukawa, H.; Britt, D.; Knobler, C.; O'Keeffe, M.; Yaghi, O. M. Control of Pore Size and Functionality in Isoreticular Zeolitic Imidazolate Frameworks and Their Carbon Dioxide Selective Capture Properties. *J. Am. Chem. Soc.* **2009**, *131* (11), 3875–3877.
- (6) Wang, X.-S.; Ma, S.; Rauch, K.; Simmons, J. M.; Yuan, D.; Wang, X.; Yildirim, T.; Cole, W. C.; Lopez, J. J.; Meijere, A. d.; Zhou, H.-C. Metal-Organic Frameworks Based on Double-Bond-Coupled Di-Isophthalate Linkers with High Hydrogen and Methane Uptakes. *Chem. Mater.* **2008**, *20* (9), 3145–3152.
- (7) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1), 1–9.
- (8) Tan, K.; Zuluaga, S.; Gong, Q.; Gao, Y.; Nijem, N.; Li, J.; Thonhauser, T.; Chabal, Y. J. Competitive Coadsorption of CO₂ with H₂O, NH₃, SO₂, NO, NO₂, N₂, O₂, and CH₄ in M-MOF-74 (M = Mg, Co, Ni): The Role of Hydrogen Bonding. *Chem. Mater.* **2015**, *27* (6), 2203–2217.
- (9) Alezi, D.; Belmabkhout, Y.; Suyetin, M.; Bhatt, P. M.; Weselinski, Lu. J.; Solovyeva, V.; Adil, K.; Spanopoulos, I.; Trikalitis, P. N.; Emwas, A.-H.; Eddaoudi, M. MOF Crystal Chemistry Paving the Way to Gas Storage Needs: Aluminum-Based Soc-MOF for CH₄, O₂, and CO₂ Storage. *J. Am. Chem. Soc.* **2015**, *137* (41), 13308–13318.
- (10) Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Velioglu, S.; Erucar, I.; Keskin, S. Database for CO₂ Separation Performances of MOFs Based on Computational Materials Screening. *ACS Appl. Mater. Interfaces* **2018**, *10* (20), 17257–17268.
- (11) Park, K. M.; Kim, H.; Murray, J.; Koo, J.; Kim, K. A Facile Preparation Method for Nanosized MOFs as a Multifunctional Material for Cellular Imaging and Drug Delivery. *Supramol. Chem.* **2017**, *29* (6), 441–445.
- (12) Müller, M.; Hermes, S.; Kähler, K.; van den Berg, M. W. E.; Muhler, M.; Fischer, R. A. Loading of MOF-5 with Cu and ZnO Nanoparticles by Gas-Phase Infiltration with Organometallic Precursors: Properties of Cu/ZnO@MOF-5 as Catalyst for Methanol Synthesis. *Chem. Mater.* **2008**, *20* (14), 4576–4587.
- (13) Gándara, F.; Gomez-Lor, B.; Gutiérrez-Puebla, E.; Iglesias, M.; Monge, M.; Proserpio, D.; Snejko, N. An Indium Layered MOF as Recyclable Lewis Acid Catalyst. *Chem. Mater.* **2008**, *20* (1), 72–76.
- (14) Chu, F.; Hu, J.; Wu, C.; Yao, Z.; Tian, J.; Li, Z.; Li, C. Metal-Organic Frameworks as Electrolyte Additives to Enable Ultrastable Plating/Stripping of Li Anode with Dendrite Inhibition. *ACS Appl. Mater. Interfaces* **2019**, *11* (4), 3869–3879.
- (15) Mohd Zain, N. K.; Vijayan, B. L.; Misnon, I. I.; Das, S.; Karuppiah, C.; Yang, C.-C.; Yusoff, M. M.; Jose, R. Direct Growth of Triple Cation Metal-Organic Framework on a Metal Substrate for Electrochemical Energy Storage. *Ind. Eng. Chem. Res.* **2019**, *58* (2), 665–674.
- (16) Gupta, V.; Mohiyuddin, S.; Sachdev, A.; Soni, P.; Gopinath, P.; Tyagi, S. Peg Functionalized Zirconium Dicarboxylate MOFs for Docetaxel Drug Delivery in Vitro. *J. Drug Delivery Sci. Technol.* **2019**, *52*, 846–855.
- (17) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**, *29* (7), 2618–2625.
- (18) Bobbitt, N. S.; Snurr, R. Q. Molecular Modelling and Machine Learning for High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage. *Mol. Simul.* **2019**, *45* (14–15), 1069–1081.
- (19) Daglar, H.; Keskin, S. Recent Advances, Opportunities, and Challenges in High-Throughput Computational Screening of MOFs for Gas Separations. *Coord. Chem. Rev.* **2020**, *422*, 213470.
- (20) Alpaydin, E. *Introduction to Machine Learning*; MIT press, 2020.
- (21) Bird, C. L.; Frey, J. G. Chemical Information Matters: An E-Research Perspective on Information and Data Sharing in the Chemical Sciences. *Chem. Soc. Rev.* **2013**, *42* (16), 6754–6776.
- (22) Beck, D. A.; Carothers, J. M.; Subramanian, V. R.; Pfandner, J. Data Science: Accelerating Innovation and Discovery in Chemical Engineering. *AIChE J.* **2016**, *62* (5), 1402–1416.
- (23) Erdem Günay, M.; Yildirim, R. Recent Advances in Knowledge Discovery for Heterogeneous Catalysis Using Machine Learning. *Catal. Rev.: Sci. Eng.* **2021**, *63*, 120.
- (24) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Adv. Energy Mater.* **2020**, *10* (8), 1903242.
- (25) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120* (16), 8066–8129.
- (26) Chong, S.; Lee, S.; Kim, B.; Kim, J. Applications of Machine Learning in Metal-Organic Frameworks. *Coord. Chem. Rev.* **2020**, *423*, 213487.
- (27) Shi, Z.; Yang, W.; Deng, X.; Cai, C.; Yan, Y.; Liang, H.; Liu, Z.; Qiao, Z. Machine-Learning-Assisted High-Throughput Computational Screening of High Performance Metal-Organic Frameworks. *Mol. Syst. Des. Eng.* **2020**, *5* (4), 725–742.
- (28) Sun, S.; Hartono, N. T. P.; Ren, Z. D.; Oviedo, F.; Buscemi, A. M.; Layurova, M.; Chen, D. X.; Ogunfunmi, T.; Thapa, J.; Ramasamy, S.; Settens, C.; DeCost, B. L.; Kusne, A. G.; Liu, Z.; Tian, S. I. P.; Peters, I. M.; Correa-Baena, J.-P.; Buonassisi, T. Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. *Joule* **2019**, *3* (6), 1437–1451.
- (29) Clayson, I. G.; Hewitt, D.; Hutereau, M.; Pope, T.; Slater, B. High Throughput Methods in the Synthesis, Characterization, and Optimization of Porous Materials. *Adv. Mater.* **2020**, *32* (44), 2002780.

- (30) Schmidt, J.; Shi, J.; Borlido, P.; Chen, L.; Botti, S.; Marques, M. A. L. Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chem. Mater.* **2017**, *29* (12), 5090–5103.
- (31) Yakutovich, A. V.; Eimre, K.; Schütt, O.; Talirz, L.; Adorf, C. S.; Andersen, C. W.; Dittler, E.; Du, D.; Passerone, D.; Smit, B.; et al. Aiidalab—an Ecosystem for Developing, Executing, and Sharing Scientific Workflows. *Comput. Mater. Sci.* **2021**, *188*, 110165–110174.
- (32) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G. M.; Hautier, G.; et al. Fireworks: A Dynamic Workflow System Designed for High-Throughput Applications. *Concurr. Comput.* **2015**, *27* (17), 5037–5059.
- (33) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (34) Ketkar, N. Introduction to Pytorch. In *Deep Learning with Python*; Springer, 2017; pp 195–208.
- (35) Ketkar, N. Introduction to Keras. In *Deep Learning with Python*; Springer, 2017; pp 97–111.
- (36) Himanen, L.; Jäger, M. O.; Morooka, E. V.; Canova, F. F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. Dscribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949–106961.
- (37) Khatib, M. E.; de Jong, W. A. M4chem: A Machine Learning Package for Chemistry and Materials Science. *arXiv.org* **2020**, 2003.13388.
- (38) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60–69.
- (39) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555.
- (40) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349* (6245), 255–260.
- (41) Jain, A.; Hautier, G.; Ong, S. P.; Persson, K. New Opportunities for Materials Informatics: Resources and Data Mining Techniques for Uncovering Hidden Relationships. *J. Mater. Res.* **2016**, *31* (8), 977–994.
- (42) Sumpter, B.; Noid, D. On the Use of Computational Neural Networks for the Prediction of Polymer Properties. *J. Therm. Anal. Calor.* **1996**, *46* (3–4), 833–851.
- (43) Carr, D. A.; Lach-hab, M.; Yang, S.; Vaisman, I. I.; Blaisten-Barojas, E. Machine Learning Approach for Structure-Based Zeolite Classification. *Microporous Mesoporous Mater.* **2009**, *117* (1), 339–349.
- (44) Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug Design by Machine Learning: Support Vector Machines for Pharmaceutical Data Analysis. *Comput. Chem.* **2001**, *26* (1), 5–14.
- (45) Ekins, S.; Shimada, J.; Chang, C. Application of Data Mining Approaches to Drug Delivery. *Adv. Drug Delivery Rev.* **2006**, *58* (12–13), 1409–30.
- (46) Fjell, C. D.; Jenssen, H.; Hilpert, K.; Cheung, W. A.; Panté, N.; Hancock, R. E. W.; Cherkasov, A. Identification of Novel Antibacterial Peptides by Chemoinformatics and Machine Learning. *J. Med. Chem.* **2009**, *52* (7), 2006–2015.
- (47) Landrum, G. A.; Penzotti, J. E.; Putta, S. Machine-Learning Models for Combinatorial Catalyst Discovery. *Meas. Sci. Technol.* **2005**, *16* (1), 270.
- (48) Winkler, D. A.; Burden, F. R.; Yan, B.; Weissleder, R.; Tassa, C.; Shaw, S.; Epa, V. C. Modelling and Predicting the Biological Effects of Nanomaterials. *SAR and QSAR in Environmental Research* **2014**, *25* (2), 161–172.
- (49) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. Large-Scale Quantitative Structure-Property Relationship (QsPr) Analysis of Methane Storage in Metal-Organic Frameworks. *J. Phys. Chem. C* **2013**, *117* (15), 7681–7689.
- (50) Moghadam, P. Z.; Rogge, S. M. J.; Li, A.; Chow, C. M.; Wieme, J.; Moharrami, N.; Aragoes-Anglada, M.; Conduit, G.; Gomez-Gualdrón, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* **2019**, *1* (1), 219–234.
- (51) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic Property Weighted Radial Distribution Functions Descriptors of Metal-Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **2013**, *117* (27), 14095–14105.
- (52) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-Based Descriptors to Rapidly Predict Hydrogen Storage in Metal-Organic Frameworks. *Mol. Syst. Des. Eng.* **2019**, *4* (1), 162–174.
- (53) Zhang, X. Y.; Zhang, K. X.; Lee, Y. J. Machine Learning Enabled Tailor-Made Design of Application-Specific Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* **2020**, *12* (1), 734–743.
- (54) Datar, A.; Chung, Y. G.; Lin, L. C. Beyond the Bet Analysis: The Surface Area Prediction of Nanoporous Materials Using a Machine Learning Method. *J. Phys. Chem. Lett.* **2020**, *11* (14), 5412–5417.
- (55) Gustafson, J. A.; Wilmer, C. E. Intelligent Selection of Metal-Organic Framework Arrays for Methane Sensing via Genetic Algorithms. *ACS Sens.* **2019**, *4* (6), 1586–1593.
- (56) Okur, S.; Zhang, Z. J.; Sarheed, M.; Nick, P.; Lemmer, U.; Heinke, L. Towards a MOF E-Nose: A Surmof Sensor Array for Detection and Discrimination of Plant Oil Scents and Their Mixtures. *Sens. Actuators, B* **2020**, *306*, 127502.
- (57) Shi, Z. N.; Liang, H.; Yang, W. Y.; Liu, J.; Liu, Z. L.; Qiao, Z. W. Machine Learning and in Silico Discovery of Metal-Organic Frameworks: Methanol as a Working Fluid in Adsorption-Driven Heat Pumps and Chillers. *Chem. Eng. Sci.* **2020**, *214*, 115430–115439.
- (58) Leperi, K. T.; Yancy-Caballero, D.; Snurr, R. Q.; You, F. Q. 110th Anniversary: Surrogate Models Based on Artificial Neural Networks to Simulate and Optimize Pressure Swing Adsorption Cycles for CO₂ Capture. *Ind. Eng. Chem. Res.* **2019**, *58* (2), 18241–18252.
- (59) Korolev, V. V.; Mitrofanov, A.; Marchenko, E. I.; Eremin, N. N.; Tkachenko, V.; Kalmykov, S. N. Transferable and Extensible Machine Learning-Derived Atomic Charges for Modeling Hybrid Nanoporous Materials. *Chem. Mater.* **2020**, *32* (18), 7822–7831.
- (60) Burns, T.; Pai, K. N.; Subraveti, S. G.; Collins, S.; Krykunov, M.; Rajendran, A.; Woo, T. K. Prediction of Mof Performance in Vacuum-Swing Adsorption Systems for Post-Combustion CO₂ Capture Based on Integrated Molecular Simulation, Process Optimizations, and Machine Learning Models. *Environ. Sci. Technol.* **2020**, *54* (7), 4536–4544.
- (61) Jablonka, K. M.; Moosavi, S. M.; Asgari, M.; Ireland, C.; Patiny, L.; Smit, B. A Data-Driven Perspective on the Colours of Metal-Organic Frameworks. *Chem. Sci.* **2021**, *12*, 3587–3598.
- (62) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Using Collective Knowledge to Assign Oxidation States. *ChemRxiv.org* **2020**, DOI: 10.26434/chemrxiv.11604129.v1.
- (63) Andrew, R.; Shaelyn, I.; Debmalya, R.; Zhenpeng, Y.; Alan, A.-G.; Laura, G.; Justin, N.; Randall, Q. S. Machine Learning the Quantum-Chemical Properties of Metal-Organic Frameworks for Accelerated Materials Discovery with a New Electronic Structure Database. *ChemRxiv.org* **2020**, DOI: 10.26434/chemrxiv.13147616.v1.
- (64) He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J. Metallic Metal-Organic Frameworks Predicted by the Combination of Machine Learning Methods and ab Initio Calculations. *J. Phys. Chem. Lett.* **2018**, *9* (16), 4562–4569.
- (65) Chen, P. C.; Tang, Z. Y.; Zeng, Z. M.; Hu, X. F.; Xiao, L. P.; Liu, Y.; Qian, X. D.; Deng, C. Y.; Huang, R. Y.; Zhang, J. Z.; Bi, Y. L.; Lin, R. K.; Zhou, Y.; Liao, H. G.; Zhou, D.; Wang, C.; Lin, W. B. Machine-Learning-Guided Morphology Engineering of Nanoscale Metal-Organic Frameworks. *Matter* **2020**, *2* (6), 1651–1666.

- (66) Yuan, X.; Deng, X.; Cai, C.; Shi, Z.; Liang, H.; Li, S.; Qiao, Z. Machine Learning and High-Throughput Computational Screening of Hydrophobic Metal-Organic Frameworks for Capture of Formaldehyde from Air. *Green Energy Environ.* **2020**, DOI: 10.1016/j.gee.2020.06.024.
- (67) Cho, E. H.; Deng, X.; Zou, C.; Lin, L.-C. Machine Learning-Aided Computational Study of Metal–Organic Frameworks for Sour Gas Sweetening. *J. Phys. Chem. C* **2020**, *124* (50), 27580–27591.
- (68) Kim, S.-Y.; Kim, S.-I.; Bae, Y.-S. Machine-Learning-Based Prediction of Methane Adsorption Isotherms at Varied Temperatures for Experimental Adsorbents. *J. Phys. Chem. C* **2020**, *124* (36), 19538–19547.
- (69) Halder, P.; Singh, J. K. High-Throughput Screening of Metal–Organic Frameworks for Ethane–Ethylene Separation Using the Machine Learning Technique. *Energy Fuels* **2020**, *34* (11), 14591–14597.
- (70) Ahmed, A.; Siegel, D. J. Predicting Hydrogen Storage in MOFs via Machine Learning. *SSRN* **2020**, DOI: 10.2139/ssrn.3751763.
- (71) Lee, S.; Kim, B.; Kim, J. Discovery of Record-Breaking Metal-Organic Frameworks for Methane Storage Using Evolutionary Algorithm and Machine Learning. *ChemRxiv.org* **2020**, DOI: 10.26434/chemrxiv.12333077.v1.
- (72) Deng, X. M.; Yang, W. Y.; Li, S. H.; Liang, H.; Shi, Z. N.; Qiao, Z. W. Large-Scale Screening and Machine Learning to Predict the Computation-Ready, Experimental Metal-Organic Frameworks for CO₂ Capture from Air. *Appl. Sci.* **2020**, *10* (2), 569–582.
- (73) Raza, A.; Sturluson, A.; Simon, C. M.; Fern, X. Message Passing Neural Networks for Partial Charge Assignment to Metal-Organic Frameworks. *J. Phys. Chem. C* **2020**, *124* (35), 19070–19082.
- (74) Zhou, M. S.; Vassallo, A.; Wu, J. Z. Toward the Inverse Design of Mof Membranes for Efficient D₂/H₂ Separation by Combination of Physics-Based and Data-Driven Modeling. *J. Membr. Sci.* **2020**, *598*, 117675.
- (75) Pardakhti, M.; Nanda, P.; Srivastava, R. Impact of Chemical Features on Methane Adsorption by Porous Materials at Varying Pressures. *J. Phys. Chem. C* **2020**, *124* (8), 4534–4544.
- (76) Ma, R. M.; Colon, Y. J.; Luo, T. F. Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* **2020**, *12* (30), 34041–34048.
- (77) Gharagheizi, F.; Tang, D.; Sholl, D. S. Selecting Adsorbents to Separate Diverse near-Azeotropic Chemicals. *J. Phys. Chem. C* **2020**, *124* (6), 3664–3670.
- (78) Cai, C. Z.; Li, L. F.; Deng, X. M.; Li, S. H.; Liang, H.; Qiao, Z. W. Machine Learning and High-Throughput Computational Screening of Metal-Organic Framework for Separation of Methane/Ethane/Propane. *Huaxue Xuebao* **2020**, *78* (5), 427–436.
- (79) Anderson, R.; Biog, A.; Gomez-Gualdrón, D. A. Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* **2020**, *16* (2), 1271–1283.
- (80) Web of Science. <https://apps.webofknowledge.com/> (accessed February 2021).
- (81) Zupan, J.; Gasteiger, J. Neural Networks: A New Method for Solving Chemical Problems or Just a Passing Phase? *Anal. Chim. Acta* **1991**, *248* (1), 1–30.
- (82) Burns, J. A.; Whitesides, G. M. Feed-Forward Neural Networks in Chemistry: Mathematical Systems for Classification and Pattern Recognition. *Chem. Rev.* **1993**, *93* (8), 2583–2601.
- (83) Himmelblau, D. M. Applications of Artificial Neural Networks in Chemical Engineering. *Korean J. Chem. Eng.* **2000**, *17* (4), 373–392.
- (84) Hoskins, B. F.; Robson, R. Design and Construction of a New Class of Scaffolding-Like Materials Comprising Infinite Polymeric Frameworks of 3d-Linked Molecular Rods. A Reappraisal of the Zinc Cyanide and Cadmium Cyanide Structures and the Synthesis and Structure of the Diamond-Related Frameworks [N(CH₃)₄]⁺[Cuⁱznⁱⁱ(Cn)₄]⁻ and Cuⁱ[4,4',4'',4'''-Tetracyanotetraphenylmethane]-BF₄·XC₆H₅NO₂. *J. Am. Chem. Soc.* **1990**, *112* (4), 1546–1554.
- (85) Yaghi, O. M.; Richardson, D.; Li, G.; Davis, C.; Groy, T. Open-Framework Solids with Diamond-Like Structures Prepared from Clusters and Metal-Organic Building Blocks. *Mater. Res. Soc. Symp. Proc.* **1994**, *371*, 15–15.
- (86) Janiak, C. Functional Organic Analogues of Zeolites Based on Metal-Organic Coordination Frameworks. *Angew. Chem., Int. Ed. Engl.* **1997**, *36* (13–14), 1431–1434.
- (87) Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. Metal–Organic Framework Materials as Catalysts. *Chem. Soc. Rev.* **2009**, *38* (5), 1450–1459.
- (88) Kim, D.; Kim, J.; Jung, D. H.; Lee, T. B.; Choi, S. B.; Yoon, J. H.; Kim, J.; Choi, K.; Choi, S. H. Quantitative Structure-Uptake Relationship of Metal-Organic Frameworks as Hydrogen Storage Material. *Catal. Today* **2007**, *120* (3–4), 317–323.
- (89) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal-Organic Frameworks. *Nat. Chem.* **2012**, *4* (2), 83–89.
- (90) Vishnyakov, A.; Ravikovitch, P. I.; Neimark, A. V.; Bülow, M.; Wang, Q. M. Nanopore Structure and Sorption Properties of Cu-Btc Metal-Organic Framework. *Nano Lett.* **2003**, *3* (6), 713–718.
- (91) Düren, T.; Millange, F.; Férey, G.; Walton, K. S.; Snurr, R. Q. Calculating Geometric Surface Areas as a Characterization Tool for Metal-Organic Frameworks. *J. Phys. Chem. C* **2007**, *111* (42), 15350–15356.
- (92) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. Structure–Property Relationships of Porous Materials for Carbon Dioxide Separation and Capture. *Energy Environ. Sci.* **2012**, *5* (12), 9849–9856.
- (93) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical Limits of Hydrogen Storage in Metal–Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* **2013**, *25* (16), 3373–3382.
- (94) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185–6192.
- (95) Wang, A. Y. T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. *Chem. Mater.* **2020**, *32* (12), 4954–4965.
- (96) Larose, D. T.; Larose, C. D. *Discovering Knowledge in Data: An Introduction to Data Mining*; John Wiley & Sons, 2014; Vol. 4.
- (97) Chibani, S.; Coudert, F.-X. Machine Learning Approaches for the Prediction of Materials Properties. *APL Mater.* **2020**, *8* (8), 080701.
- (98) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; MIT press, 2012.
- (99) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer, 2006.
- (100) Hinton, G. E.; Osindero, S.; Teh, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2006**, *18* (7), 1527–1554.
- (101) Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT press: Cambridge, 2016; Vol. 1.
- (102) Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2* (2), 165–193.
- (103) Ai, D.; Pan, H.; Li, X.; Gao, Y.; He, D. Association Rule Mining Algorithms on High-Dimensional Datasets. *Artif. Life Robot* **2018**, *23* (3), 420–427.
- (104) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016**, *2* (1), 1–7.
- (105) Butler, K. T.; Frost, J. M.; Skelton, J. M.; Svane, K. L.; Walsh, A. Computational Materials Design of Crystalline Solids. *Chem. Soc. Rev.* **2016**, *45* (22), 6138–6146.

- (106) Wang, H.; Ji, Y.; Li, Y. Simulation and Design of Energy Materials Accelerated by Machine Learning. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10* (1), 1–18.
- (107) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401–146405.
- (108) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87* (18), 1–19.
- (109) Jørgensen, M. S.; Mortensen, H. L.; Meldgaard, S. A.; Kolsbjerg, E. L.; Jacobsen, T. L.; Sørensen, K. H.; Hammer, B. Atomistic Structure Learning. *J. Chem. Phys.* **2019**, *151* (5), 054111.
- (110) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. *arXiv.org* **2017**, 1704.06439.
- (111) Eckhoff, M.; Behler, J. From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5. *J. Chem. Theory Comput.* **2019**, *15* (6), 3793–3809.
- (112) Sezginel, K. B.; Uzun, A.; Keskin, S. Multivariable Linear Models of Structural Parameters to Predict Methane Uptake in Metal-Organic Frameworks. *Chem. Eng. Sci.* **2015**, *124*, 125–134.
- (113) Gulsoy, Z.; Sezginel, K. B.; Uzun, A.; Keskin, S.; Yildirim, R. Analysis of CH₄ Uptake over Metal-Organic Frameworks Using Data-Mining Tools. *ACS Comb. Sci.* **2019**, *21* (4), 257–268.
- (114) O’Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* **2008**, *41* (12), 1782–1789.
- (115) Nazarian, D.; Camp, J. S.; Sholl, D. S. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal-Organic Frameworks. *Chem. Mater.* **2016**, *28* (3), 785–793.
- (116) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985–5998.
- (117) Boyd, P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18* (21), 3777–3792.
- (118) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal-Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* **2017**, *17* (11), 5801–5810.
- (119) Anderson, R.; Gómez-Gualdrón, D. A. Increasing Topological Diversity During Computational “Synthesis” of Porous Crystals: How and Why. *CrystEngComm* **2019**, *21* (10), 1653–1665.
- (120) Avci, G.; Velioglu, S.; Keskin, S. High-Throughput Screening of MOF Adsorbents and Membranes for H₂ Purification and CO₂ Capture. *ACS Appl. Mater. Interfaces* **2018**, *10* (39), 33693–33706.
- (121) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Stoddart, J. F.; You, F.; Hupp, J. T.; et al. In Silico Discovery of Metal-Organic Frameworks for Precombustion CO₂ Capture Using a Genetic Algorithm. *Sci. Adv.* **2016**, *2* (10), e1600909.
- (122) Dokur, D.; Keskin, S. Effects of Force Field Selection on the Computational Ranking of MOFs for CO₂ Separations. *Ind. Eng. Chem. Res.* **2018**, *57* (6), 2298–2309.
- (123) Gopalan, A.; Bucior, B. J.; Bobbitt, N. S.; Snurr, R. Q. Prediction of Hydrogen Adsorption in Nanoporous Materials from the Energy Distribution of Adsorption Sites. *Mol. Phys.* **2019**, *117*, 3683–3694.
- (124) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal-Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11* (1), 4068–4078.
- (125) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating Topologically Diverse Metal-Organic Frameworks for Cryo-Adsorbed Hydrogen Storage. *Energy Environ. Sci.* **2016**, *9* (10), 3279–3289.
- (126) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gomez-Gualdrón, D. A. Role of Pore Chemistry and Topology in the CO₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater.* **2018**, *30* (18), 6325–6337.
- (127) Prati, R. C.; Batista, G. E.; Monard, M. C. A Study with Class Imbalance and Random Sampling for a Decision Tree Learning System. *IFIP International Conference on Artificial Intelligence in Theory and Practice*, Springer, 2008; pp 131–140.
- (128) Senge, R.; Bösner, S.; Dembczyński, K.; Haasenritter, J.; Hirsch, O.; Donner-Banzhoff, N.; Hüllermeier, E. Reliable Classification: Learning Classifiers That Distinguish Aleatoric and Epistemic Uncertainty. *Inf. Sci.* **2014**, *255*, 16–29.
- (129) Hüllermeier, E.; Waegeman, W. Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods. *Machine Learning* **2021**, *110*, 457.
- (130) Evans, J. D.; Huang, D. M.; Haranczyk, M.; Thornton, A. W.; Sumbly, C. J.; Doonan, C. J. Computational Identification of Organic Porous Molecular Crystals. *CrystEngComm* **2016**, *18* (22), 4133–4141.
- (131) Moghadam, P. Z.; Rogge, S. M. J.; Li, A.; Chow, C. M.; Wieme, J.; Moharrami, N.; Aragonés-Anglada, M.; Conduit, G.; Gomez-Gualdrón, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. *MOF Mechanical Properties Explorer*. <http://aam.ceb.cam.ac.uk/mof-explorer/mechanicalproperties> (accessed 11 June 2020).
- (132) Raccuglia, P.; Elbert, K. C.; Adler, P. D.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76.
- (133) Moosavi, S. M.; Chidambaram, A.; Talirz, L.; Haranczyk, M.; Stylianou, K. C.; Smit, B. Capturing Chemical Intuition in Synthesis of Metal-Organic Frameworks. *Nat. Commun.* **2019**, DOI: 10.1038/s41467-019-08483-9.
- (134) Batra, R.; Chen, C.; Evans, T. G.; Walton, K. S.; Ramprasad, R. Prediction of Water Stability of Metal-Organic Frameworks Using Machine Learning. *Nat. Mach. Intell.* **2020**, *2* (11), 704–710.
- (135) Burtch, N. C.; Jasuja, H.; Walton, K. S. Water Stability and Adsorption in Metal–Organic Frameworks. *Chem. Rev.* **2014**, *114* (20), 10575–10612.
- (136) Fanourgakis, G. S.; Gkagkas, K.; Tylilanakis, E.; Froudakis, G. Fast Screening of Large Databases for Top Performing Nanomaterials Using a Self-Consistent, Machine Learning Based Approach. *J. Phys. Chem. C* **2020**, *124* (36), 19639–19648.
- (137) Zhang, X. Y.; Cui, J.; Zhang, K. X.; Wu, J. S.; Lee, Y. J. Machine Learning Prediction on Properties of Nanoporous Materials Utilizing Pore Geometry Barcodes. *J. Chem. Inf. Model.* **2019**, *59* (11), 4636–4644.
- (138) Simon, C. M.; Kim, J.; Gomez-Gualdrón, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The Materials Genome in Action: Identifying the Performance Limits for Methane Storage. *Energy Environ. Sci.* **2015**, *8* (4), 1190–1199.
- (139) Simon, C. M.; Kim, J.; Lin, L.-C.; Martin, R. L.; Haranczyk, M.; Smit, B. Optimizing Nanoporous Materials for Gas Storage. *Phys. Chem. Chem. Phys.* **2014**, *16* (12), 5499–5513.
- (140) Fanourgakis, G. S.; Gkagkas, K.; Tylilanakis, E.; Froudakis, G. A Generic Machine Learning Algorithm for the Prediction of Gas Adsorption in Nanoporous Materials. *J. Phys. Chem. C* **2020**, *124* (13), 7117–7126.
- (141) Fernandez, M.; Barnard, A. S. Geometrical Properties Can Predict CO₂ and N₂ Adsorption Performance of Metal-Organic Frameworks (MOFs) at Low Pressure. *ACS Comb. Sci.* **2016**, *18* (5), 243–252.
- (142) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29* (7), 2844–2854.

- (143) Anderson, G.; Schweitzer, B.; Anderson, R.; Gomez-Gualdrón, D. A. Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. *J. Phys. Chem. C* **2019**, *123* (1), 120–130.
- (144) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO₂ Capture. *J. Phys. Chem. Lett.* **2014**, *5* (17), 3056–3060.
- (145) Guda, A.; Guda, S.; Martini, A.; Bugaev, A.; Soldatov, M.; Soldatov, A.; Lamberti, C. Machine Learning Approaches to Xanes Spectra for Quantitative 3d Structural Determination: The Case of CO₂ Adsorption on CPO-27-Ni MOF. *Radiat. Phys. Chem.* **2020**, *175*, 108430–108435.
- (146) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High CO₂ Working Capacity and CO₂/H₂ Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123* (7), 4133–4139.
- (147) Yao, Z.; Sanchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; Aspuru-Guzik, A. Inverse Design of Nanoporous Crystalline Reticular Materials with Deep Generative Models. *Nature Machine Intelligence* **2021**, *3* (1), 76–86.
- (148) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification Schemes for Metal–Organic Frameworks to Enable Rapid Search and Cheminformatics Analysis. *Cryst. Growth Des.* **2019**, *19* (11), 6682–6697.
- (149) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials to Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27* (12), 4459–4475.
- (150) Wu, Y.; Duan, H. P.; Xi, H. X. Machine Learning-Driven Insights into Defects of Zirconium Metal–Organic Frameworks for Enhanced Ethane–Ethylene Separation. *Chem. Mater.* **2020**, *32* (7), 2986–2997.
- (151) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *npj Comput. Mater.* **2015**, *1* (1), 1–15.
- (152) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L.; Sanvito, S.; Buongiorno-Nardelli, M.; et al. Aflowlib.org: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.
- (153) Landis, D. D.; Hummelshøj, J. S.; Nestorov, S.; Greeley, J.; Dulak, M.; Bligaard, T.; Nørskov, J. K.; Jacobsen, K. W. The Computational Materials Repository. *Comput. Sci. Eng.* **2012**, *14* (6), 51–57.
- (154) Huber, S. P.; Zoupanos, S.; Uhrin, M.; Talirz, L.; Kahle, L.; Häuselmann, R.; Gresch, D.; Müller, T.; Yakutovich, A. V.; Andersen, C. W.; et al. Aiida 1.0, a Scalable Computational Infrastructure for Automated Reproducible Workflows and Data Provenance. *Sci. Data* **2020**, *7* (1), 1–18.
- (155) Mofdb. <https://mof.tech.northwestern.edu/> (accessed 11 June 2020).
- (156) Nanoporous Materials Adsorption Energy (NMAE) Database. <http://nmgc.umn.edu/nmae/index.html> (accessed 11 June 2020).
- (157) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *npj Comput. Mater.* **2018**, *4* (1), 1–8.
- (158) Feng, S.; Zhou, H.; Dong, H. Using Deep Neural Network with Small Dataset to Predict Material Defects. *Mater. Des.* **2019**, *162*, 300–310.
- (159) Elshawi, R.; Maher, M.; Sakr, S. Automated Machine Learning: State-of-the-Art and Open Challenges. *arXiv.org* **2019**, 1906.02287.
- (160) Tsamardinos, I.; Fanourgakis, G. S.; Greasidou, E.; Klontzas, E.; Gkagkas, K.; Froudakis, G. E. An Automated Machine Learning Architecture for the Accelerated Prediction of Metal–Organic Frameworks Performance in Energy and Environmental Applications. *Microporous Mesoporous Mater.* **2020**, *300*, 110160.
- (161) Altintas, C.; Avci, G.; Daglar, H.; Azar, A. N. V.; Erucar, I.; Velioglu, S.; Keskin, S. An Extensive Comparative Analysis of Two MOF Databases: High-Throughput Screening of Computation-Ready MOFs for CH₄ and H₂ Adsorption. *J. Mater. Chem. A* **2019**, *7* (16), 9593–9608.
- (162) Chen, T.; Manz, T. A. Identifying Misbonded Atoms in the 2019 CoRE Metal–Organic Framework Database. *RSC Adv.* **2020**, *10* (45), 26944–26951.
- (163) Velioglu, S.; Keskin, S. Revealing the Effect of Structure Curations on the Simulated CO₂ Separation Performances of MOFs. *Mater. Adv.* **2020**, *1* (3), 341–353.
- (164) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203.
- (165) Boyd, P. G.; Chidambaram, A.; García-Diez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gladysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; et al. Data-Driven Design of Metal–Organic Frameworks for Wet Flue Gas CO₂ Capture. *Nature* **2019**, *576* (7786), 253–256.
- (166) Hiszpanski, A. M.; Gallagher, B.; Chellappan, K.; Li, P.; Liu, S.; Kim, H.; Han, J.; Kailkhura, B.; Buttler, D. J.; Han, T. Y. Nanomaterial Synthesis Insights from Machine Learning of Scientific Articles by Extracting, Structuring, and Visualizing Knowledge. *J. Chem. Inf. Model.* **2020**, *60* (6), 2876–2887.
- (167) Wang, H.; Xie, Y. C.; Li, D. W.; Deng, H.; Zhao, Y. X.; Xin, M.; Lin, J. Rapid Identification of X-Ray Diffraction Patterns Based on Very Limited Data by Interpretable Convolutional Neural Networks. *J. Chem. Inf. Model.* **2020**, *60* (4), 2004–2011.
- (168) Chen, P.; Tang, Z.; Zeng, Z.; Hu, X.; Xiao, L.; Liu, Y.; Qian, X.; Deng, C.; Huang, R.; Zhang, J.; Bi, Y.; Lin, R.; Zhou, Y.; Liao, H.; Zhou, D.; Wang, C.; Lin, W. Machine-Learning-Guided Morphology Engineering of Nanoscale Metal–Organic Frameworks. *Matter* **2020**, *2* (6), 1651–1666.
- (169) Sorrenti, A.; Jones, L.; Sevim, S.; Cao, X.; deMello, A. J.; Marti-Gastaldo, C.; Puigmartí-Luis, J. Growing and Shaping Metal–Organic Framework Single Crystals at the Millimeter Scale. *J. Am. Chem. Soc.* **2020**, *142* (20), 9372–9381.
- (170) Hazama, H.; Murai, D.; Nagasako, N.; Hasegawa, M.; Ogihara, N. Optimization of Material Composition of Li-Intercalated Metal–Organic Framework Electrodes Using a Combination of Experiments and Machine Learning of X-Ray Diffraction Patterns. *Adv. Mater. Technol.* **2020**, *5* (9), 2000254.