

Review

Machine Learning Descriptors for CO₂ Capture Materials

Ibrahim B. Orhan ¹, Yuankai Zhao ² , Ravichandar Babarao ^{1,*}, Aaron W. Thornton ^{3,*}  and Tu C. Le ^{2,*} 

¹ School of Science, STEM College, RMIT University, G.P.O. Box 2476, Melbourne, VIC 3001, Australia

² School of Engineering, STEM College, RMIT University, G.P.O. Box 2476, Melbourne, VIC 3001, Australia; s3924784@student.rmit.edu.au

³ CSIRO Manufacturing Flagship, Clayton, Melbourne, VIC 3168, Australia

* Correspondence: ravichandar.babarao@rmit.edu.au (R.B.); aaron.thornton@csiro.au (A.W.T.); tu.le@rmit.edu.au (T.C.L.)

Abstract: The influence of machine learning (ML) on scientific domains continues to grow, and the number of publications at the intersection of ML, CO₂ capture, and material science is growing rapidly. Approaches for building ML models vary in both objectives and the methods through which materials are represented (i.e., featurised). Featurisation based on descriptors, being a crucial step in building ML models, is the focus of this review. Metal organic frameworks, ionic liquids, and other materials are discussed in this paper with a focus on the descriptors used in the representation of CO₂-capturing materials. It is shown that operating conditions must be included in ML models in which multiple temperatures and/or pressures are used. Material descriptors can be used to differentiate the CO₂ capture candidates through descriptors falling under the broad categories of charge and orbital, thermodynamic, structural, and chemical composition-based descriptors. Depending on the application, dataset, and ML model used, these descriptors carry varying degrees of importance in the predictions made. Design strategies can then be derived based on a selection of important features. Overall, this review predicts that ML will play an even greater role in future innovations in CO₂ capture.

Keywords: CO₂ capture; machine learning; featurization; descriptors



Academic Editor: Giorgio Vilardi

Received: 23 December 2024

Revised: 17 January 2025

Accepted: 17 January 2025

Published: 1 February 2025

Citation: Orhan, I.B.; Zhao, Y.; Babarao, R.; Thornton, A.W.; Le, T.C. Machine Learning Descriptors for CO₂ Capture Materials. *Molecules* **2025**, *30*, 650. <https://doi.org/10.3390/molecules30030650>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the 1800s, the effects of anthropogenic CO₂ on the climate have been discussed [1]. With current atmospheric levels surpassing 424 ppm as of November 2024, the Earth is experiencing concentrations comparable to those between 4.1 and 4.5 million years ago [2]. To combat this drastic change in CO₂ concentration, net CO₂ emissions must be reduced. A promising route for reducing atmospheric CO₂ emissions in the near term is carbon capture and storage (CCS) and carbon dioxide removal (CDR). CCS technologies have been proposed to meet emission reduction targets of up to 90% from fixed-point sources [3]. CDR aims to remove carbon dioxide directly from the atmosphere using biological approaches, such as enhanced mineralisation, biomass storage, soil carbon enhancement, and reforestation, or engineering approaches, such as direct air capture (DAC). However, many attempts to commercialise CCS and ensure its widespread adoption have failed [4]. Discovering better adsorbents that enhance the performance of CCS and reduce costs is necessary to achieve wider adoption.

In a 2023 review [5], Dziejarski et al. compiled a list of important properties when evaluating materials for CO₂ capture [6,7]. These include equilibrium adsorption capacity, CO₂ selectivity, adsorption kinetics, mechanical properties, the chemical nature of the

surface, pore characteristics, chemical and thermal stability, regeneration capacity, stability in adsorption/desorption cycles, production costs, and environmental implications. These properties were all influential on a candidate material being selected for CCS. While pre-combustion, post-combustion, and oxy-combustion offer different conditions under which CO₂ is captured, typically, post-combustion or direct air capture conditions are considered. Solid sorbents, compared with liquids, are brought to the forefront in their paper as they have garnered significant attention and are considered among the most promising materials for CCS [8].

Among the most widely used solid materials for CO₂ capture, with strengths and drawbacks in each of the aforementioned criteria, are zeolites, metal oxides, silica, metal-organic frameworks (MOFs), and carbon materials such as carbon nanotubes and activated carbon. Despite significant research having been conducted on these materials, the working adsorption capacity, cycle lifetime, and multi-cycle durability are the key properties that require improvement before these solid sorbents become economical [9]. Computational methods can focus on predicting these aspects of materials. For example, current trends indicate computational methods are frequently being applied to evaluate the selectivity and adsorption capacities of these materials.

Computational methods such as molecular dynamics (MD) and Monte Carlo (MC) simulations have been cornerstones of the materials science field for decades; a noteworthy addition to these methods in this field is ML. While generative models [10] have garnered significant attention in recent years, language-based models [11] have also seen rapid development and a surge of interest. Artificial intelligence (AI) and ML are not limited to the applications that have captured the public's imagination. The applications of ML and AI have been studied in a wide range of contexts ranging from art [12] and music [13,14] to finance [15–17], healthcare [18], and science [19,20]; their vulnerabilities [21] are yet another field of research. The number of research articles written on ML, CO₂ capture, and their intersection has steadily increased (Figure 1). This paper reviews the effects ML has had on the discovery of materials for CO₂ capture with a particular focus on the feature (also referred to as “descriptor”) categories used to describe candidate materials.

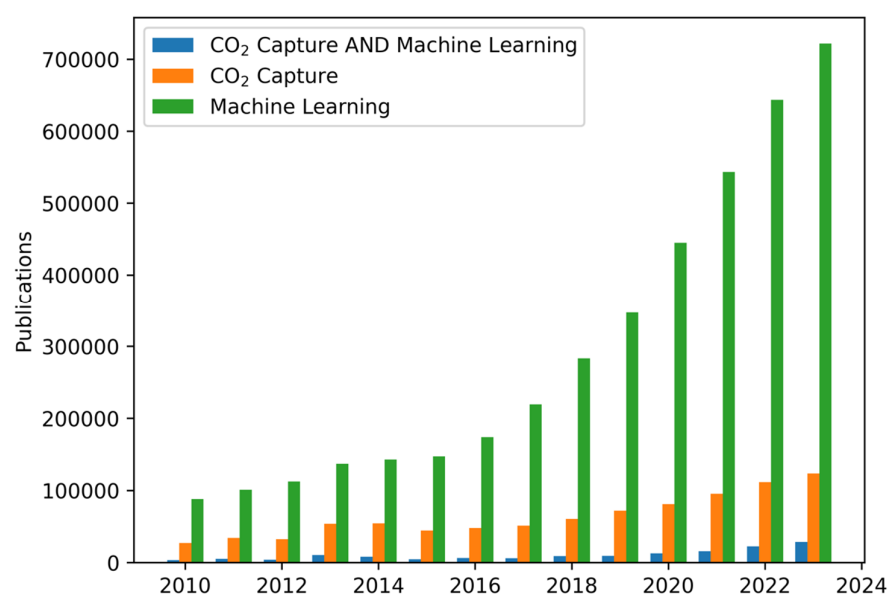


Figure 1. Number of publications on CO₂ capture, ML, and the intersection of the two fields. Data gathered from Scopus, accessed 10 November 2024. Searches were limited to the fields of chemical engineering, chemistry, and material science. The ML query was conducted using the search term “Machine Learning”; the CO₂ capture query was conducted using combinations of the following: (1) “CO₂”, “Carbon”, or “Carbon Dioxide”; and (2) “Capture”, “Storage”, or “Sequestration”.

Considered a subfield of AI, ML is a computational method that utilises data analysis to predict outcomes and plays an important role in examining “big data” [22,23]. Since its first appearance, ML has been rapidly evolving and broadening the fields to which it is applied, including material science, in which it is becoming widely used for the identification and proposal of novel materials [24,25]. ML, fundamentally, aims to use a standard set of steps and analytical methods (i.e., an algorithm) to predict the output value for any given input. Depending on how the data are presented and what procedures are used, ML algorithms can be categorised into four categories: supervised, unsupervised, semi-supervised, and reinforcement [26–28]. Within these groups of algorithms, the ML models can be further categorised into classification (for predicting categorical variables) and regression (for predicting continuous variables), clustering, and data dimensionality reduction.

In their investigation into the state of the art of ML for carbon capture, utilisation, and storage (CCUS), Yan et al. [29] formed the categories shown in Figure 2. Among some of the most commonly utilised ML algorithms are linear regressions, logistic regressions, decision trees, naive Bayes, support vector machines, random forests, artificial neural networks, *k*-means clustering, hierarchical clustering, Gaussian mixture model, AdaBoost, and principal component analysis. Most of these algorithms, falling under the umbrellas of classification, regression, or both, can be used in various contexts, including ways to predict CO₂-related material properties.

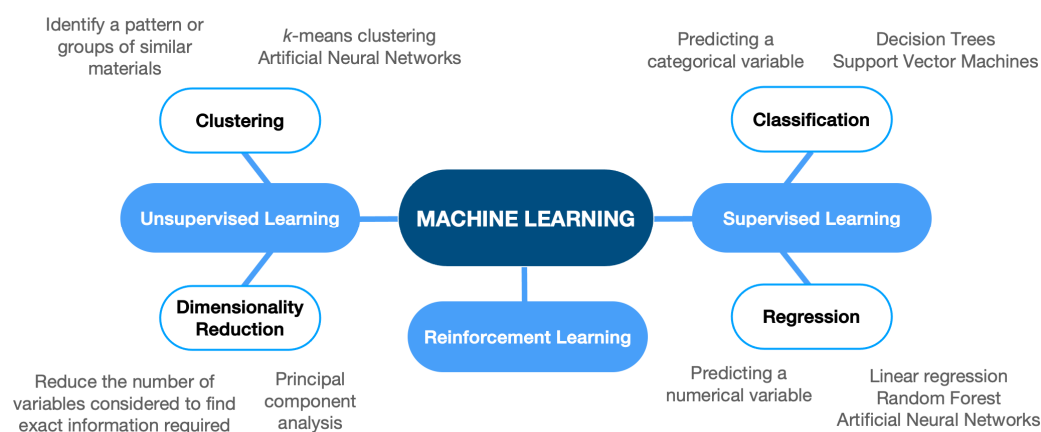


Figure 2. ML algorithms overview. ML algorithms grouped by overarching categories are shown in light blue ovals and more specific subcategories in white ovals. Examples of each subcategory’s algorithms and their function are provided in gray text above or below the relevant oval.

Yan et al. conclude that the distinct advantages of applying ML in CCS are that it provides the potential to identify links between data that are not readily identifiable, and it also provides alternative lower-computing-cost pathways. ML can be used in CCS to accelerate the design and development of materials for CCS purposes. The development of ML in CCS is expected to play a vital role in the acceleration of developing cost-effective CCS systems to tackle climate change. Being a data-driven method, it highlights the necessity for large quantities of data to develop generalised and robust ML models.

A simplified version of Yang et al.’s [30] illustration of the typical roadmap for developing and implementing an ML model is shown in Figure 3. A critical aspect of the ML process, included in the roadmap figure, is the development of descriptors. In the context of CO₂ capture, Yang et al. describe featurisation as the process in which essential properties related to CO₂ capture are transferred into numerical values that are readable by the computer. They proposed two broad subsets for these descriptors: external properties (e.g., the operating conditions and feed CO₂ concentration) and intrinsic properties (e.g.,

physical and chemical characteristics). Being a crucial step in developing ML models for CO₂ capture, this paper highlights the effectiveness of descriptor groups from both subsets.

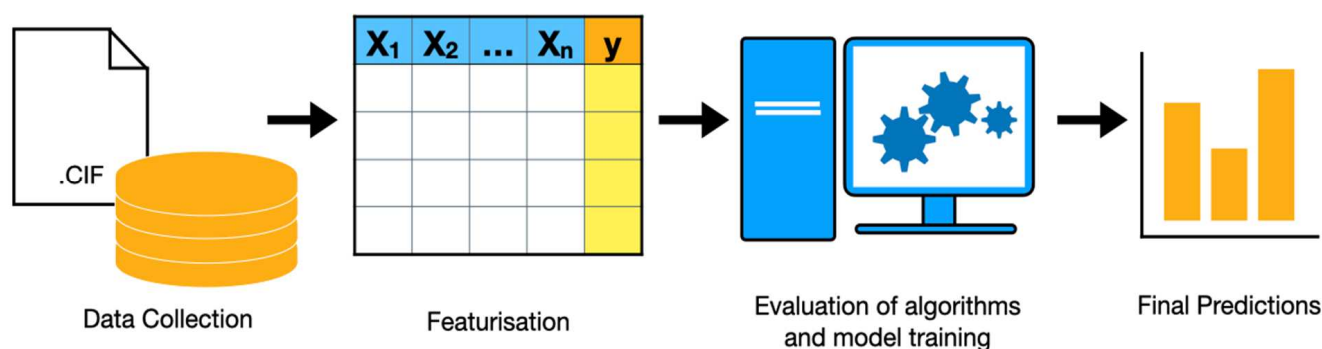


Figure 3. Illustration of the ML model building process. The process begins with the collection of data and structures, followed by gathering features for each material. The models are then trained, evaluated, and refined. Predictions are then made on unseen materials.

ML for selecting CO₂-capturing materials typically uses two approaches: (i) screening existing datasets and (ii) identifying trends and generating new structures not found in the existing datasets. While these two categories cover the main approaches to ML in CO₂ capture, there are other aspects for which ML and computational methods come into play. Large language models are assisting researchers in their data collection and experimentation [31,32], genetic algorithms and other heuristic methods are being used to explore the hypothetical material search space, and MD and MC simulations are obviating the need for physically testing materials. This review mainly focuses on publications that have utilised ML models for evaluating materials in CCS applications.

2. Descriptors in ML for CO₂ Capture

ML models can be built using experimental or simulated results; however, there is still a need to numerically or categorically describe the materials through descriptors so that an ML model can be built. Broadly, descriptors used to describe materials in CO₂ capture datasets fall under the following categories: (a) operating conditions, (b) charge and orbital, (c) thermodynamic, (d) geometric and structural, and (f) chemical composition. Depending on the type of materials, application, and ML algorithm used, certain groups may be more prevalent in constructing the ML model. The prevalence can be determined through the algorithms' internal ranking (if applicable) or through statistical methods such as Shapley values.

Considering carbon capture technologies, ML has slowly been utilised in both large-scale (industry) and small-scale (R&D and laboratory-scale) applications, including the deployment of solvent-based post-combustion capture [33], ionic liquids [34], adsorbents [35], and membranes [36]. In 2022, Hussin et al. conducted a bibliometric analysis and systematic review of ML approaches in carbon capture applications [28]. In the review, they provided a table of the most cited articles between 1999 and 2022; this was used as the starting point for evaluating the ML descriptors used in the most influential works in this review. Using the title, year, and ML model columns, the table was expanded to include a descriptors column to highlight the trends and scope of descriptors used when building ML models for CO₂ capture (see Table 1).

Table 1. Most cited articles on ML + CO₂ capture between 1999 and 2022, compiled by Hussin et al. [28]. Publications are shown in descending order based on the number of citations at the time of compiling the table. The columns highlight the various machine learning algorithms and descriptors used [37–46].

Title	Year	Type of Machine Learning	Descriptors
The use of Artificial Neural Network models for CO ₂ capture plants	2011	ANN	Temperature, Mass Flow, Mass Fraction, Solvent Lean Load, Solvent Circulation Rate, Removal Efficiency
Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO ₂ Capture	2014	Support Vector Machine (SVM)	Chemical Descriptors via Atomic Property-Weighted Radial Distribution Function (AP-RDF)
Rigorous modelling of CO ₂ equilibrium absorption in ionic liquids	2017	Least Square Support Vector Machine, Adaptive Neuro-Fuzzy Inference System, Multi-Layer Perceptron Artificial Neural Network, and Radial Basis Function Artificial Neural Network	Operating Temperature, Operating Pressure, Critical Temperature, Critical Pressure, Acentric Factor
Prediction of storage efficiency on CO ₂ sequestration in deep saline aquifers using artificial neural network	2017	ANN	Porosity, Thickness, Permeability, Depth, Time, Residual Gas Saturation
Prediction of CO ₂ loading capacities of aqueous solutions of absorbents using different computational schemes	2017	MLP-ANN, Radial Basis Function ANN, LSSVM, and ANFIS	Temperature, Concentration, Molecular Weight, Pressure
Role of Pore Chemistry and Topology in the CO ₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning	2018	Multiple Linear Regression, SVM, Decision Trees, Random Forests, Neural Networks, and Gradient Boosting Machines.	Functional Group (FG) Number Density, Void Fraction, Highest Dipole Moment of FG, Most Positive Charge, Most Negative Charge, Largest Pore Diameter, Limiting Pore Diameter, Sum of Epsilons, Gravimetric Surface Area
Prediction of CO ₂ solubility in ionic liquids using machine learning methods	2020	ANN and SVM	Temperature, Pressure, Building Groups (similar to SBUs)
Prediction of MOF Performance in Vacuum Swing Adsorption Systems for Post-combustion CO ₂ Capture Based on Integrated Molecular Simulations, Process Optimizations, and Machine Learning Models	2020	ANN and Gradient-Boosted Decision Tree Model	Adsorption Metrics (e.g., Henry's Selectivity, Heat of Adsorption, Working Capacity), Geometric Descriptors
Machine learning exploration of the critical factors for CO ₂ adsorption capacity on porous carbon materials at different pressures	2020	Random Forest	Textural Properties, Chemical Composition, Pressure
Modeling of CO ₂ capture ability of [Bmim][BF ₄] ionic liquid using connectionist smart paradigms	2021	ANN, Cascade Feed-Forward Neural Network, SVM, ANFIS	Temperature, Pressure

The studies summarised in Table 1 vary significantly in terms of the domain to which ML is applied; while some are related to CO₂ capture in materials, others look at larger, macro-scale settings and utilise considerably different descriptors than those used in the nano-scale settings of porous materials. MOFs, being assembled through distinct repeating units that can be algorithmically generated, are suitable candidates for a large range of descriptors. Clear boundaries of the repeating unit cell, known topologies, and secondary building units (SBUs) are grounds for gathering descriptors that would be inaccessible to materials with irregular, non-repeating patterns. MOFs have been studied extensively and have been used in many gas capture and storage contexts, including CO₂ capture [47].

While MOFs are dominant in the field of CO₂ capture materials, other candidates, such as ionic liquids and activated carbons, have been considered. Excluding reviews, publications with more than 10 citations (as of May 2024) that focus on the intersection of ML, CO₂ capture, and material science are listed in Table 2. While polymers, silica, alumina, and other such materials have been used for CO₂ capture, the use of ML for these materials is limited in this domain and are thus not listed in the table.

Table 2. Most cited articles for machine learning in CO₂ capture materials as of May 2024, separated by material category, providing the title of the article, ML algorithm used, descriptors used, the target variable being predicted, and the most relevant performance metrics [34,35,38,42,43,46,48–52].

Material	Title	Descriptors	Algorithm	Performance	Target
MOF and Zeolite	Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO ₂ Capture	AP-RDF	SVM	Up to 99.9% recall rate of the top 1000 MOFs at 0.15 bar and 96.8% at 1 bar	Classification of MOF CO ₂ adsorption capacity (>1 mmol/g at 0.15 bar and >4 mmol/g at 1 bar)
	Role of Pore Chemistry and Topology in the CO ₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning	Topological, geometric, charge-based	DT, RF, MLR, GBM, SVM, NN	[CO ₂ /N ₂ Selectivity] R ² = 0.905, SRCC = 0.921 [CO ₂ Loading] R ² = 0.905, SRCC = 0.950, [CO ₂ /H ₂ selectivity] R ² = 0.855, SRCC = 0.938	CO ₂ loading, CO ₂ /N ₂ selectivity, CO ₂ /H ₂ selectivity
	Prediction of MOF Performance in Vacuum Swing Adsorption Systems for Postcombustion CO ₂ Capture Based on Integrated Molecular Simulations, Process Optimizations, and Machine Learning Models	Geometric, adsorption Metrics, figures of merit (Yang's FOM, Wiersum's FOM, etc.)	Random Forest	[Productivity] Correlation R ² = 0.41 [PE] Correlation R ² = 0.18	Productivity of a material (i.e., how much CO ₂ the sorbent can extract per unit volume of the material per unit time), parasitic energy
	Robust Machine Learning Models for Predicting High CO ₂ Working Capacity and CO ₂ /H ₂ Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture	Geometric, AP-RDF	Gradient boosted trees	[CO ₂ working capacities] R ² = 0.944 [CO ₂ /H ₂ selectivities] R ² = 0.872	CO ₂ working capacities, CO ₂ /H ₂ selectivity
	A data-science approach to predict the heat capacity of nanoporous materials	Geometric, atomic, chemical	XGB	MAE = 0.02 RMAE = 2.89% SRCC = 0.98	Heat capacity (J g ^{−1} K ^{−1})

Table 2. Cont.

Material	Title	Descriptors	Algorithm	Performance	Target
MOF and Zeolite	Large-Scale Screening and Machine Learning to Predict the Computation-Ready, Experimental Metal-Organic Frameworks for CO ₂ Capture from Air	Five structural parameters: volumetric surface area (VSA), largest cavity diameter (LCD), pore-limiting diameter (PLD), porosity ϕ , density ρ , and an energy parameter: heat of adsorption	BPNN, RF, DT, SVM	Train R = 0.994, Test R = 0.981 (RF Model)	Adsorption selectivity (CO ₂ /N ₂ +O ₂)
	Design and prediction of metal organic framework-based mixed matrix membranes for CO ₂ capture via machine learning	Operating conditions, polymer type, geometric, gas adsorption metrics (selectivity, permeability, etc.)	RF	[Permeability] R ² = 0.77, RMSE = 1.45 [Selectivity] R ² = 0.7, RMSE = 0.31	CO ₂ permeability, CO ₂ /CH ₄ selectivity
	High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO ₂ Adsorption Properties of Metal Organic Frameworks	AP-RDF, chemical motif, and geometric descriptors	ANN (MLP)	[CO ₂ working capacity] Pearson r ² = 0.958, SRCC = 0.965, RMSE = 0.13 [CO ₂ /N ₂ selectivity] r ² = 0.948, SRCC = 0.975, RMSE = 10	CO ₂ /N ₂ selectivity
Ionic Liquids	Modeling of CO ₂ capture ability of [Bmim][BF ₄] ionic liquid using connectionist smart paradigms	Temperature, pressure	ANN, LS-SVM, ANFIS	AARD (%), MSE, RMSE, R ² = 7.01, 0.00115, 0.03396, 0.98408	Solubility of CO ₂ in the 1-n-butyl-3-methylimidazolium tetrafluoroborate ([Bmim][BF ₄])
	Predicting CO ₂ capture of ionic liquids using machine learning	Semi-empirical (PM6) electronic, thermodynamic, and geometrical descriptors	SVM, RF, XGB, MLP, graph-based networks	[Dataset-1] R ² , RMSE (MAE) = 0.96 0.05 (0.03) [Dataset-2] R ² , RMSE (MAE) = 0.85 0.10 (0.06)	CO ₂ solubility in 1-Butyl-3-methylimidazolium hexafluorophosphate ([Bmim][PF ₆])
	Prediction of thermo-physical properties of 1-Butyl-3-methylimidazolium hexafluorophosphate for CO ₂ capture using machine learning models	Temperature, CO ₂ partial pressure and water wt%	Gaussian process regression	R ² = 0.992; AARD% = 0.137976	CO ₂ solubility in [Bmim][PF ₆]
	Machine learning aided high-throughput prediction of ionic liquid@MOF composites for membrane-based CO ₂ capture	Structural and chemical	RF	R ² = 0.728, RMSE = 0.365, MAE = 0.277	CO ₂ /N ₂ Selectivity
	Predicting CO ₂ Absorption in Ionic Liquids with Molecular Descriptors and Explainable Graph Neural Networks	Morgan fingerprints, temperature, pressure	PLSR, CTREE, RF	MAE of 0.0137, R ² of 0.9884	CO ₂ absorption/solubility in ILs

Table 2. Cont.

Material	Title	Descriptors	Algorithm	Performance	Target
	Monitoring the effect of surface functionalization on the CO ₂ capture by graphene oxide/methyl diethanolamine nanofluids	Temperature, pressure, functionalized group, graphene oxide dosage	CFF-NN	AARD = 1.78%, MSE = 0.007, RMSE = 0.08, and R ² = 0.9906	CO ₂ solubility in graphene oxide/methyl diethanolamine
Others (Graphene, Graphite, and Activated Carbon)	Intelligent prediction models based on machine learning for CO ₂ capture performance by graphene oxide-based adsorbents	Geometric (surface area, pore volume), temperature, pressure	SVM, GBR, RF, extra trees, XGB, ANN	R ² > 0.99	CO ₂ uptake capacity
	Modeling and Optimizing N/O-Enriched Bio-Derived Adsorbents for CO ₂ Capture: Machine Learning and DFT Calculation Approaches	Physicochemical and structural features of biomass-based activated carbon	RBF-NN	R ² = 0.99, 0.974, 0.995, 0.9658, 0.9476, 0.9891 for test set predictions at (298 K and 273 K) 0.15 bar, (298 K and 273 K) 0.6 bar, and (298 K and 273 K) 1 bar, respectively	CO ₂ adsorption

Beyond the articles listed above, other publications have utilised ML through generative models built using SMILES-based descriptors [53] and evaluated them through a reward function to design new CO₂ capture materials or through ML-based analysis using more conventional descriptors, such as pore geometry- and chemistry-based features [54], to evaluate the diversity of structures in the dataset. While the specific descriptors may vary between publications, the overall information being captured through them can be grouped under broader categories. Most studies do not exclusively use a single group of descriptors; here, the descriptors used in the most-cited articles and other prevalent literature are discussed. The categories described aim to provide readers with an overview of this field in which, to the best of the authors' knowledge, a descriptor-focused review of ML for CO₂ capture has not been published.

2.1. Operating Conditions

Relevant at macro- (plant-based) and nano- (molecular-based) scales, operating conditions are key descriptors external to the adsorbent medium that determine the level of gas adsorbed. Daryayehsalameh et al., investigating 1-Butyl-3-methylimidazolium tetrafluoroborate [Bmim][BF₄] ionic liquid, used only temperature and pressure as descriptors to build an ML model. The ionic liquid and its intrinsic properties remaining constant meant temperature and pressure were the only variables in the dataset [46]. The R² displayed good agreement between the predicted values through ML and the reference values, thus proving the effectiveness of ML in predicting the CO₂ capture of the material at varying external conditions. However, the similarity of conditions under which predictions were made could have played a considerable influence on the final result and would have inflated the predictive performance of the model.

By introducing more materials and a few material-specific descriptors, Baghban et al. [39] and Song et al. [44] succeeded at predicting CO₂ capacity and solubility, respectively, in various ionic liquids and aqueous solutions. Rather than having separate temperature-specific or pressure-specific ML models, a single ML model that can make predictions on all operating conditions is beneficial. The feasibility of building such a model depends both on the algorithm's robustness and the data used to fit the model. However, these studies by Baghban et al. and Song et al., by adding descriptors relating to

the materials' unique characteristics, were able to not only make predictions on an array of materials, but they were able to do so at varying operating conditions.

Leperi et al. [55], rather than searching for new materials, instead, focused on finding the appropriate pressure-swing adsorption (PSA) conditions for a given adsorbent material, which they describe as a challenge that must be addressed to make PSA commercially competitive for carbon capture applications. Focusing on Ni-MOF-74 and zeolite 13X, the team built an artificial neural network (ANN) model that utilises variables such as pressure, column temperature, and N₂ loading as well as operating parameters such as molar feed rate, and column length to predict the column profile at the end of the feed step. They concluded that ANNs could be used as surrogate models for the rapid simulation of the individual steps of various PSA cycles. Looking at a similar macro-scale plant that utilises both adsorber and desorber columns, Sipöcz et al. [37] simulated a CO₂ removal system based on a conventional amine absorption process. From the dataset compiled through the simulations, inlet flue gas temperature, inlet flue gas mass flow, inlet flue gas CO₂ mass fraction, solvent lean load, solvent circulation rate, and removal efficiency were used as parameters to build an ANN that predicts the amount of CO₂ captured, among other outputs. Using this model, the predictions of the CO₂ captured had a maximum error that did not exceed 0.17%.

As evidenced by the studies discussed, when the adsorbent material or the macro-scale adsorption system remains constant, the role of pressure and temperature becomes critical. Unlike in ML models that search through libraries of candidate materials, in ML models built for a single material, pressure and temperature often become the only independent variables. As such, these models can be expected to have a much better performance than those looking at multiple materials. Often a simple interpolation or linear regression could yield performances sufficient for a desired application. The use of more advanced ML methods must, therefore, be scrutinized for such scenarios.

2.2. Charge- and Orbital-Based Descriptors

When more than one CO₂-capturing material candidate is considered, it is crucial to describe the intrinsic characteristics of materials to allow the ML algorithms to distinguish between candidates. Adding material-specific characteristics allows researchers, through the use of statistical methods and ML, to find trends among the characteristics that make them suitable for CCS applications. Examples of such descriptors that are particularly important at lower partial pressures are charge-based and orbital-based descriptors. One example comes from Venkatraman and Alsberg, who considered applying ML to predict CO₂ capture in ionic liquids (IL) [34]. A significant portion of their gas solubility data came from a review by Lei et al. [56], which summarised recent research produced for a number of ILs with different combinations of cations and anions. The highly tuneable nature of ILs was a leading factor in a large array of data being readily available. Data from other articles combined with the data from the review by Lei et al. resulted in the duo's dataset containing over 10,000 IL- CO₂ solubility datapoints. The study used molecular descriptors that were computed using the software KRAKENX [57]. Various quantum chemical and molecular orbital-based descriptors were calculated at the PM6 level of theory. Utilising a partial least squares regression (PLSR), conditional inference trees (CTREES), and random forests (RFs), they trained algorithms using approximately 100 descriptors. Performing two 50:50 random splits on the dataset, they aimed to accurately assess the performance of the ML algorithms. The PLSR model performed considerably worse than the other two algorithms, with the PLSR model having an R² of less than 0.5, while the RF and CTREE models had R² values above 0.8.

While the CTREE algorithm highlighted the *charge partial surface areas* (CPSAs) as the most important descriptors, one of the molecular orbital descriptors was highlighted as the most important by the RF model. The authors highlighted that, in agreement with previous studies, CO₂ adsorption is heavily influenced by the electron-donating ability, position, type, and number of substituents on the solvent. This aligns with the variables shown to be highly important as they correspond to geometrical and physicochemical properties of both cations and anions, in particular, those relevant to intermolecular interactions.

Furthermore, it was found that, according to frontier molecular orbital theory [58], the strength of the donor–acceptor interactions is typically determined by the overlap between the highest occupied molecular orbital (HOMO) on the donor and the lowest unoccupied molecular orbital (LUMO) on the acceptor. One such orbital-based descriptor was highlighted by the RF algorithm. The HOMO and LUMO energy descriptors are indicative of the cation–anion electrostatic interactions that are key to CO₂ solvation abilities [59,60]. As a more comprehensive analysis on the topic of ILs (spanning 185 ILs), this study enforces the robustness of ML used to predict CO₂ capture in another family of materials. The RF achieving an R² up to 0.96 on the test set was found to be more accurate than those obtained with the quantum-chemistry-based COSMOtherm predictions.

Charge-based and orbital-based descriptors are not uncommon for ML models built for materials such as MOFs, in which charges can be encoded to the crystallographic files that represent the structures. Along with topological and chemical descriptors, Anderson et al. [42] included the highest dipole moment of functional groups in the MOF, the most positive charge, and the most negative charge in building ML models to predict CO₂ adsorption. Other publications have created further descriptors to quantify the effects of each charge in an MOF framework through a polynomial function fitted to MC simulations using a fictitious massless atom of varying charges at different pressures [61]. Such descriptors, while improving ML models, require the structures to have the partial charges on atom sites readily available. As such, a drawback to using charge-based descriptors is the necessity of calculating the partial charges within the frameworks. While methods such as QEq are expedient, they lack the necessary precision. On the other hand, *ab initio* simulations based on density functional theory, while resulting in more accurate charge assignment on the framework, are more computationally expensive. Researchers, therefore, must weigh the costs and benefits of the trade-off between charge calculation methods before deriving any charge-based descriptors.

2.3. Thermodynamic Descriptors

Quantifying various aspects of adsorbent–adsorbate interactions, thermodynamic descriptors provide useful insight into CO₂-capturing materials. Burns et al. [43], utilising the CoRE MOF dataset [62], conducted atomistic simulations fully integrated with a detailed vacuum swing adsorption simulator and validated them at the pilot-scale to screen 1632 experimentally characterised MOFs. A total of 482 materials were found to meet the 95% CO₂ purity and 90% CO₂ recovery targets (95/90-PRTs), 365 of which had parasitic energies with a low value of 217 kWh_e/MT CO₂ below that of solvent-based capture (~290 kWh_e/MT CO₂). For post-combustion CO₂ capture, they concluded that N₂ rather than CO₂ adsorption is found to be the key metric to predict whether a material can meet the 95/90 purity–recovery requirement. Using the adsorption metrics, ML models were built to predict whether a material can meet the 95/90-PRT. These models achieved an overall prediction accuracy of up to 91%. The CO₂ and N₂ uptake capacities, working capacities, heats of adsorption, and selectivities of each gas were among the descriptors used to build the models. Other studies have also shown thermodynamic descriptors, such as the Henry coefficient of CO₂, to be important descriptors in ML predictions [61].

Thermodynamic descriptors are evidently highly influential; however, they can require significant computational resources. Therefore, similar to charge-based descriptors, weighing costs and benefits is required when considering the incorporation of thermodynamic descriptors. The appeal of ML methods is their ability to rapidly sift through large datasets and to lower the computational expense of evaluating materials. Some descriptors can only be derived through computationally expensive methods; in these cases, researchers must determine if calculating the metrics will be more beneficial than running simulations that calculate the target variable directly.

2.4. Geometric and Other Structural Descriptors

As defined by the researchers who developed ToposPro (software used for topological analysis of crystal structures), the term *topology*, when applied to a chemical structure, carries the meaning of a set of chemical bonds or other links between structural groups (atoms, molecules, and residues) [63]. Therefore, any parameter that relates to these properties and characterises them can be considered topological. Geometric descriptors conventionally relate to, among other characteristics, the surface areas, volumes, channels within the structures, and pore characteristics (Figure 4). Geometric descriptors such as these can be considered a consequence of the composition of the materials' building units (including the presence of functional groups located at nodes or linkers) and the topology; with that being the case, topological, structural, and geometric descriptors were grouped under the same header in this paper.

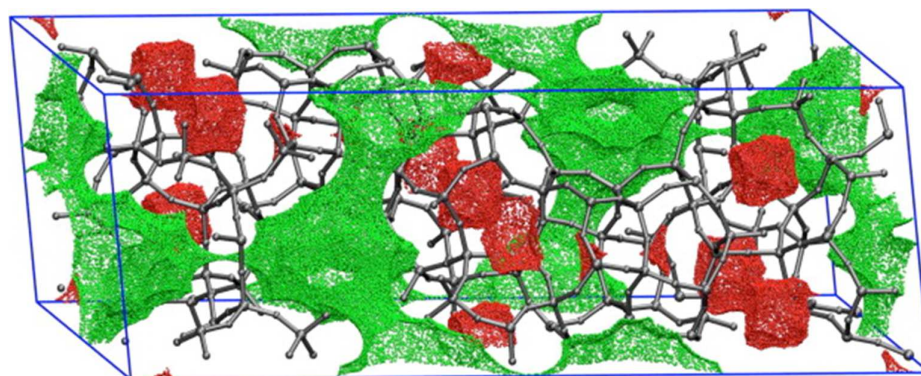


Figure 4. Sampled points on the surface of the DDR zeolite structure. Green and red points are accessible and inaccessible to a spherical probe of radius 3.2 Å, respectively [64].

MOFs' discrete repeating structures provide grounds for a wide array of topological and structural descriptors to be derived. Focusing on the MOF family of materials, Anderson et al. [42] computationally synthesised over 400 MOFs using ToBaCCo [65] by combining MOF SBUs. Calculating the adsorption loading in MOFs via grand canonical Monte Carlo (GCMC) simulations using the RASPA code [66], gathering geometrically calculated descriptors through Zeo++ [64] (accessible and inaccessible surface areas, as exemplified in Figure 5), and other descriptors such as the highest dipole bond moment led to the team compiling a dataset containing over 3000 simulation results.

Utilising the R programming language, the team implemented the following ML algorithms: multiple linear regression (MLR), support vector machine (SVM), conditional inference decision trees (DTs), random forests (RFs), neural networks (NNs), and gradient boosting machines (GBMs). Functionalisation affecting CO₂ adsorption was highlighted, focusing on the pores specifically. The authors reported that some functionalisations result in a pore chemistry that typically enhances CO₂ loadings relative to the parent MOF,

whereas some other functionalisations usually have the opposite effect. Further insight into the topological effects was provided through the use of ML.

GBM models were then used to investigate optimal chemical and topological property combinations for CO₂ capture metrics, and a genetic algorithm was used to maximise the predictions of all five metrics for each topology by the GBM models. The study found that targeting an LPD of ~nine Å and a PLD of ~four Å would yield desirable performances in the metrics the team considered. It was also found that to maximise working capacities, much higher VSA values were needed than to maximise selectivities and pure CO₂ loading. Additionally, the team noted that the selectivities and working capacities of CO₂:N₂ required higher bond dipole moments of the functional groups than for the CO₂:H₂ mixture. The use of genetic algorithms in combination with other ML algorithms, as demonstrated by the study, is an example of how design strategies can be derived through such computational methods.

Ma et al. applied ML to study the effects of pore structure, chemical properties, and adsorption conditions on CO₂ adsorption performance based on 1594 CO₂ adsorption data points gathered from the previously published literature [67]. A distinguishing feature of this paper is that the group did not focus on a single, niche family of materials but considered a broader category, termed porous carbons. These included MOFs, porous organic polymers, biomass (e.g., tobacco stem and glucose), and organic salt, among others. In their study, they selected random forest as their ML algorithm due to its proven track record in predicting heavy metal adsorption by biochar [68] and the adsorption of CO₂/CH₄ by various types of coal [36]. Using R² and RMSE as their performance metrics, the team utilised a *k*-fold cross-validation technique (70% train and 30% test) to tune the parameters of the ML model. Because the dataset consisted of adsorption data at varying temperatures and pressures, the adsorption conditions were added as descriptors alongside pore structure and chemical properties.

They found that, depending on the adsorption conditions of the test set, the model's performance ranged from an R² of 0.978 to 0.995, with RMSEs ranging from 0.057 to 0.150 mmol/g. These ML results are important for the topic of CO₂ capture in two important ways: (i) the study demonstrates how the existing literature on a collection of different materials can be effectively utilised in an ML model; and (ii) the paper provides actionable insight into trends that enable high CO₂ uptake through a feature importance analysis of the ML algorithm, experimental verification, and molecular simulations. At pressures below 0.5 bar, the study concluded that pressure was an important factor in predicting the adsorption of CO₂, while its influence waned as the pressure increased. It was also found that CO₂ adsorption density decreased with increasing pore size; this was attributed to smaller pores providing a stronger adsorption potential between the pore wall and the gas molecules [69]. Finally, they concluded that the prevalence of functional groups also plays a role in a material's adsorption capacity; the doping of oxygen and nitrogen groups can enhance CO₂ adsorption capacity.

2.5. Chemical Composition-Based Descriptors

The elements and chemical features that compose a structure are characteristics that vary widely between materials and are thus ideal for creating descriptors for ML algorithms to discern between materials. The creation of chemical descriptors has varied in both approach and context [70,71]. Pardakhti et al., while predicting methane adsorption in MOFs, created chemical descriptors that relied on the number and ratio of elements present in the adsorbent. This descriptor was later used in other contexts, including CO₂ capture [61,72,73]. Moosavi et al., on the other hand, wanted to inspect the structural building blocks of MOFs rather than an expedient but less informative approach. They analysed

the contents of the complete structure (Figure 5). Simulating the CO₂ adsorptions at 0.15 bar and 16 bar for CO₂ and utilising three databases of MOF structures, the team trained ML models based on the chemistry of the metal nodes, the chemistry of the linkers, and the functional groups as the chemical descriptors. Along with the chemical descriptors, the descriptors relating to the pore geometry were used in the training of the ML models. The study obtained a Spearman rank correlation coefficient (SRCC) of above 0.9. It was noted that for properties that are less dependent on the chemistry (e.g., high-pressure applications), the geometric descriptors are sufficient to describe the materials with the average relative error. However, for applications in which chemistry plays a role, such as predicting the Henry coefficient of CO₂, the chemical descriptors are essential to accurately predict the material properties. Similarly, predicting the maximum positive charge and minimum negative charge was possible using the chemical descriptors but not the geometric descriptors.

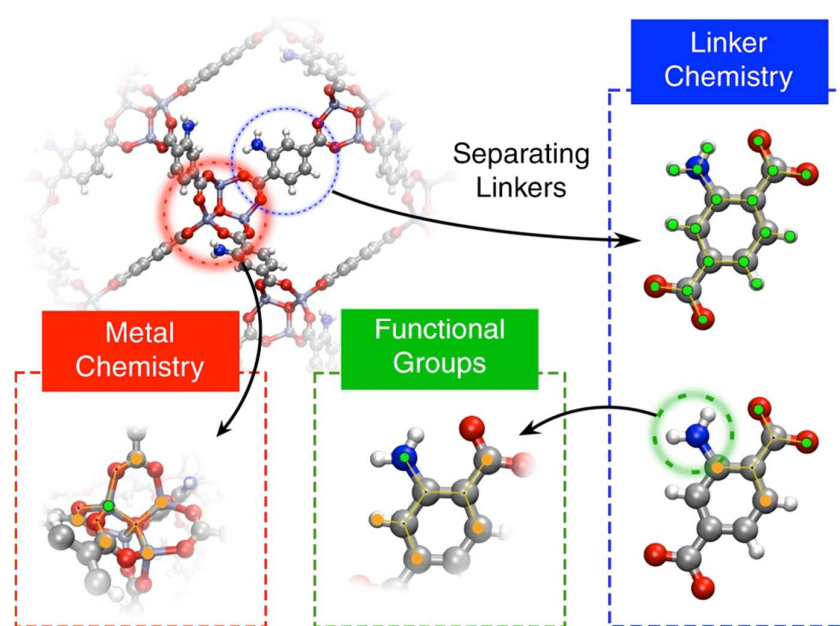


Figure 5. Description of the three domains of MOF chemistry by Moosavi et al. [74]: Metal centre revised autocorrelations (RACs) are computed on the crystal graph. Linker and functional-group RACs are computed on the corresponding linker molecular graph. Linker chemistry includes two types of RACs, namely full linker and linker connecting atoms. The graphs show the start atom (in green) and the nearby atom (in orange) used to define the RAC descriptors.

Using a similar algorithm to that introduced by Wilmer et al. [75], the authors generated approximately 1800 unfunctionalised base structures, combining 66 SBUs and 19 functional groups, resulting in a total of 324,500 hypothetical MOF structures. Utilising SVMs, the structures in the test set were predicted as having either a high or low uptake (by determining a cutoff value). Depending on the threshold used, the algorithm could discard between 67% and 95% of the dataset, greatly reducing the number of candidate materials to be simulated. The study derived the AP-RDF descriptors that would encapsulate the chemistry of periodic structures such as MOFs. It was suggested that the AP-RDF profile of an MOF framework could be interpreted as the weighted probability distribution to find an atom pair in a spherical volume of radius R inside the unit cell. In their work, the use of classifiers necessitated the employment of receiver operating characteristic (ROC) plots to evaluate their performance. The area under the curve (AUC) of the ROC plot provides a basis for comparing classifiers. The AUC values range between 0 and 1, where AUC = 1 indicates a perfect classifier. The AUCs of the ML models trained in this study were

0.979 and 0.978 for the 0.15 and 1 bar classifiers, respectively, when predicting CO₂ capture. Dureckova et al. [48] used this descriptor in their work to predict both the CO₂ capacities ($R^2 = 0.944$) and CO₂/H₂ selectivities ($R^2 = 0.872$).

Moosavi et al. utilised regression models to predict the heat capacity of materials for CO₂ capture [76]. The models were built using atomic, geometric, and chemical descriptors, and it was able to yield good agreement between the predictions and DFT simulation results (Table 2). Their simulations on ZIFs showed that changes in the topology have a minor effect on the heat capacity; thus, it was concluded that the relevant chemical environment is relatively short-ranged; this led the team to consider only the local environment surrounding the atoms in a framework, rather than the complete structure for chemical features. Achieving a good performance from a model based on these descriptors, the study showed that it was not only possible to predict the heat capacity of materials but also that the chemical featurisation of these materials could be achieved by considering only the local environment.

3. Descriptor Selection Strategies

As highlighted in the previous section, numerous descriptors can be used to represent the candidate materials; however, only a limited number are significantly relevant to the target property [77]. Feature selection algorithms enhance model performance by reducing the dimensionality of the descriptor space to prevent overfitting and decrease training time by minimising the number of input variables [70]. Feature selection algorithms can be categorised into the filter, wrapper, and embedded methods [71].

The filter method intuitively ranks features based on their relevance to the property and selects the top ones. A feature's relevance can be quantified using metrics such as the Pearson correlation coefficient, Spearman's rank correlation coefficient, and Kendall's tau. In the study by Venkatraman and Alsberg [34], using only descriptors highly correlated with the target variable, the authors built an ML model that screened suitable ILs.

Instead of selecting descriptors individually based on their relevance, the wrapper method aims to identify a subset of descriptors that collectively optimise the performance of the ML models [78]. This method is ideal for achieving a high accuracy with a specific model [79] while having the drawback of substantial computational costs. In their study [80], Dong-Hoon Oh et al. used a stepwise backward elimination algorithm to select the best descriptor set to successfully predict the CO₂ capture potential of amine-based capture processes.

The embedded method is a combination of filter and wrapper methods. Once a preliminary set of relevant features is established, these features are then used in conjunction with ML models. During the training process, the model evaluates the performance of different subsets of selected features. By incorporating regularisation techniques, such as Lasso or Ridge regression, the embedded method can further refine the descriptor set by penalising less important features, thereby promoting the selection of a more optimal subset. Through iterative training and validation, the embedded method identifies the best-performing feature subset that maximises model accuracy while minimising complexity. This method can reduce the computational cost and enhance the algorithm efficiency, becoming the main choice for feature selection algorithms [81].

4. Machine Learning Model Optimisation

4.1. Hyperparameter Tuning

The optimisation of hyperparameters in ML models is a crucial step in enhancing model performance [82]. Hyperparameters are values that configure the learning process,

and selecting an appropriate set is essential for achieving optimal results. An appropriate optimisation algorithm can significantly enhance the model's performance [77].

Manual testing is the most direct strategy among the hyperparameter tuning methods, in which researchers select and test hyperparameters based on intuition and experience. The effectiveness of hyperparameters often depends on how well they align with the chosen descriptors, as these descriptors shape the input features that the model relies on for learning. While this fast approach allows researchers to gain insights into how a model is affected by different hyperparameters, the method is unreliable in finding the best hyperparameters for a given ML model, and it relies heavily on an understanding of ML models. The relationship between hyperparameters and model performance is complex; it can therefore be time-consuming for researchers to find the hyperparameters that lead to a sufficient performance for the desired application. In the study by Xiaomei Deng et al. [83], hyperparameters such as the training function and transfer were set manually. The study, utilising the RF algorithm, successfully identified 14 MOFs with optimal CO₂ adsorption and diffusion properties. In the study by Yue Jian et al. [84], the authors manually set the hyperparameters for their graph convolutional network model, including the dimension of four convolutions, the activation algorithm, and the dropout ratio. Their model, utilising the graph neural network (GNN), achieved a mean absolute error (MAE) of 0.0137 and an R² value of 0.9884.

In contrast to manual testing, the grid search algorithm exhaustively explores a specific range of values for hyperparameters [85]. All possible hyperparameter combinations within the given range of values form a grid that can, in turn, be tested to determine the best hyperparameter combination. The model is trained and evaluated based on each combination in the grid and the best performing one is selected. A grid search can find the best hyperparameter combination within the tested combinations but requires significant time and computational resources to do so. This is, therefore, more suitable for smaller hyperparameter sets. In the publication by Jake Burner et al. [86], a grid search was conducted to optimise the dropout probability, the number of hidden layers, the number of nodes in each hidden layer, and the learning rate. The optimized model predicted for CO₂ working capacity and CO₂/N₂ selectivity in MOFs under low-pressure conditions. It achieved R² values of 0.96 for CO₂ working capacity and 0.95 for selectivity, successfully identifying 994 of the top 1000 MOFs from a test set, thereby demonstrating over a tenfold speed-up for pre-screening materials for more computationally intensive simulations.

In contrast to the systematic approach of the grid search method, a random search is a technique that avoids the computational expense of the grid search method by randomly sampling the hyperparameter search space [87]. While it can reduce computational expenses, the randomness of the algorithm cannot guarantee the best combination will be found.

4.2. Evaluation of the Performance of Models

Evaluation metrics are measurements used to quantify the performance of ML models [88]. They offer insights into how closely a model's predictions match actual outcomes, allowing the assessment of the model's effectiveness, accuracy, and reliability. These metrics help guide decisions related to model selection, tuning, and optimization [89].

4.2.1. Coefficient of Determination

R square (R²) is a statistical measure used to determine how well the regression model prediction fits the data [90]. R² values range from zero to one, and a higher value indicates a better fit of the predictions. R² represents the relative relation between the original data

and a prediction and, thus, is more informative than other metrics counting absolute errors, such as MAE and RMSE [91].

$$R^2 = 1 - \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{\sum_{i=0}^n (y_i - \bar{y})^2} \quad (1)$$

Equation (1) demonstrates the method for calculating R^2 for predictions made using a regression model. In the equation, n denotes the count of data, y_i denotes the original data, \hat{y}_i denotes the predictions, and \bar{y} denotes the average value of the original data.

4.2.2. Mean Absolute Error

The mean absolute error (MAE) calculates the average gap between the original data and predictions. As this is a measure of error, more precise predictions are closer to zero [92]. Equation (2) demonstrates the method for calculating MAE.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

where n denotes the data count, y_i denotes the original data, and \hat{y}_i denotes the predictions.

4.2.3. Root Mean Square Error

The root mean square error (RMSE) measures the square root of the average squared differences between predicted values and original values, measuring how well the model's predictions align with the observed data [93]. Similar to MAE, the closer to zero, the more accurate the prediction. However, the squaring of the difference between the true and predicted values results in outliers having a greater impact on the metric.

$$RSME = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

4.2.4. Recall Rate

The recall rate is a metric used in classification tasks to measure the ML model's ability to correctly identify instances [94]. It focuses on how well the model identifies positive cases and is particularly important when missing positive instances have significant consequences. The recall rate ranges from zero to one, with values closer to one indicating a better performance in identifying positive instances. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Here TP (true positives) is the number of correctly predicted positive instances and FN (false negatives) is the number of actual positive instances that the model incorrectly classified as negative.

4.2.5. Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient (SRCC) is a non-parametric measure of rank correlation that assesses the strength and direction of the monotonic relationship between two ranked variables. Its values range from negative one to one. A high positive SRCC (close to one) indicates a strong positive correlation, a value near zero suggests no correla-

tion, and a value close to negative one represents a strong negative correlation [95]. It is calculated as shown in Equation (5).

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (5)$$

where d is the difference between the ranks of the corresponding values of the two variables.

4.2.6. Average Absolute Relative Deviation

The average absolute relative deviation (AARD) is a metric that measures the average relative difference between predicted and actual values, indicating how distant the predicted values are to the actual values [96].

$$AARD = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y} \right| \quad (6)$$

AARD can be calculated through Equation (6). Lower AARD values indicate a better performance by the model; AARD values can be as low as zero; however, there is no upper limit to this value.

Depending on the application, researchers may opt to use one metric over another. As highlighted by Alexander et al. [90], even the commonly used metric of R^2 has its shortcomings; they encourage the reporting of root mean square error or equivalent measures of dispersion, which they argue is of more practical importance than R^2 . Opting for one evaluation metric over another or removing certain datapoints from a dataset without true justification has the possibility of falsely boosting the model's performance.

4.3. Descriptor Importance and Design Strategies

The reticular nature of MOFs and Zeolites, along with their virtually infinite number of possible configurations, allows for the production of descriptors that would otherwise be difficult. Distinct and identifiable topologies, SBUs, and tuneable functional groups within frameworks all contribute to these materials having expansive lists of descriptors. As with most adsorbents, the adsorption metrics such as permeability and heat of adsorption could be used for these materials as well; however, these descriptors are typically more expensive (computationally or materially) to obtain. MOFs, in particular, have been studied extensively for CO₂ capture purposes, and based on the descriptors used, certain design strategies have been derived, such as including specific functional groups, targeting specific pore sizes, and utilising pre-determined topologies.

Ionic liquids, on the other hand, have fewer readily available descriptor types. Daryayehsalameh et al., when investigating the solubility of CO₂ in 1-n-butyl-3-methylimidazolium tetrafluoroborate [46], utilised only the operating conditions (temperature and pressure) as the descriptors. Since the material remained constant, these were the only independent variables and thus were the only ones necessary to build the model. Due to similar conditions, the model's performance was inflated compared to those that made predictions on a wide range of candidate ILs. ML models considering more than a single ionic liquid could include thermodynamic, geometric, structural, and chemical descriptors to differentiate between the ILs. Categorical descriptors for the anions and cations, SMILES-based descriptors, and molecular fingerprints could also be included in the featurisation of these materials. For direct CO₂ capture, intermolecular cation–anion interactions were found to be important [34], while for CO₂/N₂ selectivity and permeability, accessible volume and accessible surface area were shown to have the greatest relative importance in the building of ML models [35]. Publications on the use of ML to predict

CO₂ capture properties in other materials remain limited; however, the available literature on such materials indicates the use of similar descriptors.

It is not just the materials in the dataset that influence feature importance; the context in which models are built also plays a role. Anderson et al. [42] demonstrated the varying relative importance of descriptors based on the type of ML models (Figure 6). In some instances, a specific descriptor can be an overwhelming contributor to the predictions made. For example, in some studies, it was shown that the importance of the heat of adsorption for predicting selectivity in a model utilising only six descriptors was significantly greater than all others.

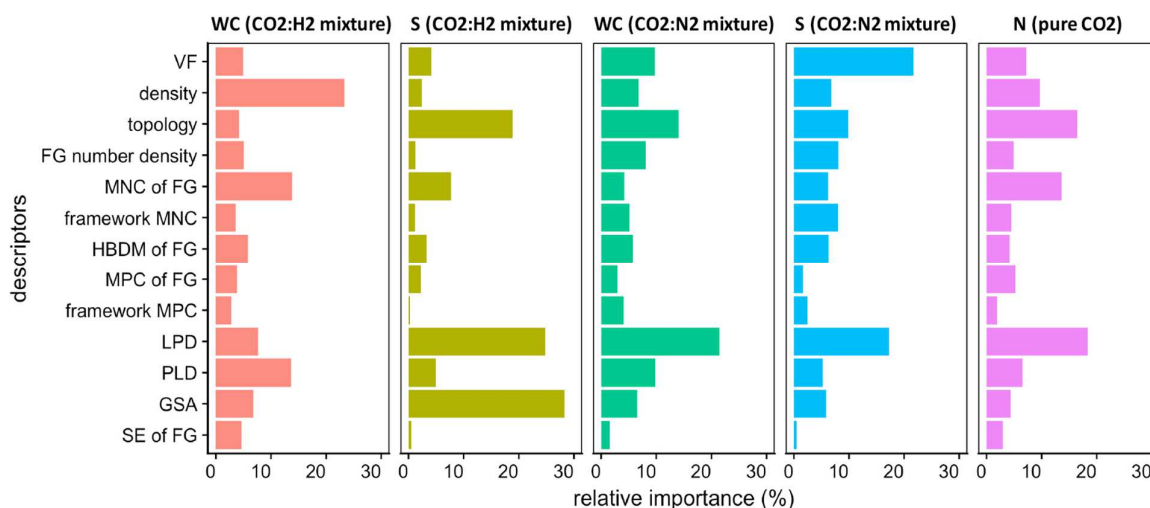


Figure 6. Influence of descriptors for predictions in various contexts [42], where the feature group S denotes selectivity, WC denotes working capacity, N denotes adsorption loading, FG denotes functional group, VF denotes void fraction, HBDM denotes highest dipole moment, MPC denotes most positive charge, MNC denotes most negative charge, LPD denotes largest pore diameter, PLD denotes limiting pore diameter, SE denotes sum of epsilons, and GSA denotes gravimetric surface area.

Moosavi et al. [74], utilising three datasets, made predictions on CO₂ uptake in MOFs. The ML algorithm, finding different trends in each dataset, gave different weights to the importance of the descriptor group (Figure 7). Taking into account how the values are distributed for the descriptor categories and their relative importance in the prediction-making process, inferences can be drawn to derive design strategies.

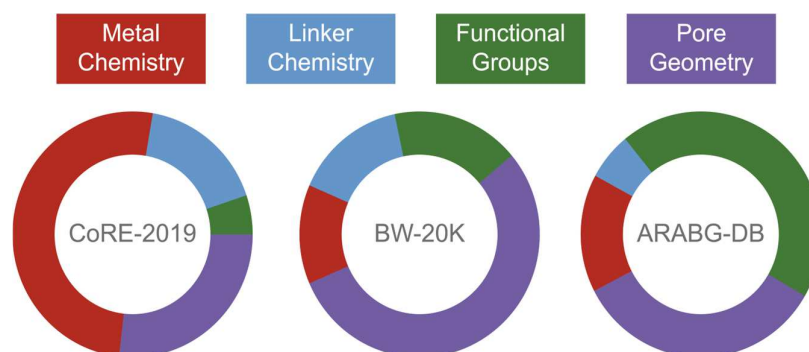


Figure 7. The relative importance of feature groups for CO₂ adsorption predictions adapted from Moosavi et al. [74], based on the work by Rahimi et al.[97]. The charts depict the relative importance of the metal chemistry (shown in red), linker chemistry (shown in blue), functional groups (shown in green), and pore geometry (shown in purple) descriptor groups for making predictions on the CoRE MOF, BW-20 K, and ARABG databases [42,62,98].

Trained using different datasets, studying different working conditions, or for different applications entirely, ML will display greater benefits from varying descriptors. In the context of CO₂ capture, external conditions were shown to be influential for both macro-scale (e.g., CCUS plants) and nano-scale (e.g., MOF) systems. Acknowledging the existence of other descriptor categories, among the most used in ML for CO₂ capture materials were the chemical, topological, thermodynamic, and charge/orbital-based descriptors.

An essential utility of trained ML models is to derive design strategies for specific applications; this can be guided by the knowledge of which descriptors are most important; however, it is not always necessary. A simple approach is to screen out optimal candidates through accurate predictions and to find common traits among the screened group. This approach is intuitive and easy to apply, but the diversity and volume of the candidate set will significantly influence the outcome. Researchers can manually select promising candidates, such as when Xiangyu Zhang et al. manually selected 10 combinations of metal nodes and topologies [53].

Genetic algorithms are another approach to deriving design strategies [99]. These algorithms generate a diverse population of candidate structures, each representing a potential solution to the optimisation problem. The candidate structures are evaluated based on a fitness function, which measures how well they meet the desired criteria. Through selection, crossover, and mutation, the population evolves over multiple generations, gradually improving the quality of the solutions. A design strategy can be derived by observing which traits are selected through the generations.

Another approach is to derive trends through an interpretable ML model. Model interpretation aims to provide transparency in decision making, revealing how input descriptors influence the outcomes of the models. This approach offers valuable insights into the underlying relationships between features and the target variable. In the context of CO₂ capture, ML model interpretation can offer a clearer understanding of how various descriptors contribute to the CO₂ capture capacity. The article by Jian Guan et al. [100], provides an example of such an application, in which their team determined that a pore size greater than one nm and a surface area of approximately 800 m²g^{−1} were necessary for their desired application.

The quality of data in the dataset used plays an important role, both in the accuracy of the model and in which descriptors are highlighted as being the most important in capturing CO₂. Inaccurate models caused by low-quality data can lead to imprecise and flawed design strategies being derived. Therefore, to gain true insight through descriptors, diligence is required in scrutinizing not only the methods used (such as ML models and evaluation metrics) but also the data quality.

5. Conclusions and Perspectives

This paper reviewed the use of ML in CO₂ capture with a specific focus on the descriptors used. ML is becoming an ever-increasingly important tool in evolving the CCS technology landscape. New approaches are frequently considered for creating descriptors relating to gas capture and storage [101].

MOFs, zeolites, and ionic liquids currently dominate the research into applications of ML on carbon-capturing materials. Some publications have demonstrated the effectiveness of using only thermodynamic conditions to make predictions on a specific material's CO₂-capture properties. In datasets involving more than one material, structural, chemical, adsorption metric-based, and charge-based descriptors have all played roles in predicting CO₂ capture to varying levels of importance. Applicable also to MOFs, SMILES-based and molecular fingerprint descriptors were among those used in the most cited ML studies focusing on ionic liquids [102].

Beyond the ability to predict which materials would be most suitable for the application on which the ML model was trained, trained models can provide insight into which descriptors play a greater role in the suitability of materials; design strategies can be derived with such insights. Researchers must pay close attention to which metrics are used to evaluate the performance of ML models; overstating the performance of an ML model is possible by favouring certain metrics over others. Therefore, it is critical to examine whether ML models would truly meet the expectations for a given application.

We speculate that the breadth of materials evaluated through ML will continue to increase, and novel approaches to featurising materials will continue to be developed to effectively represent and differentiate data points within a dataset. Materials such as eutectic solvents [103] are examples of emerging materials in CO₂ capture that can utilise ML. The overlap of material science, CO₂ capture, and ML is an area of research that will undoubtedly continue to expand. All of these factors will allow for the intersection of CO₂ capture technologies and ML to amplify. The domain of ML applications is rapidly enlarging, and as such, the role it plays in designing and evaluating CO₂ capture materials will gain further importance; through generative AI methods, research into ML for the generation of materials is already being conducted [104].

What will undoubtedly guide the adoption of materials discovered through ML and other computational methods is their feasibility in synthesis. The cost of synthesizing discovered materials and the stability of those materials will determine whether scaling the production of any candidate highlighted by ML is feasible. Affecting the economic aspect of this field of research is whether useful byproducts can be generated through the captured CO₂; formic acid is one such material that can be generated from the captured CO₂ if it can be efficiently released and used in its formation.

Author Contributions: Conceptualization, R.B. and T.C.L.; Data curation, I.B.O.; Investigation: I.B.O.; Writing—original draft preparation, I.B.O.; Writing—Reviewing and editing: Y.Z., A.W.T., R.B. and T.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Foote, E. On the Heat in the Sun's Rays. *Amer. J. Sci. Arts* **1856**, XXII, 377–382.
2. *Carbon Dioxide Now More Than 50% Higher Than Pre-Industrial Levels*; National Oceanic and Atmospheric Administration—U.S. Department of Commerce; NOAA: Silver Spring, MD, USA.
3. MacDowell, N.; Florin, N.; Buchard, A.; Hallett, J.; Galindo, A.; Jackson, G.; Adjiman, C.S.; Williams, C.K.; Shah, N.; Fennell, P. An overview of CO₂ capture technologies. *Energy Environ. Sci.* **2010**, *3*, 1645–1669. [[CrossRef](#)]
4. Abdulla, A.; Hanna, R.; Schell, K.R.; Babacan, O.; Victor, D.G. Explaining successful and failed investments in U.S. carbon capture and storage using empirical and expert assessments. *Environ. Res. Lett.* **2020**, *15*, 014036. [[CrossRef](#)]
5. Dziejarski, B.; Serafin, J.; Andersson, K.; Krzyżyńska, R. CO₂ capture materials: A review of current trends and future challenges. *Mater. Today Sustain.* **2023**, *24*, 100483. [[CrossRef](#)]
6. Yang, R.T. *Gas Separation by Adsorption Processes*; Series on Chemical Engineering; World Scientific Publishing CO. 364; Imperial College Press: London, UK, 1997; Volume 1.
7. Rouquerol, F.; Rouquerol, J.; Sing, K.S.W.; Llewellyn, P.; Maurin, G. *Adsorption by Powders and Porous Solids*, 2nd ed.; Oxford Academic Press: Oxford, UK, 2014.
8. Figueroa, J.D.; Fout, T.; Plasynski, S.; McIlvried, H.; Srivastava, R.D. Advances in CO₂ capture technology—The U.S. Department of Energy's Carbon Sequestration Program. *Int. J. Greenh. Gas Control* **2008**, *2*, 9–20. [[CrossRef](#)]
9. Lee, S.-Y.; Park, S.-J. A review on solid adsorbents for carbon dioxide capture. *J. Ind. Eng. Chem.* **2015**, *23*, 1–11. [[CrossRef](#)]
10. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.

11. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
12. Falomir, Z.; Museros, L.; Sanz, I.; Gonzalez-Abril, L. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-Learn). *Expert Syst. Appl.* **2018**, *97*, 83–94. [[CrossRef](#)]
13. Bahuleyan, H. Music Genre Classification using Machine Learning Techniques. *arXiv* **2018**, arXiv:1804.01149.
14. Sturm, B.L.; Ben-Tal, O.; Monaghan, Ú.; Collins, N.; Herremans, D.; Chew, E.; Hadjeres, G.; Deruty, E.; Pachet, F. Machine learning research that matters for music creation: A case study. *J. New Music Res.* **2019**, *48*, 36–55. [[CrossRef](#)]
15. Karachun, I.; Vinnichuk, L.; Tuskov, A. Machine learning methods in finance. *SHS Web Conf.* **2021**, *110*, 05012. [[CrossRef](#)]
16. Warin, T.; Stojkov, A. Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature. *J. Risk Financ. Manag.* **2021**, *14*, 302. [[CrossRef](#)]
17. Sahu, S.K.; Mokhade, A.; Bokde, N.D. An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. *Appl. Sci.* **2023**, *13*, 1956. [[CrossRef](#)]
18. Shailaja, K.; Seetharamulu, B.; Jabbar, M.A. Machine Learning in Healthcare: A Review. In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 910–914.
19. Tarca, A.L.; Carey, V.J.; Chen, X.-w.; Romero, R.; Drăghici, S. Machine learning and its applications to biology. *PLoS Comput. Biol.* **2007**, *3*, e116. [[CrossRef](#)]
20. Gao, C.; Min, X.; Fang, M.; Tao, T.; Zheng, X.; Liu, Y.; Wu, X.; Huang, Z. Innovative Materials Science via Machine Learning. *Adv. Funct. Mater.* **2022**, *32*, 2108044. [[CrossRef](#)]
21. Newen, C.; Müller, E. On the Independence of Adversarial Transferability to Topological Changes in the Dataset. In Proceedings of the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece, 9–13 October 2023; pp. 1–8.
22. Thai, H.-T. Machine learning for structural engineering: A state-of-the-art review. *Structures* **2022**, *38*, 448–491. [[CrossRef](#)]
23. Liu, Y.; Esan, O.C.; Pan, Z.; An, L. Machine learning for advanced energy materials. *Energy AI* **2021**, *3*, 100049. [[CrossRef](#)]
24. Lu, W.; Xiao, R.; Yang, J.; Li, H.; Zhang, W. Data mining-aided materials discovery and optimization. *J. Mater.* **2017**, *3*, 191–201. [[CrossRef](#)]
25. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: Recent applications and prospects. *NPJ Comput. Mater.* **2017**, *3*, 54. [[CrossRef](#)]
26. Khan, A.I.; Al-Badi, A. Open Source Machine Learning Frameworks for Industrial Internet of Things. *Procedia Comput. Sci.* **2020**, *170*, 571–577. [[CrossRef](#)]
27. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)] [[PubMed](#)]
28. Hussin, F.; Md Rahim, S.A.N.; Hatta, N.S.M.; Aroua, M.K.; Mazari, S.A. A systematic review of machine learning approaches in carbon capture applications. *J. CO₂ Util.* **2023**, *71*, 102474. [[CrossRef](#)]
29. Yan, Y.; Borhani, T.N.; Subraveti, S.G.; Pai, K.N.; Prasad, V.; Rajendran, A.; Nkulikiyinka, P.; Asibor, J.O.; Zhang, Z.; Shao, D.; et al. Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS)—A state-of-the-art review. *Energy Environ. Sci.* **2021**, *14*, 6122–6157. [[CrossRef](#)]
30. Yang, Z.; Chen, B.; Chen, H.; Li, H. A critical review on machine-learning-assisted screening and design of effective sorbents for carbon dioxide (CO₂) capture. *Front. Energy Res.* **2023**, *10*, 1043064. [[CrossRef](#)]
31. Naseem, U.; Dunn, A.G.; Khushi, M.; Kim, J. Benchmarking for biomedical natural language processing tasks with a domain specific ALBERT. *BMC Bioinform.* **2022**, *23*, 144. [[CrossRef](#)]
32. Boiko, D.A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570–578. [[CrossRef](#)]
33. Shalaby, A.; Elkamel, A.; Douglas, P.L.; Zhu, Q.; Zheng, Q.P. A machine learning approach for modeling and optimization of a CO₂ post-combustion capture unit. *Energy* **2021**, *215*, 119113. [[CrossRef](#)]
34. Venkatraman, V.; Alsberg, B.K. Predicting CO₂ capture of ionic liquids using machine learning. *J. CO₂ Util.* **2017**, *21*, 162–168. [[CrossRef](#)]
35. Zhang, Z.; Cao, X.; Geng, C.; Sun, Y.; He, Y.; Qiao, Z.; Zhong, C. Machine learning aided high-throughput prediction of ionic liquid@MOF composites for membrane-based CO₂ capture. *J. Membr. Sci.* **2022**, *650*, 120399. [[CrossRef](#)]
36. Meng, M.; Qiu, Z.; Zhong, R.; Liu, Z.; Liu, Y.; Chen, P. Adsorption characteristics of supercritical CO₂/CH₄ on different types of coal and a machine learning approach. *Chem. Eng. J.* **2019**, *368*, 847–864. [[CrossRef](#)]
37. Sipöcz, N.; Tobiesen, F.A.; Assadi, M. The use of Artificial Neural Network models for CO₂ capture plants. *Appl. Energy* **2011**, *88*, 2368. [[CrossRef](#)]
38. 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*, 3056. [[CrossRef](#)] [[PubMed](#)]

39. Baghban, A.; Bahadori, A.; Mohammadi, A.H.; Behbahani, A. Prediction of CO₂ loading capacities of aqueous solutions of absorbents using different computational schemes. *Int. J. Greenh. Gas Control* **2017**, *57*, 143–161. [\[CrossRef\]](#)
40. Baghban, A.; Mohammadi, A.H.; Taleghani, M.S. Rigorous modeling of CO₂ equilibrium absorption in ionic liquids. *Int. J. Greenh. Gas Control* **2017**, *58*, 19–41. [\[CrossRef\]](#)
41. Kim, Y.; Jang, H.; Kim, J.; Lee, J. Prediction of storage efficiency on CO₂ sequestration in deep saline aquifers using artificial neural network. *Appl. Energy* **2017**, *185*, 916–928. [\[CrossRef\]](#)
42. Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-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*, 6325–6337. [\[CrossRef\]](#)
43. Burns, T.D.; Pai, K.N.; Subraveti, S.G.; Collins, S.P.; Krykunov, M.; Rajendran, A.; Woo, T.K. Prediction of MOF Performance in Vacuum Swing Adsorption Systems for Postcombustion CO₂ Capture Based on Integrated Molecular Simulations, Process Optimizations, and Machine Learning Models. *Environ. Sci. Technol.* **2020**, *54*, 4536–4544. [\[CrossRef\]](#)
44. Song, Z.; Shi, H.; Zhang, X.; Zhou, T. Prediction of CO₂ solubility in ionic liquids using machine learning methods. *Chem. Eng. Sci.* **2020**, *223*, 115752. [\[CrossRef\]](#)
45. Zhu, X.; Tsang, D.C.W.; Wang, L.; Su, Z.; Hou, D.; Li, L.; Shang, J. Machine learning exploration of the critical factors for CO₂ adsorption capacity on porous carbon materials at different pressures. *J. Clean. Prod.* **2020**, *273*, 122915. [\[CrossRef\]](#)
46. Daryayehsalameh, B.; Nabavi, M.; Vaferi, B. Modeling of CO₂ capture ability of [Bmim][BF₄] ionic liquid using connectionist smart paradigms. *Environ. Technol. Innov.* **2021**, *22*, 101484. [\[CrossRef\]](#)
47. Demir, H.; Aksu, G.O.; Gulbalkan, H.C.; Keskin, S. MOF Membranes for CO₂ Capture: Past, Present and Future. *Carbon Capture Sci. Technol.* **2022**, *2*, 100026. [\[CrossRef\]](#)
48. 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*, 4133. [\[CrossRef\]](#)
49. Rahimi, M.; Abbaspour-Fard, M.H.; Rohani, A.; Yuksel Orhan, O.; Li, X. Modeling and Optimizing N/O-Enriched Bio-Derived Adsorbents for CO₂ Capture: Machine Learning and DFT Calculation Approaches. *Ind. Eng. Chem. Res.* **2022**, *61*, 10670–10688. [\[CrossRef\]](#)
50. Mazari, S.A.; Siyal, A.R.; Solangi, N.H.; Ahmed, S.; Griffin, G.; Abro, R.; Mubarak, N.M.; Ahmed, M.; Sabzoi, N. Prediction of thermo-physical properties of 1-Butyl-3-methylimidazolium hexafluorophosphate for CO₂ capture using machine learning models. *J. Mol. Liq.* **2021**, *327*, 114785. [\[CrossRef\]](#)
51. Zhou, Z.; Davoudi, E.; Vaferi, B. Monitoring the effect of surface functionalization on the CO₂ capture by graphene oxide/methyl diethanolamine nanofluids. *J. Environ. Chem. Eng.* **2021**, *9*, 106202. [\[CrossRef\]](#)
52. Fathalian, F.; Aarabi, S.; Ghaemi, A.; Hemmati, A. Intelligent prediction models based on machine learning for CO₂ capture performance by graphene oxide-based adsorbents. *Sci. Rep.* **2022**, *12*, 21507. [\[CrossRef\]](#)
53. Zhang, X.; Zhang, K.; Lee, Y. Machine Learning Enabled Tailor-Made Design of Application-Specific Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* **2020**, *12*, 734–743. [\[CrossRef\]](#)
54. Majumdar, S.; Moosavi, S.M.; Jablonka, K.M.; Ongari, D.; Smit, B. Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening. *ACS Appl. Mater. Interfaces* **2021**, *13*, 61004–61014. [\[CrossRef\]](#)
55. Leperi, K.T.; Yancy-Caballero, D.; Snurr, R.Q.; You, F. 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*, 18241–18252. [\[CrossRef\]](#)
56. Lei, Z.; Dai, C.; Chen, B. Gas solubility in ionic liquids. *Chem. Rev.* **2014**, *114*, 1289–1326. [\[CrossRef\]](#)
57. Venkatraman, V.; Alsberg, B.K. KRAKENX: Software for the generation of alignment-independent 3D descriptors. *J. Mol. Model.* **2016**, *22*, 93. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Fukui, K. Theory of Orientation and Stereoselection. In *Orientation and Stereoselection*; Springer: Berlin/Heidelberg, Germany, 1970; pp. 1–85.
59. Carvalho, P.J.; Kurnia, K.A.; Coutinho, J.A.P. Dispelling some myths about the CO₂ solubility in ionic liquids. *Phys. Chem. Chem. Phys.* **2016**, *18*, 14757–14771. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Klähn, M.; Seduraman, A. What Determines CO₂ Solubility in Ionic Liquids? A Molecular Simulation Study. *J. Phys. Chem. B* **2015**, *119*, 10066–10078. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Orhan, I.B.; Le, T.C.; Babarao, R.; Thornton, A.W. Accelerating the prediction of CO₂ capture at low partial pressures in metal-organic frameworks using new machine learning descriptors. *Commun. Chem.* **2023**, *6*, 214. [\[CrossRef\]](#)
62. Chung, Y.G.; Haldoupis, E.; Bucior, B.J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K.D.; Milisavljevic, M.; Ling, S.; Camp, J.S.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985. [\[CrossRef\]](#)
63. Blatov, V.A.; Shevchenko, A.P.; Proserpio, D.M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Cryst. Growth Des.* **2014**, *14*, 3576. [\[CrossRef\]](#)

64. Willems, T.F.; Rycroft, C.H.; Kazi, M.; Meza, J.C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141. [\[CrossRef\]](#)
65. 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*, 5801. [\[CrossRef\]](#)
66. Dubbeldam, D.; Calero, S.; Ellis, D.E.; Snurr, R.Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42*, 81. [\[CrossRef\]](#)
67. Ma, X.; Xu, W.; Su, R.; Shao, L.; Zeng, Z.; Li, L.; Wang, H. Insights into CO₂ capture in porous carbons from machine learning, experiments and molecular simulation. *Sep. Purif. Technol.* **2023**, *306*, 122521. [\[CrossRef\]](#)
68. Zhu, X.; Wang, X.; Ok, Y.S. The application of machine learning methods for prediction of metal sorption onto biochars. *J. Hazard. Mater.* **2019**, *378*, 120727. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Ma, X.; Chen, R.; Zhou, K.; Wu, Q.; Li, H.; Zeng, Z.; Li, L. Activated Porous Carbon with an Ultrahigh Surface Area Derived from Waste Biomass for Acetone Adsorption, CO₂ Capture, and Light Hydrocarbon Separation. *ACS Sustain. Chem. Eng.* **2020**, *8*, 11721–11728. [\[CrossRef\]](#)
70. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 94. [\[CrossRef\]](#)
71. Jović, A.; Brkić, K.; Bogunović, N. A review of feature selection methods with applications. In Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 25–29 May 2015; pp. 1200–1205.
72. Orhan, I.B.; Daglar, H.; Keskin, S.; Le, T.C.; Babarao, R. Prediction of O₂/N₂ Selectivity in Metal–Organic Frameworks via High-Throughput Computational Screening and Machine Learning. *ACS Appl. Mater. Interfaces* **2022**, *14*, 736–749. [\[CrossRef\]](#)
73. Pardakhti, M.; Moharrer, E.; Wanik, D.; Suib, S.L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19*, 640–645. [\[CrossRef\]](#)
74. 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*, 4068. [\[CrossRef\]](#)
75. 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*, 83. [\[CrossRef\]](#)
76. Moosavi, S.M.; Novotny, B.Á.; Ongari, D.; Moubarak, E.; Asgari, M.; Kadioglu, Ö.; Charalambous, C.; Ortega-Guerrero, A.; Farmahini, A.H.; Sarkisov, L.; et al. A data-science approach to predict the heat capacity of nanoporous materials. *Nat. Mater.* **2022**, *21*, 1419–1425. [\[CrossRef\]](#)
77. Kira, K.; Rendell, L.A. A Practical Approach to Feature Selection. In *Machine Learning Proceedings 1992*; Sleeman, D., Edwards, P., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1992; pp. 249–256.
78. Venkatesh, B.; Anuradha, J. A Review of Feature Selection and Its Methods. *Cybern. Inf. Technol.* **2019**, *19*, 3. [\[CrossRef\]](#)
79. Colaco, S.; Kumar, S.; Tamang, A.; Biju, V.G. A Review on Feature Selection Algorithms. In Proceedings of the Emerging Research in Computing, Information, Communication and Applications, Singapore, 11 September 2019; pp. 133–153.
80. Oh, D.-H.; Dat Vo, N.; Lee, J.-C.; You, J.K.; Lee, D.; Lee, C.-H. Prediction of CO₂ capture capability of 0.5 MW MEA demo plant using three different deep learning pipelines. *Fuel* **2022**, *315*, 123229. [\[CrossRef\]](#)
81. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 1060–1073. [\[CrossRef\]](#)
82. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [\[CrossRef\]](#)
83. Deng, X.; Yang, W.; Li, S.; Liang, H.; Shi, Z.; Qiao, Z. Large-Scale Screening and Machine Learning to Predict the Computation-Ready, Experimental Metal–Organic Frameworks for CO₂ Capture from Air. *Appl. Sci.* **2020**, *10*, 569. [\[CrossRef\]](#)
84. Jian, Y.; Wang, Y.; Barati Farimani, A. Predicting CO₂ Absorption in Ionic Liquids with Molecular Descriptors and Explainable Graph Neural Networks. *ACS Sustain. Chem. Eng.* **2022**, *10*, 16681–16691. [\[CrossRef\]](#)
85. Liashchynskiy, P.; Liashchynskiy, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.
86. Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P.G.; Woo, T.K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO₂ Adsorption Properties of Metal–Organic Frameworks. *J. Phys. Chem. C* **2020**, *124*, 27996–28005. [\[CrossRef\]](#)
87. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
88. Handelman, G.S.; Kok, H.K.; Chandra, R.V.; Razavi, A.H.; Huang, S.; Brooks, M.; Lee, M.J.; Asadi, H. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *Am. J. Roentgenol.* **2019**, *212*, 38–43. [\[CrossRef\]](#)
89. Shcherbakov, M.; Brebels, A.; Shcherbakova, N.L.; Tyukov, A.; Janovsky, T.A.; Kamaev, V.A. A survey of forecast error measures. *World Appl. Sci. J.* **2013**, *24*, 171–176. [\[CrossRef\]](#)

90. Alexander, D.L.J.; Tropsha, A.; Winkler, D.A. Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322. [[CrossRef](#)]
91. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [[CrossRef](#)]
92. Cort, J.W.; Kenji, M. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82.
93. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [[CrossRef](#)]
94. Powers, D. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach. Learn. Technol.* **2008**, *2*. [[CrossRef](#)]
95. Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
96. Kader, G.D. Means and MADs. *Math. Teach. Middle Sch.* **1999**, *4*, 398–403. [[CrossRef](#)]
97. Rahimi, M.; Moosavi, S.M.; Smit, B.; Hatton, T.A. Toward smart carbon capture with machine learning. *Cell Rep. Phys. Sci.* **2021**, *2*, 100396. [[CrossRef](#)]
98. Boyd, P.G.; Chidambaram, A.; García-Díez, E.; Ireland, C.P.; Daff, T.D.; Bounds, R.; Gładysiak, 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*, 253–256. [[CrossRef](#)]
99. Collins, S.P.; Daff, T.D.; Piotrkowski, S.S.; Woo, T.K. Materials Design by Evolutionary Optimization of Functional Groups in Metal–Organic Frameworks. *Sci. Adv.* **2016**, *2*, e1600954. [[CrossRef](#)]
100. Guan, J.; Huang, T.; Liu, W.; Feng, F.; Japip, S.; Li, J.; Wu, J.; Wang, X.; Zhang, S. Design and prediction of metal organic framework-based mixed matrix membranes for CO₂ capture via machine learning. *Cell Rep. Phys. Sci.* **2022**, *3*, 100864. [[CrossRef](#)]
101. Lu, C.; Wan, X.; Ma, X.; Guan, X.; Zhu, A. Deep-Learning-Based End-to-End Predictions of CO₂ Capture in Metal–Organic Frameworks. *J. Chem. Inf. Model.* **2022**, *62*, 3281–3290. [[CrossRef](#)] [[PubMed](#)]
102. Zhang, K.; Wu, J.; Yoo, H.; Lee, Y. Machine Learning-based approach for Tailor-Made design of ionic Liquids: Application to CO₂ capture. *Sep. Purif. Technol.* **2021**, *275*, 119117. [[CrossRef](#)]
103. Makarov, D.M.; Krestyaninov, M.A.; Dyshin, A.A.; Golubev, V.A.; Kolker, A.M. CO₂ capture using choline chloride-based eutectic solvents. An experimental and theoretical investigation. *J. Mol. Liq.* **2024**, *413*, 125910. [[CrossRef](#)]
104. Choudhary, K.; Yildirim, T.; Siderius, D.W.; Kusne, A.G.; McDannald, A.; Ortiz-Montalvo, D.L. Graph neural network predictions of metal organic framework CO₂ adsorption properties. *Comput. Mater. Sci.* **2022**, *210*, 111388. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.