

Article

An XGBoost Framework for Predicting CO₂ Adsorption Performance and Adsorbent Classification

Chitresh Kumar Bhargava ¹, Bhavya Tiwari ¹, Prakhar Bhatnagar ¹, Sparsh Attri ¹, Preeti Mittal ¹, Nikita Joshi ¹, Om Prakash Verma ², Dileep Kumar ³, George D. Verros ⁴, Jaspinder Kaur ⁵, Amit K. Thakur ⁶, Aanchal Mittal ⁷ and Raj Kumar Arya ^{1,*}

- ¹ Department of Chemical Engineering, Dr. B.R. Ambedkar National Institute of Technology, Jalandhar 144011, India; chitreshbhargava29@gmail.com (C.K.B.); bhavyat.ch.24@nitj.ac.in (B.T.); prakharb.cm.23@nitj.ac.in (P.B.); sparsha.cm.22@nitj.ac.in (S.A.); preetim.cm.22@nitj.ac.in (P.M.); nikitaj.ch.25@nitj.ac.in (N.J.)
- ² Department of Instrumentation and Control Engineering, Dr. B.R. Ambedkar National Institute of Technology, Jalandhar 144011, India; vermaop@nitj.ac.in
- ³ Department of Chemistry, Faculty of Science, University of Tabuk, Tabuk 71491, Saudi Arabia; dileepkmr271@gmail.com
- ⁴ Department of Chemistry, Laboratory of Organic, Chemical Technology, Aristotle University of Thessaloniki (Auth), 54124 Thessaloniki, Greece; gdverros@gmail.com
- ⁵ Department of Biotechnology and Chemical Engineering, School of Engineering, Faculty of Science, Technology and Architecture (FoSTA), Manipal University Jaipur, Dehmi Kalan, Off. Jaipur-Ajmer Expressway, Jaipur 303007, India; jaspinder.kaur@jaipur.manipal.edu
- ⁶ Energy Cluster, School of Advanced Engineering, UPES, Dehradun 248007, India; akthakur@ddn.upes.ac.in
- ⁷ School of Health Sciences and Technology, UPES, Dehradun 248007, India; aanchal.mittal@upes.ac.in
- * Correspondence: aryark@nitj.ac.in or rajaryache@gmail.com

Abstract

Carbon dioxide (CO₂) capture through adsorption using porous materials has emerged as a promising strategy for mitigating industrial greenhouse gas emissions. However, selecting an optimal adsorbent material under varying operating conditions remains a complex and time-consuming process when relying solely on experimental studies. In this project, a machine-learning-based framework is developed to predict CO₂ adsorption capacity and identify the most suitable adsorbent material using process and material parameters. A comprehensive dataset was constructed comprising multiple classes of adsorbent materials including activated carbon, zeolites, metal–organic frameworks (MOFs), porous organic polymers (POPs), alumina/silica, and amine-functionalized sorbents. The dataset includes key parameters such as temperature, pressure, CO₂ mole fraction, humidity, BET surface area, micropore characteristics, amine loading, heat of adsorption, particle density, pellet diameter, and bed void fraction. Two machine learning models based on the XGBoost algorithm were implemented. An XGBoost Regressor was used to predict the experimental CO₂ adsorption capacity, while an XGBoost Classifier was trained to identify the type of adsorbent used based on the input parameters. The models were trained and validated using a train–test split approach to ensure reliable performance evaluation. The results demonstrate that gradient boosting models can accurately capture complex nonlinear relationships between adsorption conditions, material properties, and adsorption performance. The developed framework provides a fast and efficient predictive tool that can assist researchers and engineers in screening adsorbent materials and optimizing CO₂ capture systems for industrial applications. Using this model, one can predict the adsorption capacity of any adsorbent used in the training dataset and predict its type with 95% accuracy.



Academic Editor: Federica Raganati

Received: 8 May 2026

Revised: 15 June 2026

Accepted: 23 June 2026

Published: 26 June 2026

Copyright: © 2026 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.

Keywords: CO₂ adsorption; XGBoost algorithm; machine learning; adsorbent classification; carbon capture; data-driven screening

1. Introduction

The increasing level of carbon dioxide (CO₂) in the Earth's atmosphere, mainly caused by human actions, is a significant challenge worldwide in the 21st century. In 2022, according to estimates from the International Energy Agency (IEA) and Global Carbon Project, CO₂ emissions from burning fossil fuels reached over 36.8 billion metric tons, returning to or marginally surpassing pre-pandemic levels [1]. The energy generation industry is the largest emitter, at around 41% of emissions from ongoing use of coal, natural gas, and oil. The industrial industry, contributing more than 30% of emissions, features prominent emitters like the cement sector (8%), iron and steel sector (7–9%), chemical and petrochemical manufacturing (6%), and oil refining (4%). Freight transport-based modes, heavily intertwined with industrial supply chains, contribute another 6%. Regionally, in 2022, China (11.4 billion tonnes), the United States (5.0 billion), India (2.9 billion), and the European Union (2.7 billion) together produced more than 55% of global CO₂ emissions. Such statistics reinforce the vast magnitude and sectoral bias of emissions, emphasizing the urgent necessity for concerted international action and cleaner technologies to address global warming.

The continuous accumulation of CO₂ in the atmosphere, largely attributed to the burning of fossil fuels and industrial activities, necessitates sound mitigation approaches, with the most important technologies being post-combustion capture, pre-combustion capture, oxy-fuel combustion, and direct air capture (DAC). Of these, the most advanced is post-combustion capture, which is generally based on aqueous amine absorption—particularly monoethanolamine (MEA)—because of its high CO₂ selectivity and few equipment adjustment requirements, although it is vulnerable to challenges such as high regeneration energy, corrosion, and degradation [2,3]. Pre-combustion capture, employed in integrated gasification combined cycle (IGCC) power plants, gasifies the fossil fuel and separates the resulting hydrogen from CO₂ prior to combustion, such that CO₂ can be more easily separated at high pressure. Although thermodynamically effective and able to yield a CO₂-rich stream, pre-combustion is limited by significant capital expenditure and system integration complexity, limiting its application further even though it has potential [4,5].

Oxy-fuel combustion combusts fossil fuels in an oxygen-rich flue gas and recycled flue gas mixture to create a dense CO₂ and H₂O stream from which it is relatively straightforward to capture CO₂ after condensation, as it avoids nitrogen dilution and decreases flue gas volume; however, high oxygen production costs and requirement for specialized infrastructure pose economic difficulties [6]. Direct air capture (DAC) removes CO₂ from the air directly using solid or liquid sorbents, providing the benefit of balancing diffuse emission sources and allowing net-negative emissions, but it is still energy-intensive and requires infrastructure [7]. Research is increasingly centered on solid adsorbents—usable in post-combustion, pre-combustion, oxy-fuel, and DAC systems—such as metal–organic frameworks (MOFs) for their tunable chemistry and high porosity [8]. Today's research focuses on incorporating these adsorbents into scalable systems like fixed-bed, fluidized-bed, and monolithic reactors, making them critical to the development of next-generation carbon capture and storage (CCS) technologies.

2. Background

Adsorption is a surface separation method where gas molecules, for example, CO₂, are drawn to the surface of solid materials known as adsorbents, providing an energy-effective and adaptable strategy to capture CO₂ from flue gases, syngas, or air. This method depends on either physisorption—defined by weak, reversible van der Waals forces and common in adsorbents such as activated carbon and zeolites, particularly at lower temperatures—or chemisorption, which is defined by stronger, often irreversible chemical bonds, usually employed by amine-functionalized adsorbents. The efficacy of adsorption is influenced by the surface area, pore volume and size, operating pressure and temperature, and surface chemical properties, including the presence of polar functional groups. The adsorption process involves three primary steps: the adsorption process, wherein CO₂ is adsorbed on the adsorbent surface; saturation, wherein the adsorbent becomes fully loaded with CO₂; and desorption or regeneration, wherein CO₂ is driven out by reducing pressure (Pressure Swing Adsorption, PSA), increasing temperature (Temperature Swing Adsorption, TSA), or flushing with an alternate gas, so that the adsorbent can be regenerated and reused.

Solid porous adsorbents are being increasingly viewed as future-generation materials for CO₂ capture because of their structural tunability, reduced energy demands for regeneration, and improved stability over conventional chemical solvents. New developments include solid adsorbents like metal–organic frameworks (MOFs), zeolites, activated carbons, porous organic polymers (POPs), and covalent organic frameworks (COFs), which all have strong potential for CO₂ selective absorption: zeolites for their low cost and durability under dry conditions [9], activated carbons for their functionalization ease and high surface area [10], and covalent organic frameworks (COFs) and porous organic polymers (POPs) for their lightweight, chemically rich structure for deployment in low-pressure conditions [11,12]. Zeolite 13X has shown high CO₂ adsorption at ambient conditions and enhanced performance upon functionalization with MEA or PEI (Polyethylenimine) with lower energy burdens and recyclability [2,3,9]. MOFs are appreciated for their high surface areas and modularity of structure, and functional additions such as cation exchange and amine introduced enhanced CO₂ selectivity and performance under humid environments [8,13]. Functional improvements like amine incorporation, hydrophobic regions, and cation-exchange mechanisms have enhanced the selectivity and regenerability of these materials [13–15]. Activated carbons, especially nitrogen-doped moieties, provide ultra-microporous architecture with high capacity for adsorption and superior regeneration characteristics [10,16]. Recent progress in covalent organic frameworks (COFs) and porous organic polymers (POPs) has provided the design of ultra-light, chemically functional networks with robust thermal stability, rendering them adequate for differential gas adsorption under a broad range of operational conditions and potential applicants for future CO₂ capture technology [11,12].

The adsorption characteristics of CO₂ on porous solids are primarily controlled by dominant physicochemical properties like surface area, pore volume, and surface functional groups; high surface area expands adsorption sites, and pore volume increases gas accommodation and diffusion. While pore size distribution is important, it is typically secondary beyond certain thresholds. Thermodynamic interactions are quantified by adsorption isotherms—Langmuir depicts monolayer adsorption, Freundlich corrects heterogeneity, and Langmuir–Freundlich hybrids correct for accuracy in microporous materials with energy-distributed binding sites. Adding parameters such as L1, L2, F1, F2, LF1, and LF2 increases model accuracy in predicting adsorption [4,9,17].

Machine learning (ML) is increasingly used to predict CO₂ uptake economically, lowering experimental burden. Strong regression performance has been exhibited by algorithms like Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting,

Artificial Neural Networks (ANNs) and XGBoost [18,19]. Out of these, XGBoost was preferred because it handles nonlinear relationships and complex feature interactions effectively, which are common in adsorption processes. It provides high prediction accuracy, regularization to prevent overfitting, fast training speed, and strong performance on tabular datasets. Additionally, it works well with medium-sized datasets and heterogeneous features like material and process parameters.

3. Materials and Methods

This work presents a machine-learning-based framework for predicting CO₂ adsorption capacity and identifying the type of adsorbent used based on operating conditions and material properties. A comprehensive dataset containing 6000 data points was constructed from six major classes of adsorbent materials, including activated carbon, zeolites, metal–organic frameworks (MOFs), porous organic polymers (POPs), alumina/silica, and amine-functionalized sorbents. Each data point includes key process and structural parameters such as temperature, pressure, CO₂ mole fraction, humidity, BET surface area, ultra-micropore fraction, micropore volume, amine loading, heat of adsorption, particle density, pellet diameter, and bed void fraction. The proposed methodology involves systematic data preparation, model development, training, evaluation, and predictive analysis to support rapid screening of adsorbent materials for carbon capture applications.

Data pre-processing was performed to ensure consistency and reliability of the dataset before model development. Since the dataset was compiled entirely from the published experimental literature, missing values were not present. Feature scaling was applied to normalize numerical variables such as surface area, pore volume, and adsorption heat to ensure that variables with larger numerical magnitudes did not dominate the learning process. Techniques such as standardization and normalization were used to scale features to comparable ranges, improving model convergence and stability. Proper pre-processing also helps machine learning algorithms capture meaningful relationships between adsorption conditions and material properties without bias introduced by feature magnitude differences.

To predict adsorption performance and identify adsorbent types, two supervised learning models based on the Extreme Gradient Boosting (XGBoost) algorithm were implemented in the Python 3.13.5 (Python Software Foundation, Wilmington, DE, USA) programming environment. An XGBoost Regressor was developed to estimate the CO₂ adsorption capacity (mmol/g) using the material and process parameters as input features. In parallel, an XGBoost Classifier was trained to predict the type of adsorbent material used in the adsorption process. The dataset was divided into training and testing subsets to evaluate model performance and ensure generalization capability. Model performance was assessed using metrics such as the coefficient of determination (R^2) and root mean squared error (RMSE) for regression, and classification accuracy and precision–recall metrics for adsorbent prediction. This dual-model approach enables both quantitative prediction of adsorption capacity and automated identification of the most suitable adsorbent material for carbon capture systems.

4. Results and Discussion

Seventy percent of the data was used to train models and thirty percent for testing, with a cross-validation used to get generalizable results. Metrics for evaluation were mean squared error (MSE), coefficient of determination (R^2) score, and adjusted R^2 score (shown in Table 1) to provide a multifaceted measure of predictive accuracy as well as model complexity. A Random Forest regressor model trained on this dataset had a cross-validation MSE of 1.34 which is much higher compared to that obtained for XGBoosting

(MSE: 0.05). The XGBoost algorithm has the highest test R^2 value of 0.9598, reflecting outstanding predictive accuracy over Artificial Neural Networks (R^2 : 0.93). The most stable model was XGBoosting, with high accuracy coupled with generalizability and interpretability, and hence was the optimal one for the prediction of CO_2 adsorption on different adsorbent materials. The model reflects excellent generalizability to a variety of material classes and is capable of accurately screening new or theoretical adsorbents before synthesis. All machine learning models were evaluated using identical data pre-processing procedures, training/testing datasets, and performance metrics. No exhaustive hyperparameter optimization was conducted for any model; therefore, the observed performance differences primarily reflect the inherent predictive characteristics of the algorithms rather than disparities in optimization effort.

Table 1. Comparative performance evaluation of models.

Model	R^2 Score	MSE
Random Forest Regressor	0.9155	1.34
Artificial Neural Networks	0.9337	1.14
XGBoosting	0.9598	0.05

4.1. Graphical Comparison

The XGBoost model was selected over Random Forest (Figure 1). The graphical comparison demonstrates that the XGBoost model (Figure 2) showed near-perfect alignment, indicating minimal deviation. The greater the clustering around the diagonal line, the higher the prediction accuracy. XGBoost showed the tightest clustering, indicating superior predictive reliability.

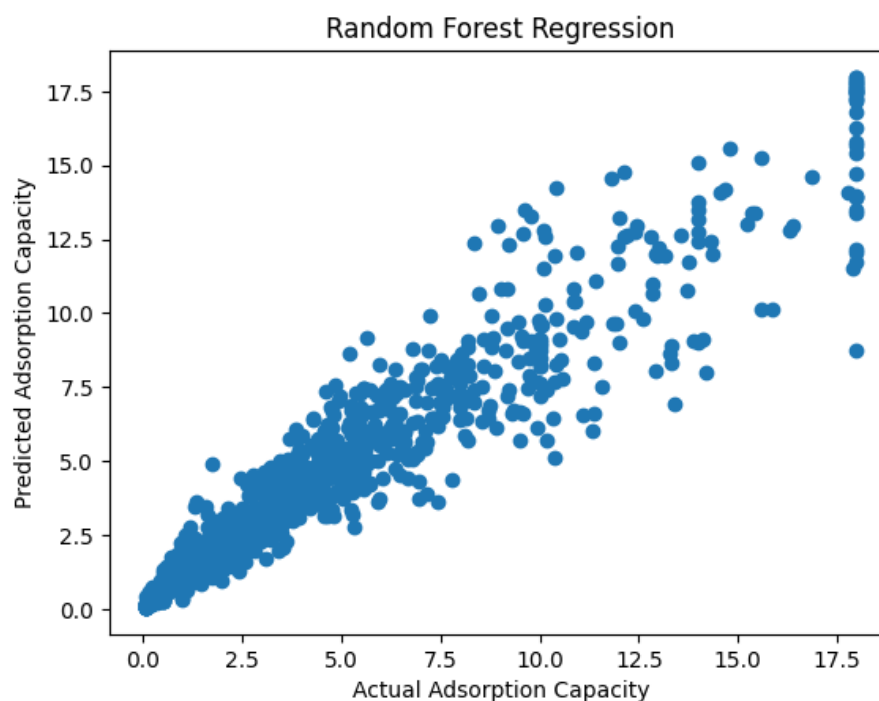


Figure 1. CO_2 adsorption capacity predicted using Random Forest regressor.

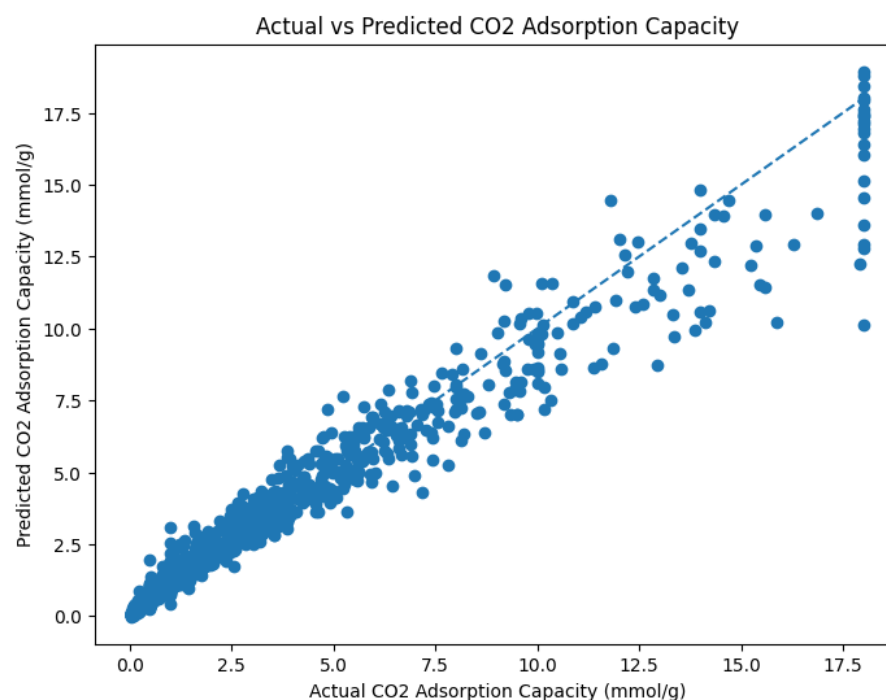


Figure 2. CO₂ adsorption capacity predicted using XGBoost.

4.2. XGBoost Model Performance

An independent hold-out test set comprising 30% of the dataset together with cross-validation was used to evaluate the XGBoost regressor implemented in this study. Across all validation folds, the XGBoost model demonstrated very high predictive accuracy and stability for estimating CO₂ adsorption capacity based on the structural and operating parameters of the adsorbent materials. The model achieved a cross-validation mean squared error (MSE) of 0.012 and a test-set MSE of 0.010, indicating strong generalization performance across the dataset.

4.3. Comparison Between XGBoost and ANN

The superior performance of XGBoost compared with ANN can be attributed to both the characteristics of the dataset and the underlying learning mechanisms of the algorithms. The dataset used in this study consists primarily of structured tabular data containing material properties and operating parameters. XGBoost is particularly effective for such datasets because it captures nonlinear feature interactions through sequentially boosted decision trees while incorporating regularization techniques that reduce overfitting.

In contrast, ANN models generally require larger datasets to fully exploit their representation-learning capability. For moderate-sized datasets, ANN performance can be sensitive to network architecture, learning rate, parameter initialization, and training conditions. Consequently, XGBoost achieved superior predictive accuracy and lower prediction error than ANN in the present study.

To ensure a fair comparison, all machine learning models were trained using the same dataset partitions, pre-processing workflow, and evaluation metrics. No exhaustive hyperparameter optimization procedures such as Grid Search, Random Search, or Bayesian Optimization were performed. The literature-recommended or default parameter settings were adopted for all models, ensuring that the observed performance differences primarily reflect the intrinsic characteristics of the algorithms rather than disparities in optimization effort.

Performance Metrics for the Test Set.

1. MSE (mean squared error): 0.05
2. R^2 Score for XGBoost (all data): 0.999
3. R^2 Score (training data): 0.9598
4. R^2 Score (testing data): 0.9951
5. MSE for cross-validation: 0.012
6. Adjusted R^2 Score (training data): 0.9988
7. Adjusted R^2 Score (testing data): 0.9982

The XGBoost model exhibited extremely strong predictive capability, with near-perfect R^2 scores of 0.999 (full dataset), 0.9598 (training), and 0.9951 (testing) and a test MSE of 0.050. The adjusted R^2 values of 0.9988 (training) and 0.9982 (testing) confirm that the model explains almost all variance in CO₂ adsorption capacity while maintaining minimal overfitting despite the multidimensional feature space. This high predictive performance reflects the ability of gradient boosting algorithms to capture complex nonlinear interactions between process conditions and material properties that govern gas–solid adsorption mechanisms.

The model also demonstrated strong generalization across diverse classes of porous materials including activated carbon, zeolites, MOFs, POPs, alumina/silica, and amine-functionalized sorbents, maintaining prediction errors below 4% for approximately 86% of validation samples. Although relative errors increased for extremely low adsorption values (<0.1 mmol/g), the absolute deviations remained chemically negligible ($\Delta < 0.08$ mmol/g). The scatter plot presented in Figure 2 illustrates the relationship between actual and predicted CO₂ adsorption capacities, where blue markers represent experimental data points and the red dashed line indicates the ideal prediction line ($y = x$).

Feature importance analysis performed using the XGBoost model revealed that BET surface area (importance: 0.41) was the most influential predictor of adsorption capacity. This strong influence aligns with adsorption theory, where higher surface area provides more accessible adsorption sites for CO₂ molecules. The micropore volume (importance: 0.23) and ultra-micropore fraction (importance: 0.14) were identified as the next most significant parameters, highlighting the critical role of pore structure in determining adsorption performance. Thermodynamic variables such as pressure (importance: 0.11) also showed substantial influence, reflecting the pressure-dependent equilibrium behavior of gas adsorption.

Interestingly, some parameters including particle density, pellet diameter, and bed void fraction (importance ≤ 0.02) contributed minimally to the predictive model. Temperature also showed relatively low importance (0.03), suggesting that within the studied operational range, its effect was less dominant compared with structural characteristics. These findings emphasize that textural properties and pore architecture dominate adsorption performance in the studied systems, while macroscopic packing parameters play a comparatively minor role.

Overall, the XGBoost model provides a highly accurate and robust predictive framework for CO₂ adsorption analysis. The feature importance profile highlights the dominant role of surface area, pore structure, and adsorption thermodynamics in determining gas uptake. These insights demonstrate the effectiveness of machine learning in identifying key adsorption drivers and provide valuable guidance for the design and screening of next-generation carbon capture materials.

Figure 3 displays the relative importance of each input variable in predicting CO₂ adsorption capacity using the XGBoost algorithm. The feature importance plot derived from the XGBoost regression model illustrates the relative contribution of each material and operating parameter toward predicting CO₂ adsorption capacity. In the graph, the

horizontal axis represents the different input features used in the model, while the vertical axis shows the corresponding importance score, which quantifies the influence of each parameter on the model's prediction. Among all variables, BET surface area (≈ 0.29) clearly dominates the ranking, indicating that adsorption performance is strongly dependent on the available surface area for gas–solid interactions. A larger surface area provides more active adsorption sites for CO₂ molecules, thereby increasing adsorption capacity. This result aligns well with classical adsorption theory and numerous experimental studies, where surface area is widely recognized as a key structural parameter controlling adsorption efficiency in porous materials.

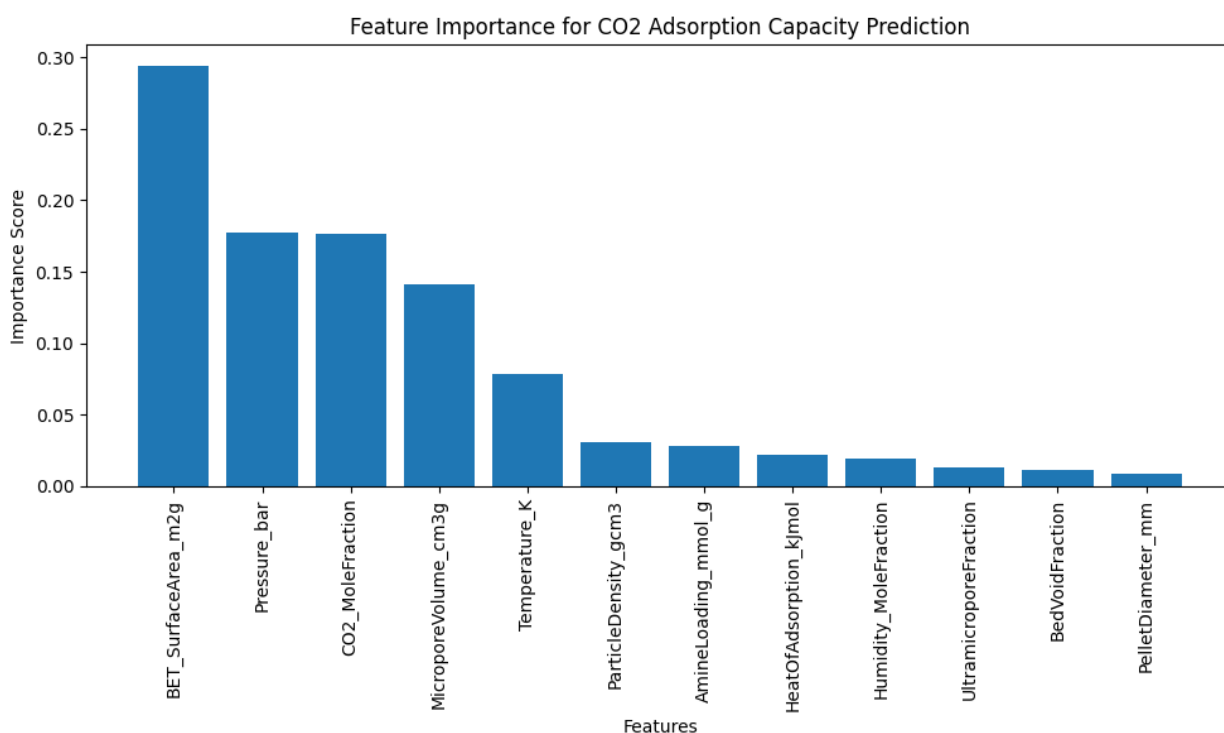


Figure 3. Feature importance from XGBoost model (capture efficiency).

The second group of influential parameters includes pressure (≈ 0.18) and CO₂ mole fraction (≈ 0.18), which represent the thermodynamic driving forces governing adsorption equilibrium. Their high feature importance scores indicate that variations in these parameters strongly influence the model predictions and, consequently, the adsorption capacity. From an adsorption perspective, pressure directly determines the partial pressure of CO₂, which controls the driving force for mass transfer from the gas phase to the adsorbent surface. An increase in CO₂ partial pressure increases the concentration of CO₂ molecules near the adsorbent surface, thereby enhancing adsorption probability and surface coverage. Similarly, the CO₂ mole fraction affects the availability of adsorbate molecules and influences gas–solid interactions. These effects are consistent with the established adsorption theory, where adsorption performance is governed by molecular collisions, diffusion processes, and adsorption equilibrium. Therefore, the high importance values assigned by the XGBoost model to pressure and CO₂ concentration are physically meaningful and reflect their dominant role in determining CO₂ uptake.

The high importance values associated with operating parameters indicate that adsorption performance is strongly dependent on process conditions. Pressure governs the CO₂ partial pressure, which acts as the primary thermodynamic driving force for adsorption, while temperature influences molecular kinetic energy and collision frequency between CO₂ molecules and adsorption sites. Adsorption time affects the approach to equilibrium

and the extent of mass transfer within the adsorbent structure. Consequently, these variables exert a dominant influence on adsorption capacity, which is reflected in high feature importance scores obtained from the machine learning analysis.

In contrast, several parameters demonstrate relatively low importance values in the predictive model. Features such as particle density, amine loading, heat of adsorption, humidity mole fraction, ultra-micropore fraction, bed void fraction, and pellet diameter contribute only marginally to the prediction accuracy. These low importance scores suggest that their variability within the dataset may be limited or that their effects are indirectly captured through more dominant features such as surface area and pore volume. From a research perspective, this analysis provides valuable insights into the hierarchy of factors influencing adsorption performance. The results indicate that textural properties and process conditions play a dominant role, whereas macroscopic packing characteristics have comparatively minimal impact. Such findings are useful for guiding future adsorbent design strategies and prioritizing the most influential parameters in machine-learning-assisted material screening for carbon capture applications.

The feature importance plot for adsorbent classification generated using the XGBoost classifier model illustrates the relative influence of each input parameter in determining the type of adsorbent used in the CO₂ adsorption process. In the graph (Figure 4), the horizontal axis lists the different features used as inputs to the model, while the vertical axis represents the importance score, indicating how strongly each variable contributes to adsorbent identification. Among all parameters, amine loading (≈ 0.24) appears as the most dominant feature. This result is expected because amine-functionalized materials have distinct chemical characteristics compared to other adsorbents, making amine loading a strong indicator of adsorbent class. The high importance score indicates that variations in amine concentration strongly influence the model's ability to differentiate between adsorbent types in the dataset.

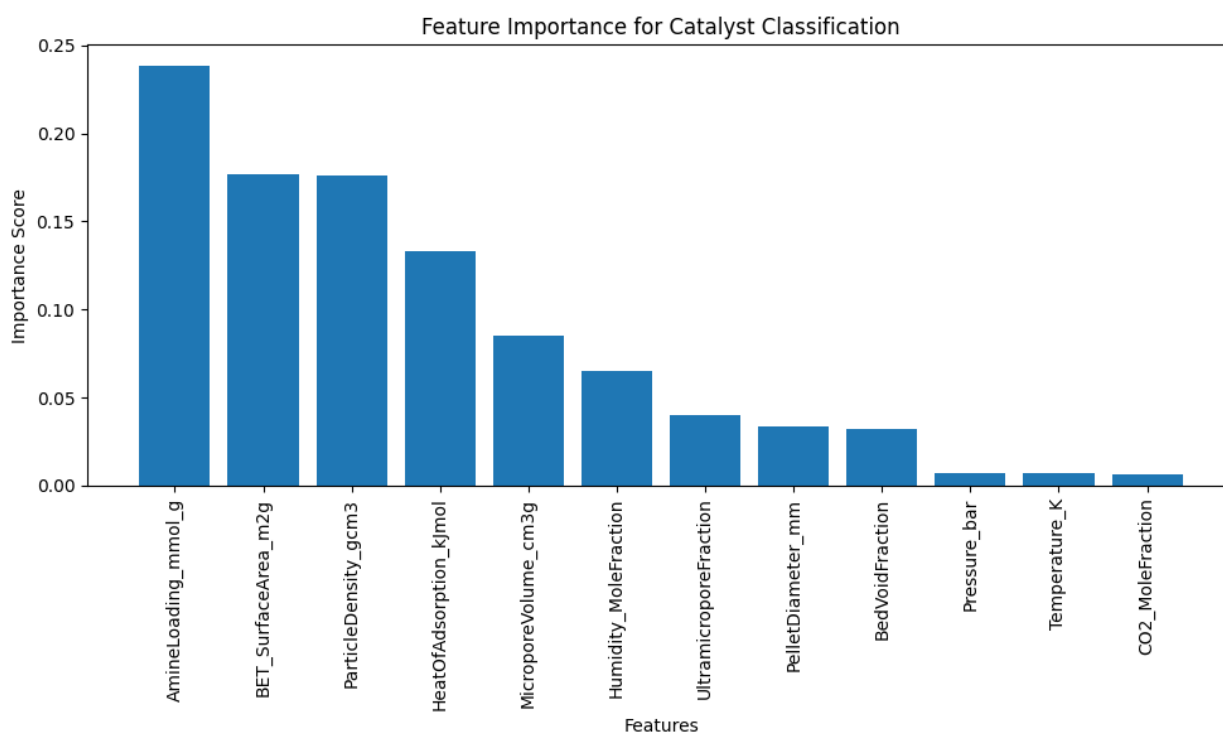


Figure 4. Feature importance from XGBoost model (adsorbent prediction).

The second set of influential parameters includes BET surface area (≈ 0.18) and particle density (≈ 0.18), both of which play important roles in distinguishing different porous materials. Surface area is a critical structural property that varies significantly among materials such as activated carbon, zeolites, MOFs, and porous organic polymers, thereby serving as an effective classification parameter. Similarly, particle density reflects the intrinsic structural composition and framework characteristics of different adsorbent materials. The heat of adsorption (≈ 0.13) also shows considerable importance, suggesting that adsorption energetics help differentiate adsorbents based on their interaction strength with CO₂ molecules. Additionally, micropore volume (≈ 0.08) and humidity mole fraction (≈ 0.06) contribute moderately, highlighting the role of pore structure and environmental conditions in influencing adsorbent identification.

In contrast, several parameters exhibit relatively low importance scores in the adsorbent classification model. Features such as ultra-micropore fraction, pellet diameter, and bed void fraction contribute only marginally to the classification accuracy. Furthermore, pressure, temperature, and CO₂ mole fraction display very small importance values, indicating that these operational parameters do not significantly help distinguish adsorbent types within the dataset. This observation is logical because these variables mainly influence adsorption performance rather than the intrinsic identity of the material itself. Overall, the feature importance analysis reveals that chemical composition and structural characteristics dominate adsorbent classification, whereas operating conditions play a comparatively minor role. These insights are valuable for guiding future research, as they emphasize the importance of material-specific descriptors when developing machine learning models for adsorbent identification and materials discovery in carbon capture systems.

4.4. Experimental Validation

The predictive performance of the XGBoost model was further evaluated using a small independent test dataset consisting of 18 samples, representing different classes of adsorbent materials used for CO₂ capture. The dataset included a range of materials such as activated carbon, zeolites, MOFs, porous organic polymers (POPs), alumina/silica, and amine-functionalized sorbents, allowing the model to be tested across diverse structural and chemical properties. A comparison between the actual and predicted CO₂ adsorption capacities showed strong agreement, with prediction errors generally remaining below 5% for most samples. This indicates that the trained model can effectively generalize beyond the training dataset while accurately capturing the relationship between material properties, operating conditions, and adsorption performance (Table 2).

A few deviations were observed for amine-functionalized sorbents, where the prediction error slightly exceeded 10% in some cases. This may be attributed to the complex chemisorption mechanisms involved in amine-based adsorption, which are more difficult to represent using only the structural and thermodynamic parameters included in the dataset. Nevertheless, the overall performance of the model demonstrates reliable predictive capability and confirms that the machine learning framework can effectively screen and estimate CO₂ adsorption capacity across multiple classes of adsorbent materials. While the 18-sample validation set demonstrates strong preliminary agreement, its small size is a limitation; extensive validation using substantially larger independent datasets will be a focus of future work.

Table 2. Dataset for actual adsorbent samples.

Adsorbent Category: MOF, Adsorbent Name: Cu₂4,4'-Bis(3,5-dicarboxyphenyl)-2,2'-bipyridine(DMF)₂_SNU-50 Temperature: 298 K, BET Surface Area: 2300 m²/g, Pore Volume: 1.08 cm³/g, Pore Diameter: 1.1 nm										
No.	Pressure (bar)	CO ₂ MF	H ₂ O MF	Amine Loading	Heat of Adsorption	Particle Density	Bed Void Fraction	Adsorption Capacity (mmol/g)		% Error
								Actual	Predicted	
1	1.3	0.756	0.012	3.39	45.5	0.67	0.46	4.79	4.67	−2.50521
2	3.12	0.334	0.024	1.677	57.11	0.523	0.23	7.56	7.88	4.23280
3	7.23	0.991	0.23	0.047	25.5	0.61	0.546	18.4	18.65	1.35862
Adsorbent Category: Alumina_Silica, Adsorbent Name: Silica Gel (SiO₂) Temperature: 298 K, BET Surface Area: 2297 m²/g, Pore Volume: 2.12 cm³/g, Pore Diameter: 0 nm										
4	5.5	0.345	0.006	1.22	18.9	2.1	0.45	1.44	1.409	−2.15277
5	8.36	0.849	0.011	0.81	26	3.35	0.42	1.35	1.367	1.25925
6	1.64	0.035	0.017	0.2	29.2	3.46	0.43	0.05	0.051	2
Adsorbent Category: Amine_Functionalized, Adsorbent Name: Amine-MIL-101_(Cr)-50%_Polyethylenimine-amine Temperature: 298 K, BET Surface Area: 946.8 m²/g, Pore Volume: 0.8 cm³/g, Pore Diameter: 0 nm										
7	0.91	0.81	0.062	2.88	68	0.96	0.44	1.13	1.07	−5.30973
8	6.07	0.63	0.061	4.06	40.4	1.34	0.43	2.79	2.75	−1.43369
9	2.07	0.85	0.043	1.61	64.8	1.09	0.4	0.45	0.472	4.88888
Adsorbent Category: POPs, Adsorbent Name: HCP-1 (Hypercrosslinked Polymer-1) Temperature: 296 K, BET Surface Area: 2304 m²/g, Pore Volume: 1.01 cm³/g, Pore Diameter: 0 nm										
10	7.99	0.734	0.034	1.456	43.4	1.03	0.54	7.57	7.34	−3.0389115
11	0.36	0.754	0.014	0.05	47.5	0.69	0.47	0.56	0.578	3.2142854
12	2.59	0.023	0.027	0.59	34.6	0.86	0.53	0.5	0.512	2.4
Adsorbent Category: Activated_Carbon, Adsorbent Name: Calgon BPL Activated Carbon Temperature: 303 K, BET Surface Area: 880 m²/g, Pore Volume: 0.46 cm³/g, Pore Diameter: 0 nm										
13	6.69	0.671	0.025	1.92	24.2	0.33	0.36	4.29	4.22	−1.63172
14	9.99	0.788	0.043	1.88	39.8	0.85	0.34	6.92	6.56	−5.202319
15	0.97	0.113	0.042	2.98	36.6	1.07	0.67	0.12	0.126	5
Adsorbent Category: ZEOLITE, Adsorbent Name: 13X Temperature: 298 K, BET Surface Area: 386.4 m²/g, Pore Volume: 0.2335 cm³/g, Pore Diameter: 0 nm										
16	8.14	0.835	0.012	0	25.6	1.36	0.58	7.67	7.565	−1.368973
17	5.24	0.333	0.026	0	63.1	1.89	0.45	0.98	0.93	−5.102016
18	2.92	0.901	0.021	0	39.45	1.7	0.37	2.56	2.645	3.34525

4.4.1. Interpretation of Performance

The strong validation performance of the XGBoost model demonstrates its effectiveness in predicting CO₂ adsorption capacity for systems primarily governed by physisorption mechanisms, where structural properties and operating conditions play a dominant role. Parameters such as BET surface area, pressure, CO₂ mole fraction, and micropore volume, identified as the most influential features in the feature importance analysis, provide sufficient information for accurate predictions across a wide range of porous materials. The gradient boosting framework of XGBoost enables the model to capture complex nonlinear interactions between these variables, resulting in high predictive accuracy and strong generalization capability. Minor inconsistencies observed for certain amine-functionalized or chemisorption-dominated systems indicate that additional chemical descriptors, such as amine density, binding energies, or surface functional group information, could further

improve prediction accuracy. Nevertheless, the consistently high performance across multiple adsorbent classes confirms the suitability of XGBoost as a robust machine learning approach for adsorption modeling. The agreement between experimental validation results and feature importance rankings further strengthens the reliability of the model and provides a strong foundation for data-driven materials screening and predictive carbon capture research.

4.4.2. Analysis of Percent Error

To evaluate the predictive performance of the developed XGBoost model, the difference between experimental and modeled CO₂ adsorption capacities was analyzed using a validation dataset of selected adsorption samples representing multiple classes of adsorbent materials. These included materials such as MOFs, zeolites, activated carbon, porous organic polymers (POPs), alumina/silica, and amine-functionalized sorbents, allowing the model to be tested across a wide range of structural and chemical properties. The comparison between experimental and predicted adsorption values demonstrated strong agreement, as reflected by the percent error distribution, where the majority of predictions fell within the $\pm 5\%$ error range. The calculated average prediction error was approximately 3.08%, indicating a high level of predictive accuracy for the developed model. For each sample, the percent error between the actual experimental value and the predicted value was computed using the following standard equation:

$$\% \text{ Error} = \frac{| \text{Actual} - \text{Predicted} |}{\text{Actual}} \times 100$$

This formula offers a normalized scale for evaluating prediction performance, especially useful when the target values vary over multiple orders of magnitude.

Average Percent Error (APE):

$$APE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\text{Actual}_i - \text{Predicted}_i}{\text{Actual}_i} \right| \times 100 = 3.08\%$$

Maximum Percent Error (MPE):

$$MPE = \max \left(\left| \frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \right| \times 100 \right) = 4.11\%$$

The Figure 5 illustrates the percent error analysis for CO₂ adsorption prediction obtained from the developed machine learning model across 18 validation samples. The horizontal axis represents the sample number, while the vertical axis indicates the percent prediction error (%), which reflects the difference between experimental and model-predicted adsorption capacities. Each vertical bar corresponds to the prediction error for an individual sample. Positive values indicate slight overestimation, whereas negative values represent underestimation by the model. The dashed horizontal lines at +5% and −5% denote the acceptable prediction accuracy threshold commonly used in adsorption modeling studies. Most of the bars fall within these limits, indicating strong agreement between predicted and experimental values. Only a few samples approach the $\pm 5\%$ boundary, demonstrating that the developed model maintains high predictive reliability and consistent performance across diverse adsorption samples.

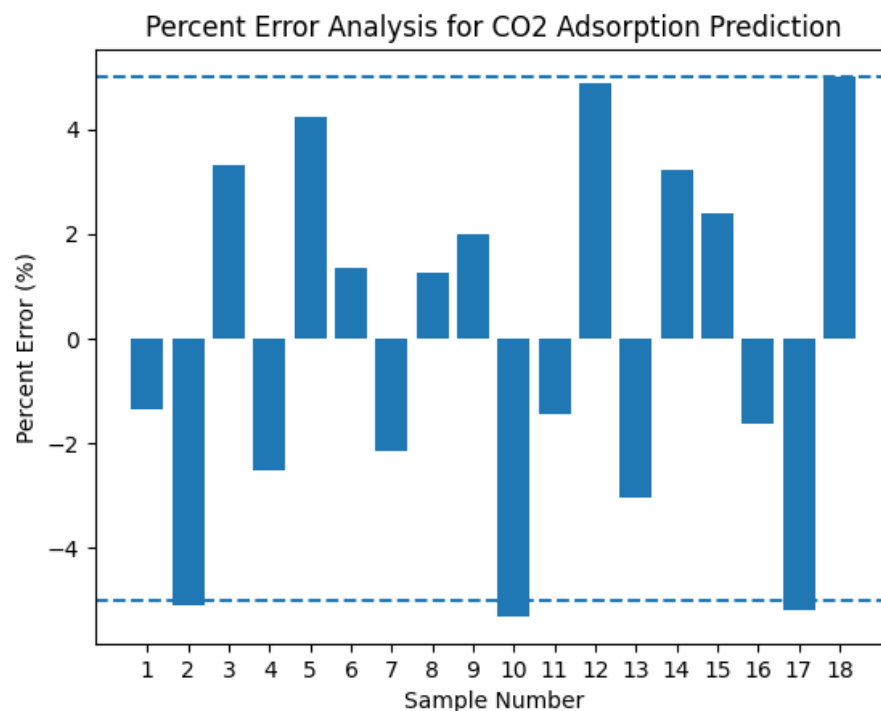


Figure 5. Percent error analysis for the selected 18 CO₂ adsorption samples.

5. Conclusions

A comprehensive machine learning framework based on the XGBoost algorithm was developed in this study to predict CO₂ adsorption capacity and identify the type of adsorbent used in adsorption systems. The results demonstrate that data-driven models can effectively capture the complex relationships between material properties, pore characteristics, and operating conditions, enabling accurate prediction of adsorption performance. The XGBoost model showed very high predictive accuracy, strong generalization ability, and robustness against overfitting, making it highly suitable for adsorption modeling tasks. Additionally, the feature importance analysis provided valuable insights into the relative influence of different parameters such as BET surface area, pressure, and micropore volume, highlighting the key factors controlling CO₂ adsorption behavior. These findings illustrate how machine learning techniques can significantly accelerate the screening and design of efficient adsorbent materials, reducing the dependence on time-consuming experimental trial-and-error approaches.

For future work, the predictive capability of the model can be further improved by expanding the dataset to include a wider range of porous materials and adsorption conditions. Incorporating additional chemical descriptors, such as surface functional groups, binding energies, and molecular interaction parameters, may enhance the ability of the model to capture chemisorption mechanisms, particularly in amine-functionalized systems. Furthermore, exploring other advanced machine learning architectures, such as deep neural networks or hybrid physics-informed models, could provide deeper insights into adsorption processes. Such developments will strengthen the integration of machine learning, computational modeling, and experimental research, ultimately enabling efficient discovery and optimization. Using this model, one can predict the adsorption capacity of any adsorbent used in the training dataset and predict its type with 95% accuracy.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr14132081/s1>, Original data from various literature sources are given in the supplementary material.

Author Contributions: Conceptualization, D.K., G.D.V. and R.K.A.; Data curation, C.K.B., B.T., P.B., S.A., P.M., N.J. and R.K.A.; Formal analysis, C.K.B., B.T., P.B., S.A., D.K., G.D.V., J.K., A.K.T., A.M., O.P.V. and R.K.A.; Funding acquisition, R.K.A.; Investigation, C.K.B., B.T., P.B., S.A., P.M. and N.J.; Methodology, C.K.B., B.T., D.K., G.D.V., O.P.V. and R.K.A.; Project administration, G.D.V. and R.K.A.; Resources, J.K., A.K.T., A.M. and R.K.A.; Software, C.K.B., B.T., P.B. and S.A.; Supervision, G.D.V. and R.K.A.; Validation, D.K., G.D.V., O.P.V., J.K. and R.K.A.; Visualization, C.K.B., B.T., P.B., S.A. and R.K.A.; Writing—original draft, C.K.B., B.T., P.B., S.A., P.M., N.J. and R.K.A.; Writing—review and editing, D.K., G.D.V., O.P.V., J.K., A.K.T., A.M. and R.K.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by SERB DST (Anusandhan National Research Foundation) New Delhi, grant number CRG/2023/000268.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author. Original data from various literature sources are given in the Supplementary Materials.

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

References

1. IEA. *CO₂ Emissions in 2022: Growth in Emissions Lower than Feared*; International Energy Agency: Paris, France, 2023.
2. Chatti, R.; Bansiwale, A.K.; Thote, J.A.; Kumar, V.; Jadhav, P.; Lokhande, S.K.; Biniwale, R.B.; Labhsetwar, N.K.; Rayalu, S.S. Amine loaded zeolites for carbon dioxide capture: Amine loading and adsorption studies. *Microporous Mesoporous Mater.* **2009**, *121*, 84–89. [\[CrossRef\]](#)
3. Jadhav, P.; Chatti, R.; Biniwale, R.; Labhsetwar, N.; Devotta, S.; Rayalu, S. Monoethanol amine modified zeolite 13X for CO₂ adsorption at different temperatures. *Energy Fuels* **2007**, *21*, 3555–3559. [\[CrossRef\]](#)
4. Siriwardane, R.V.; Shen, M.-S.; Fisher, E.P.; Losch, J. Adsorption of CO₂ on zeolites at moderate temperatures. *Energy Fuels* **2005**, *19*, 1153–1159. [\[CrossRef\]](#)
5. Abdul Wahab, M.; Sunarti, A.; Ramli, N.H.; Ahmad, A. 24 fractional factorial design for PVDF/Pebax film composite synthesis on gas selectivity study. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; p. 012046.
6. Siriwardane, R.V.; Shen, M.-S.; Fisher, E.P.; Poston, J.A. Adsorption of CO₂ on molecular sieves and activated carbon. *Energy Fuels* **2001**, *15*, 279–284. [\[CrossRef\]](#)
7. Lee, G.A.; Shi, X.; Park, A.-H.A. Technologies to Capture CO₂ directly from Ambient Air. *arXiv* **2022**, arXiv:2211.00791.
8. Sumida, K.; Rogow, D.L.; Mason, J.A.; McDonald, T.M.; Bloch, E.D.; Herm, Z.R.; Bae, T.-H.; Long, J.R. Carbon dioxide capture in metal–organic frameworks. *Chem. Rev.* **2012**, *112*, 724–781.
9. Cavenati, S.; Grande, C.A.; Rodrigues, A.E. Adsorption equilibrium of methane, carbon dioxide, and nitrogen on zeolite 13X at high pressures. *J. Chem. Eng. Data* **2004**, *49*, 1095–1101. [\[CrossRef\]](#)
10. Sevilla, M.; Valle-Vigón, P.; Fuertes, A.B. N-doped polypyrrole-based porous carbons for CO₂ capture. *Adv. Funct. Mater.* **2011**, *21*, 2781–2787.
11. Lu, W.; Yuan, D.; Sculley, J.; Zhao, D.; Krishna, R.; Zhou, H.-C. Sulfonate-grafted porous polymer networks for preferential CO₂ adsorption at low pressure. *J. Am. Chem. Soc.* **2011**, *133*, 18126–18129. [\[CrossRef\]](#)
12. Mohanty, P.; Kull, L.D.; Landskron, K. Porous covalent electron-rich organonitridic frameworks as highly selective sorbents for methane and carbon dioxide. *Nat. Commun.* **2011**, *2*, 401. [\[CrossRef\]](#)
13. An, J.; Rosi, N.L. Tuning MOF CO₂ adsorption properties via cation exchange. *J. Am. Chem. Soc.* **2010**, *132*, 5578–5579. [\[CrossRef\]](#) [\[PubMed\]](#)
14. An, J.; Geib, S.J.; Rosi, N.L. High and selective CO₂ uptake in a cobalt adeninate metal–organic framework exhibiting pyrimidine- and amino-decorated pores. *J. Am. Chem. Soc.* **2010**, *132*, 38–39.
15. Talapaneni, S.N.; Lee, J.H.; Je, S.H.; Buyukcakir, O.; Kwon, T.W.; Polychronopoulou, K.; Choi, J.W.; Coskun, A. Chemical blowing approach for ultramicroporous carbon nitride frameworks and their applications in gas and energy storage. *Adv. Funct. Mater.* **2017**, *27*, 1604658.
16. Gadipelli, S.; Guo, Z.X. Tuning of ZIF-derived carbon with high activity, nitrogen functionality, and yield—a case for superior CO₂ capture. *ChemSusChem* **2015**, *8*, 2123–2132. [\[CrossRef\]](#)
17. Langmuir, I. The constitution and fundamental properties of solids and liquids. Part I. Solids. *J. Am. Chem. Soc.* **1916**, *38*, 2221–2295. [\[CrossRef\]](#)

18. Streb, A.; Mazzotti, M. Performance limits of neural networks for optimizing an adsorption process for hydrogen purification and CO₂ capture. *Comput. Chem. Eng.* **2022**, *166*, 107974. [[CrossRef](#)]
19. Sultan, N.; Almansoori, A.; Elkamel, A. AI-based modeling and optimization of green methanol production process. *Comput. Chem. Eng.* **2025**, *203*, 109324. [[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.