scientific reports

OPEN

Modeling of CO_2 adsorption capacity by porous metal organic frameworks using advanced decision tree-based models

Jafar Abdi[1], Fahimeh Hadavimoghaddam[2], Masoud Hadipoor[3] & Abdolhossein Hemmati-Sarapardeh[4,5]✉

In recent years, metal organic frameworks (MOFs) have been distinguished as a very promising and efficient group of materials which can be used in carbon capture and storage (CCS) projects. In the present study, the potential ability of modern and powerful decision tree-based methods such as Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Random Forest (RF) was investigated to predict carbon dioxide adsorption by 19 different MOFs. Reviewing the literature, a comprehensive databank was gathered including 1191 data points related to the adsorption capacity of different MOFs in various conditions. The inputs of the implemented models were selected as temperature (K), pressure (bar), specific surface area (m^2/g) and pore volume (cm^3/g) of the MOFs and the output was CO_2 uptake capacity (mmol/g). Root mean square error (RMSE) values of 0.5682, 1.5712, 1.0853, and 1.9667 were obtained for XGBoost, CatBoost, LightGBM, and RF models, respectively. The sensitivity analysis showed that among all investigated parameters, only the temperature negatively impacts the CO_2 adsorption capacity and the pressure and specific surface area of the MOFs had the most significant effects. Among all implemented models, the XGBoost was found to be the most trustable model. Moreover, this model showed well-fitting with experimental data in comparison with different isotherm models. The accurate prediction of CO_2 adsorption capacity by MOFs using the XGBoost approach confirmed that it is capable of handling a wide range of data, cost-efficient and straightforward to apply in environmental applications.

Carbon dioxide (CO_2) plays a high influencing role in global warming[1]. As a result of consumption of fossil fuels especially in electricity generation, transportation and other industrial activities, CO_2 emission into the atmosphere is surging[2,3]. According to an investigation by Pachauri et al. in 2014, carbon dioxide's concentration in the atmosphere has increased from 280 to 400 ppm with 0.8 °C[1,4]. It is estimated that concentration of CO_2 would touch a peak of 600–700 ppm at the dawn of 22th century and resultantly it will lead in 4.5–5 °C growth in the average temperature of earth[5]. Thus, as to put a halt on rapid growth of CO_2 emission rates, United States Department of Energy (DOE) provided the world with a program aiming for reduction of CO_2 concentrations by utilization of high efficiency CO_2 capture plans. According to the issued program 90% of emitted CO_2 could be captured just with less than 35% of additional budget allocation to carbon capture and storage (CCS) programs[6]. These programs could be contemplated as promising approaches for the separation and sequestration of CO_2. As soon as carbon dioxide is separated, it could be stored underground, or alternatively it can be utilized in various industries such as oil industries in order to enhance the recovery[7–9]. Up to date, scientists have proposed numerous methods for CO_2 capture among which absorption[10–12], membranes[13,14], and carbon-based adsorbents[5,13] are well developed. Nevertheless; high total costs, low capacities, and challenging regeneration processes are some of their limitations[15]. A material which could be used successfully for an efficient CO_2 capture process, not only should possess good characteristics for CO_2 uptake, but also must release the captured carbon dioxide in the

[1]Faculty of Chemical and Materials Engineering, Shahrood University of Technology, 3619995161 Shahrood, Iran. [2]Ufa State Petroleum Technological University, Ufa, Russia 450064. [3]Department of Petroleum Engineering, Ahwaz Faculty of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran. [4]Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. [5]College of Construction Engineering, Jilin University, Changchun 130600, China. ✉email: hemmati@uk.ac.ir

regeneration step. Furthermore, the effectiveness of the capturing process could be maximized provided that the structure of the material could be modified using various functional groups and molecular tuning approaches.

Metal organic frameworks (MOFs) have been distinguished as a very promising and efficient group of materials which can be used in CCS projects because of their unique properties of being modifiable, stable in high temperatures, and having a chemical structure which can be easily adjusted[16–20]. Therefore, different advantages and disadvantages of MOFs have been investigated by several researchers such as an investigation done by Le et al., which reported how the architecture and active functional groups of MOFs could be controlled[17]. Moreover, having conducted various investigations on the thermal stability of MOFs, scientists have reported the impact of morphology and crystalline shape of these materials on their thermal stability[21–24]. Taking MOFs synthesized from strontium and calcium by the way of example, Yeh et al.[25] found that, as a result of their micro-porosity, these materials were thermally stable in all temperatures lower than 450 °C. Some other scientific investigations have found that MOFs could be modified for various applications just by changing the functional groups, which are located on their pore walls[26,27]. It is wieldy known that MOFs are comprised of metal ions clusters and linkers which are organic molecules. These linkers have a 3D-structure of pores and connecting channels. While 3D-structurevoids adsorb molecules as their host, the primary structure of these materials provides reversible channels and pores as soon as desorption step take place[26–28]. Properties of MOFs is determined by the selected linker and the metal. For example, zinc in the structure of IRMOF-1 is the metal which is located at the center of structures and is connected to the terephthalic molecules. Having this structure, IRMOF-1 benefits from existence of pores with high capacity of adsorption. Furthermore, there are various groups of MOFs with their unique structures, properties and application. Some MOFs like UMCs have unsaturated metallic centers[29], which provides carbon dioxide molecules with more active sites, hence facilitates formation of strong bonds between CO_2 and the structure. Other advantage of MOFs in comparison to other materials like zeolite is that the MOFs have typically wider pores, which boost diffusion rate of molecules not only in a single structure, but also between different crystals[30–33]. That is why scientists believe that MOFs are promising for adsorption of CO_2, emphasizing on their adjustable porosities and having a modifiable surface chemistry[5,34–36]. Regarding developing and testing MOF, there are some important challenges and obstacles to overcome or be solved using alternative methods. Firstly, experimental investigation of adsorption capacity of MOFs not only is time consuming, but also is not cost-effective. Secondly, data cannot be matched with the developed isotherms because typically these isotherms are proposed for a specific range of data[31]. So as to address the problem, many scientists have been trying to use soft computing methods by which not only the time and money could be saved, but also there is no need for simplifying assumptions and the data could be modeled more precisely. Artificial intelligence (AI) approaches are useful tools which enable us to estimate and develop representative models in various disciplines[37–42]. These powerful algorithms are able to model non-linear relationship which exist between influencing parameters. Up to the date, a plethora of models have been developed such as fuzzy logic, radial basis networks, support vector machine, and colony optimization[43–47].

In the current investigation, authors tried utilizing smart models for the prediction of nonlinear adsorption of CO_2 by MOFs. According to the literature, scientists have done fewer researches on the modeling of CO_2 adsorption by MOFs using AI methods. To model the CO_2 uptake capacity of different MOFs, new and powerful methods of CatBoost, LightGBM, Random Forest (RF), and XGBoost were employed. CatBoost (which is the abbreviated form of categorical boosting) is an open-source and modern gradient boosting library and it can deal with problems which are intrinsically heterogeneous thought handling categorical features. XGboost and LightGBM belong to GBDTs (Gradient Boosted Decision Trees) and in these methods a tree structure includes two separate steps. Firstly, the appropriate structure for the tree must be found. Secondly, leaf values must be set as soon as the tree structure is finalized[48]. Another approach which was used in this investigation was Random Forest. Since it was introduced, the accuracy of classifications improved significantly because not only growth of various trees were allowed, but also the program is able to vote for best and most distinguished class[49]. Reviewing literature, the authors gathered 1191 data points related to the adsorption capacity of different MOFs at various temperatures and pressures. The investigated MOFs are ZIF-8, Zn-MOF-74, Mg-MOF-74, PCN-16, MOF-5, PCN-11, BeBTB, Co-BDP, Mg_2(dobdc), Cu-BTTri, MOF-177, IRMOF-1, IRMOF-6, IRMOF-3, IRMOF-11, Cu-BTC, MOF-505, MOF-74, and MOF-2[50–53].

Implementation of models

Extreme gradient boosting model (XGBoost). In a tree-based ensemble method, a group of various classification and regression trees (CARTs) are utilized to minimize a set of objective functions applied to a training dataset. The XGBoost approach could be contemplated as a tree-based model which basically belongs to a gradient boosting decision tree (GBDT). So as to explain the CART's basic structure, it is comprised of three various nodes namely (a) the main node (root node), (b) internal nodes, and (c) leaf nodes like as illustrated in Fig. 1. The binary decision-making processes will split the root node into internal nodes. Doing so, the dataset which is located in the root will be classified into various nodes in the internal nodes and the final classification will take place in the leaf nodes as the final classes. Aiming for developing a powerful set according to the gradient booting model, an ensemble of CATRs are introduced and developed using determination of their influence by giving them a specific weight during the training process[54].

In a dataset where m dimension features and n examples exist, the modeling output (y) would be trained according to the following expression to form n tree nodes[55]:
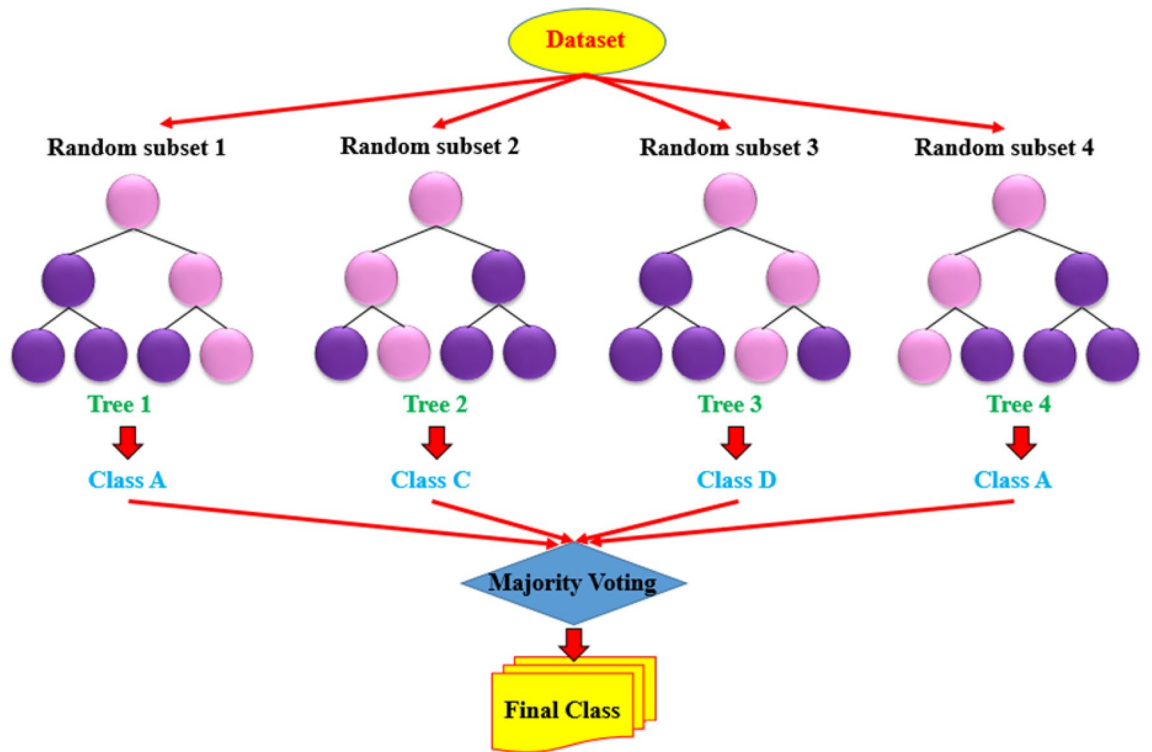
Figure 1. A schematic illustration of XGBoost depicting the main node, interior nodes, and the leaves.

$$\hat{y}_i = \sum_{k=1}^{N} f_k(X_i), \quad f_k \in f$$

$$where \; f = \left\{ f(X) = \omega_{q(x)} \right\}, \left(q : m \rightarrow T, \omega \in T \right) \tag{1}$$

where a binary leaf index will be formed by mapping an example X using a defined decision rule q(x). In Eqs. (1) and (2), the corresponding space of each regression tree is depicted by 'f'. Accordingly, f_k represents the kth tree, tree leaves are denoted by T, and their corresponding weight is determined by ω.

As the next step in the modeling, tree sets will be determined by minimizing an objective function denoted by L[55]:

$$L = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{N} \Omega(f_k)$$

$$with \; \Omega(f) = \gamma T + \frac{1}{2} \lambda \omega^2 \tag{2}$$

In the given formulation, Ω represents the regularization function and limits the model complexity by reducing the overfitting issues; loss function is shown by l and intrinsically is a differentiable convex; the minimum loss is denoted by γ and it is necessary in division of a new ultimate class as a leaf, and λ stands for the regulation coefficient. γ and λ facilitates the growth of the variance of the model and resultantly plummet the overfitting issue[55]. Every leaf in the boosting model has its own objective function, which should be minimized iteratively as follows[55]:

$$L^{(t)} = \sum_{i=1}^{n} \left\{ l\left(y_i, \hat{y}_i^{(t-1)} \right) + f_t(X_i) \right\} + \Omega(f_t) \tag{3}$$

In the presented formula, t is the iteration number for the minimization of a leaf objective function in the training step. So as to improve the model, an algorithm known as the greedy algorithm is utilized, which is designed to provide enough space for regression trees. Doing so, XGBoost model can continuously update its final results through improving the preciseness of the objective functions[55]:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(X_i) \tag{4}$$

Shrinkage strategy is another strategy that XGBoost method utilizes properly. In this strategy, a learning factor is defined and its learning rate is regulated in every gradient boosting step through definition of additional weights. Shrinkage strategy prevents overfitting problem by restricting future trees from affecting previously formed trees[56].

Light gradient boosting machine (LightGBM). On the basis of gradient learning theories, a novel learning machine was developed which is known as LightGBM[57]. The LightGBM approach, in comparison to XGBoost, needs lower memory spaces and speeds up the training step by using a histogram[58]. LightGBM can form a histogram with a width of 'k' by discretizing eigenvalues into 'k' different bins. Furthermore, the aforementioned approach diminishes the need for a set of pre-sorted results and values will be saved in an integer with a size of eight bits, which results in a drastic reduction of memory consumption. That said, such as kind of approach unfortunately results in dipping the model's preciseness. Leaf-wise approach has also been utilized in LightGBM. Drawing comparison between traditional growth strategies and the leaf-wise strategy, it must be admitted that this approach is considerably more efficient than the others. What makes the leaf-wise strategy more efficient than the alternative level-wise strategy is this fact that the leaves existing in the same layer are properly taken into consideration, which diminish the need for unnecessary allocation of memory. Therefore, by finding leaves with the maximum branching gain, errors could be minimized and a better accurateness could be achieved. In Fig. S1, the leaf-wise strategy of tree development is illustrated. An important drawback of leaf orientation is that as the decision trees grow deeper, unfavorable overfitting gets exacerbated. That said, simultaneously when LightGBM is resulting in overfitting, by definition of an upper limit on depth of the leaf top, a high efficiency will be achieved[57,58].

Regarding the formulation of a LightGBM model, parameters and calculations could be introduced as follows[59]:

Being given a training dataset of $X = \{x_i, y_i\}_{i=1}^{m}$, the LightGBM approach approximates a $\hat{f}(x)$ according to a $f^*(x)$ aiming for the minimization of the desirable values of a loss function shown by $L(y, f(x))$:

$$\hat{f}(x) = \arg \min_f E_{y,x} L(y, f(x)) \tag{5}$$

A wide variety of T regression trees with the formulation of $\sum_{t=1}^{T} f_t(x)$ will be formed as LightGBM sets which can be used for the model's approximation. In a defined regression tree ($W_{q(x)}, q \in \{1, 2, \ldots, N\}$), w stands for a vector which represents the weight of each leaf node, N depicts how many leaves exist in a tree, and the decision rules applied to trees are shown by q. The training step of the model development, at step t, is formulated as follows[59]:

$$G_t \cong \sum_{i=1}^{N} L(y_i, F_{t-1}(x_i) + f_t(x_i)) \tag{6}$$

The objective functions are determined by utilizing Newton's method.

Gradient boosting with categorical features support (CatBoost). In the CatBoost approach aiming for successful application of categorical boosting technique, categorical columns must be utilized. The aforementioned column benefit from a range of processing techniques. The target-based statistics and the one_hot_max_size (OHMS) are the most important ones. In the growth of every branch of the running tree, a greedy approach will be applied to facilitate finding changes in the combination of features of CatBoost method[60]. In the CatBoost method, the following steps must progress properly for every feature of various categories[48,61]:

1. To form a random subgroup of the available records
2. Convert labels to integer
3. According to Eq. (7), features related to categories must be transformed into numeric form:

$$avgTarget = \frac{countInClass + prior}{totalCount + 1} \tag{7}$$

In the given formulation, targets are counted by *countInClass*. Each target is assigned with some categorical features, each having a value of one, and all the previous objects will be counted in *totalCount* (the *prior* ones which are needed in counting the objects are determined by the initiating parameters)[48,61].

Random forest (RF). An ensemble of various decision trees forms a random forest, in which trees are being trained in parallel. The superiority and the importance of each decision tree is determined by the algorithm[62]. Additionally, a built-in property of RF classifier which is utilized to select various features makes the RF able to manage different inputted features without need for removing a number of parameters for reduction of dimension[63]. In the modeling, in order to enhance the diversity of trees of the forest, the RF technique employs a method known as Bagging (which stands for bootstrap aggregating). Typically, the population of trees are given as an input to the model, and accordingly the model will divide data points into various sets. Being a type of random sampling methods, bagging employs just a third of data points for the training step of a subtree development process and the remaining data points are referred as the out-of-bag (OOB). Furthermore, in the model
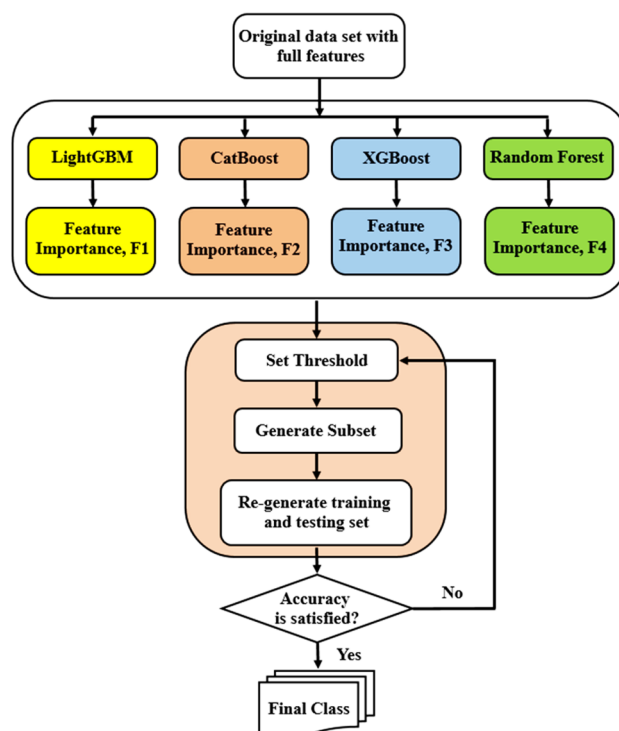
Figure 2. Feature selection and classification using different algorithm for predicting CO_2 adsorption on MOFs.

development cross-validation of results are not required in the RF, as the accuracy of the model is assessable using the OOB's errors[49]. In Fig. S2, the strategy of RF method is illustrated. A successful training process will happen if a training sample is given to the model as a requirement. If there is a training data set in the form of $D = [(x_1.y_1).(x_2.y_2).\cdots\cdot(x_n.y_n)]$, the defined training dataset for tree h_t will be denoted by D_t, and the resulting prediction of the out-of-bag dataset of sample x will be H^{oob}, as it is given as follows[49]:

$$H^{oob}(x) = argmax \sum_{t=1}^{T} I(h_t(x)) = y \tag{8}$$

For the purpose of the modeling, the error of the OOB dataset is generalized as follows[49]:

$$\varepsilon^{oob}(x) = \frac{1}{|D|} \sum_{(x.y)D} I\left(H^{oob}(x) \neq y\right) \tag{9}$$

The RF's operation should be random and this feature is controlled by a parameter formulated as $k = log_2 d$[49]. The significance of a characteristic of a variable X_i could be measured using the following expression[49]:

$$I(X_i) = \frac{1}{B} \sum_{t}^{B} \widetilde{OOBerr_{t^i}} - OOBerr_t \tag{10}$$

Accordingly, in the X vector, the ith factor is denoted by X_i, B depicts how many trees exist in the current RF, the predicted error of the OOB samples is defined by $\overline{OOBerr_{t^i}}$, which stands for the feature X_i of tree t, and finally the initial OOB data samples are given as the $OOBerr_t$, which includes the permuted variables.

The significance of the character permutation process illustrates how much a feature is salient for the estimation. Resultantly characteristic permutation severely changes the model's estimation and it can be conclusively observed that an insignificant feature possess a very limited power for changing the prediction of the model[64]. Figure 2 shows the feature selection and classification using different algorithm for predicting CO_2 adsorption on MOFs.

Data gathering

A dataset comprising of approximately 1191 experimental results of CO_2 uptake on various MOFs including ZIF-8, Zn-MOF-74, Mg-MOF-74, PCN-16, MOF-5, PCN-11, BeBTB, Co-BDP, Mg_2(dobdc), Cu-BTTri, MOF-177, IRMOF-1, IRMOF-6, IRMOF-3, IRMOF-11, Cu-BTC, MOF-505, MOF-74, and MOF-2 was collected reviewing the literature[51–53]. Table 1 represents the details of the data used in this study. Moreover, Table 2 lists down the statistical details of the whole dataset which was collected for the purpose of this investigation. Trying

No	MOF	Surface area (m²/g)	Pore volume (cm³/g)	Temperature (K)	Pressure (bar)	CO₂ uptake (mmol/g)	No. data	Refs.
1	MOF-2	345	0.227	298	0–42.2	0–3.20	39	51
2	MOF-74	816	0.4	298	0–42.4	0–10.4	31	51
3	MOF-505	1547	1.83	298	0–42.5	0–10.4	31	51
4	Cu₃(BTC)₂	1781	0.43	298	0–42.4	0–10.7	30	51
5	IRMOF-11	2096	0.92	298	0–42.4	0–14.8	32	51
6	IRMOF-3	2160	1.07	298	0–42.2	0–18.9	37	51
7	IRMOF-6	2516	1.14	298	0–42.5	0–19.7	37	51
8	IRMOF-1	2833	0.18	298	0–42.2	0–22.0	35	51
9	MOF-177	4508	1.59	298	0–42.5	0–33.9	70	52
10	CuBTTri	1750	0.713	313	0.546–39.86	1.16–16.99	43	52
11	Mg₂(dobdc)	1800	0.5727	313	0.0005–35.32	0.09–15.15	51	52
12	CoBDP	2030	0.93	313	1.31–39.87	0.28–16.56	31	52
13	BeBTB	4030	1.701	313	1.02–38.83	1.79–30.17	40	52
14	PCN-11	1931	0.91	220–310	0–30.80	0 -22.84	104	53
15	MOF-5	3500	1.31	220–310	0–31.43	0–30.34	96	53
16	PCN-16	2273	1.06	220–310	0–31.66	0–21.51	120	53
17	HKUST-1	1690	0.66	220–310	0–24.99	0–19.36	130	53
18	Mg-MOF-74	1332	0.61	280–310	0–31.63	0–13.55	63	53
19	Zn-MOF-74	885	0.41	280–310	0–31.39	0–10.65	63	53
20	ZIF-8	1980	0.65	220–310	0–31.6	0–11.83	108	53
	Total			220–313	0–42.5	0–33.9	1191	

Table 1. Details of the experimental data used in this study.

|	Pressure (bar)	Temperature (K)	Surface area: S (m²/g)	Pore volume: V_P (cm³/g)	CO₂ uptake (mmol/g)
Mean	10.469	286.319	2156.937	0.877	10.910
STD	11.539	28.238	969.463	0.409	8.512
Min	0	220	345	0.180	0
25%	1	280	1690	0.610	4.9
50%	5.2	298	1980	0.910	9.899
75%	17.714	310	2273	1.070	16.065
Max	42.5	313	4508	1.830	141

Table 2. Statistical details of the dataset gathered in this paper.

to conceptualize and perceive the effect of various parameters on the MOF's CO₂ uptake capacity, the authors have incorporated temperature (K), pressure (bar), surface area (m²/g) and pore volume (cm³/g). To provide a precise and reliable set of models, deliberately about 80% of data points were devoted to model establishment and training phase and just about 20% were considered for testing phase. Therefore, the models' preciseness and trustworthiness were ensured by utilizing two statistical factors, namely root mean-square error (RMSE) and coefficient of determination (R²) were used[65–67]:

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^{N} \left(y_i^{estimated} - y_i^{exp.} \right)^2 \right)^{0.5} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i^{estimated} - y_i^{exp.} \right)^2}{\sum_{i=1}^{N} \left(y_i^{etimated} - \overline{y}^{exp.} \right)^2} \tag{12}$$

Outlier detection. Having a range of uncertainties and being associated with outliers, experimental data can introduce errors in the modeling process. To prevent any undesirable and not reliable outcome, the experimental models must undergo data evaluation and outlier detection. For the case of CO₂ adsorption on various MOFs, the faulty data points will enormously impact the preciseness of the predicting models. In the outlier detection process, as soon as a notable deviation of a data point from the others is detected, it will be recognized as an outlier. This study seeks benefits from the well-known leverage value procedure for spotting an outlier[68].
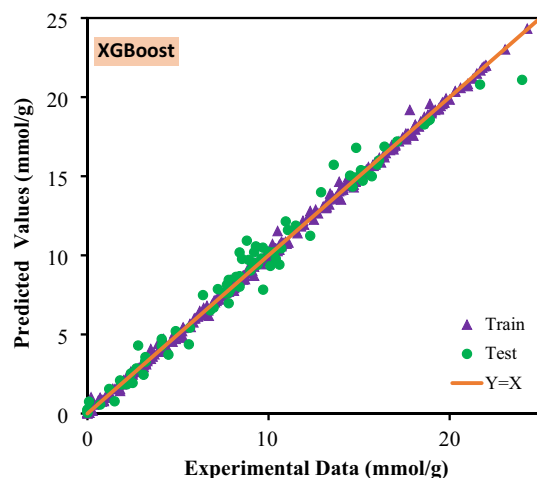
Figure 3. Crossplots of the proposed machine learning models in this study.

Firstly, the standardized residual and the Hat and value corresponding to any input data were calculated. The following formulation was used for the determination of Hat matrix[37]:

$$H = X \left(X^T X \right)^{-1} X^T \tag{13}$$

In the abovementioned formula, X represents a matrix with a size of $N \times P$, where the total number of data points is depicted by N and the number of inputs' features is denoted by P. Additionally, an alarming leverage value was determined as follows[37]:

$$H^* = \frac{3(P + 1)}{N} \tag{14}$$

Results and discussion

Model development. For predicting the amount of CO_2 absorbed on the surface of various MOFs, we developed various models using XGboost, LightGBM, Catboost and RF methods. In order to avoid overfitting, the grid search was used to optimize hyperparameters of models. Each model's grid search hyperparameters were different, and the importance of the hyperparameters was determined by theoretical and practical considerations. For each model, the following hyperparameters were employed. Table S1 shows the optimal values of the major hyperparameters, as well as the search intervals for the hyperparameters set for the machine learning models used in this study.

In Table S1, n_estimators represents the number of trees, subsample shows subsample ratio of columns when constructing a tree, C denotes a degree of importance that is given to misclassifications, max_depth is maximum depth of a tree, feature_fraction is parameters randomly selected in each iteration for building trees, learning_rate controls the impact of each tree on the final outcome.

Model implementation and accuracy evaluation. The most salient feature of each model is its accuracy and trustworthiness. In Fig. 3 and Fig. S3, the predicted CO_2 adsorption capacity is illustrated versus the experimental data for the implemented models, in which the closer predicted data to the experimentally obtained data (forming a y = x line), the more accurate the model is. In the given figure, almost all models illustrate a slope very close to unity which prove that all of them can be reliable when anticipation of real data is needed. As an important statistical criterion, the calculated error corresponding to every data point is depicted in Fig. 4 and Fig. S4, which provide the readers with comprehensive information about the preciseness of the proposed models. Delineated in the given illustrations, data points fluctuate closely around the zero-error line which indicates that the models were developed well. According to this figure, the calculated error for RF, Catboost, LightGBM, and XGBoost are limited to $\pm 10, \pm 8, \pm 7$, and ± 12.5, respectively. The cumulative frequency of errors of each developed model is illustrated in Fig. 5. As it can be seen, 95% of the data predicted by XGBoost model has an error less than 1%. Moreover, Catboost model could predict 89% data with errors less than 1%.

In Table 3, the accuracy of each model is reported based on statistical criteria. As it can be perceived from this table, having an R^2 of 0.9992 and 0.9733 in the training and testing steps of modeling, the XGboost is the most trustworthy model followed closely by the Catboost model.

Moreover, the calculated RSME for every model is given in Fig. S5 from which it is obvious that the XGBoost model is the most aureate model having an RSME of 0.568. Following closely, the LightGBM approach had the second lowest RSME value. According to both of RSME and R^2 criteria, RF was determined as the poorest predicting model.
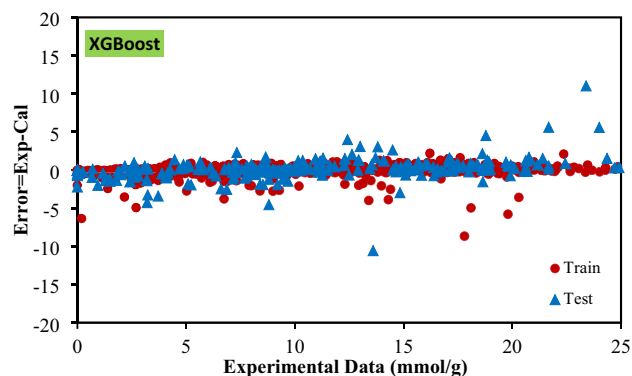
**Figure 4.** Error distribution plots of machine learning models for training and test sets.
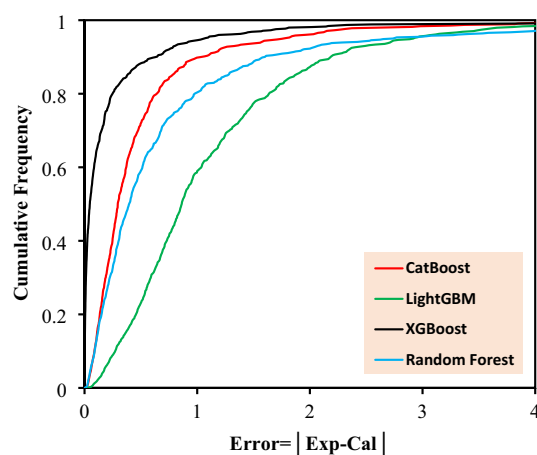
Figure 5. The cumulative frequency plot for the developed predictive models.

| Statistical criteria		R²	RSME
XGBoost	Train	0.9992	0.2515
	Test	0.9733	1.1649
	Total	0.9955	0.5682
LightGBM	Train	0.9876	0.9819
	Test	0.9599	1.4245
	Total	0.9837	1.0853
CatBoost	Train	0.9660	1.6270
	Test	0.9653	1.3259
	Total	0.9659	1.5712
Random Forest	Train	0.9467	2.0364
	Test	0.9457	1.6600
	Total	0.9466	1.9667

Table 3. Calculated statistical criteria for the developed models.

With increasing pressure, the saturation process of the MOFs' pores will be facilitated according to the room-temperature isotherms. From these isotherms, it can be perceived that the uptake capacity of each MOF is qualitatively related to its surface area and among all types of assessed MOFs in the current study, the MOF-177, BeBTB, MOF-5 and PCN-11 showed outstanding and considerable CO_2 adsorption capacity. As the pore sizes are greater and more efficient, the contribution of the pressure regime appearance gets more influence on the CO_2 uptake capacity[69,70]. Comparison among all mentioned porous materials, the MOF-177 had a maximum capacity of 33.5 mmol/g, because of its high surface area (4508 m^2/g) nearly twice of IRMOF-11. Figure 6a illustrates a detailed representation of uptake capacities in different pressures at 298 K for the investigated MOFs. Also, Fig. 6b
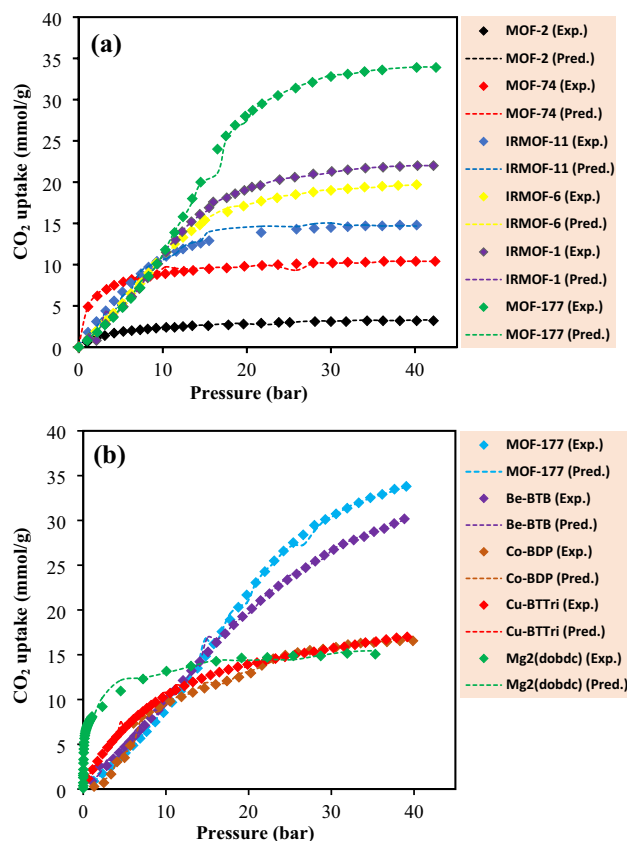
Figure 6. Comparison between experimental and predicted CO_2 uptake capacities by XGBoost model for the investigated MOFs at different temperatures: (**a**) 298 K and (**b**) 313 K.
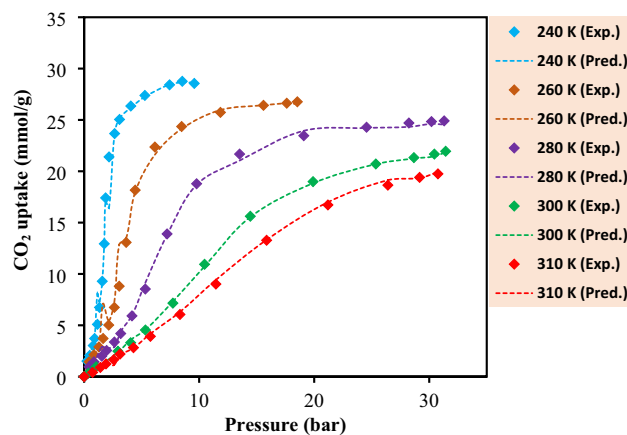
Figure 7. The effect of temperature on the CO_2 adsorption capacity.

draws a comparison between the CO_2 adsorption efficiency of MOF-177 at 313 K and a range of other MOFs. It has also been discovered that the amount of CO_2 adsorption on the MOF structure is directly related to the surface area and the polarity of the surface. As these features are higher, more CO_2 will be absorbed and even in low pressures a high adsorption will be achieved using some MOFs like Cu-BTTri and Mg_2(dobdc). Moreover, operating at a high pressure significantly affects the CO_2 adsorption capacity.

Regarding Co-BDP, a step-like appearance for the adsorption isotherm can be seen, which is potentially resulted from a phenomenon known as the gate opening. This event takes place due to the notable flexibility of the MOF's framework[71,72]. This claim must not be interpreted as if the other MOFs are not efficient and practical in CO_2 removal endeavors, but it means that MOF-177 has an excellent CO_2 removal capacity at 35 bar[51]. Furthermore, as presented in Fig. 7, dependency of adsorption capacity on the operational temperature is considerable,
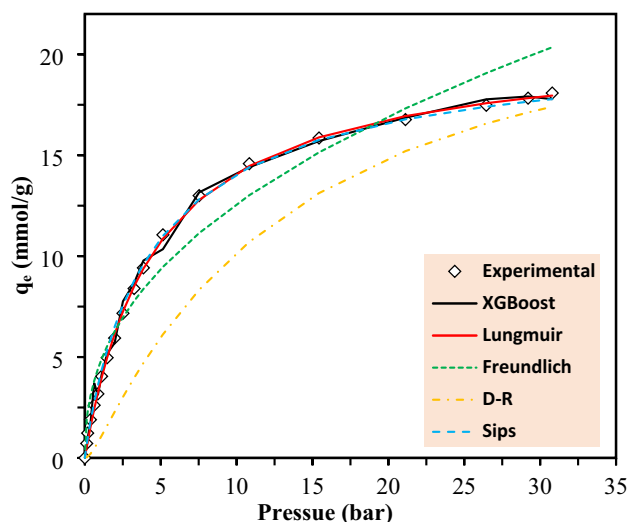
Figure 8. Comparison between different isotherms and XGBoost model for CO_2 adsorption by PCN-11 at 313 K.

and generally for MOF-5, the CO_2 uptake capacity was higher at room temperature than its capacity at 313 K. Thus, temperature negatively affects the carbon dioxide adsorption efficiency. In other words, with increasing temperature, the adsorption capacity decreases.

Ability of various models in prediction of CO_2 adsorption capacity at low pressure ranges is shown in Fig. S6, in which the models were utilized to estimate CO_2 removal efficiency of Mg_2(dobdc) at 313 K. Among all models, the XGBoost and CatBoost were found to be the most trustable models. Moreover, this model can be successfully compared with different isotherm models. As illustrated in Fig. 8 and Fig. S7, XGBoost as the best model along with Langmuir, Freundlich, Dubinin-Radushkevitch (D-R) and Sips isotherm models were fitted with experimental data for PCN-11 and Mg-MOF-74. Figures show that both Langmuir and XGBoost models predicted the results well. Langmuir isotherm model showed well-fitting with correlation coefficient of 0.999 and 0.980 for PCN-11 and Mg-MOF-174, respectively, but totally the XGBoost models represented better fitting correlation with R^2 values of 0.998 for both MOFs.

Microporous MOFs are new emerging generation of promising adsorbents which can drastically improve the CO_2 removal efficiency due to their high selectivity ranges. Using these materials, a higher CO_2 uptake capacity could be achieved[73] specially between 5 and 40 bar which is appropriate for the separation of CO_2 and H_2[74,75]. In power plants in which coal is mainly consumed for generation of power, a wide range of pollutants will be emitted to the environment including carbon dioxide. As a highly efficient process, pre-combustion capture of CO_2 can take place in the integrated gasification and combined cycle method. In this system, carbon dioxide could be separated from H_2 and removed from the process[5,65,76]. Additionally, the intrinsic feature of the surface of these materials will result in a better and more powerful interaction between CO_2 and the surface of the MOFs[77]. Due to their promising removal efficiency, many scientists have paid attention to integration of these materials into various industries including gas separation and purification processes in the petroleum industry specially when separation of CO_2/N_2[78], CO_2/CH_4[71], and O_2/N_2[79] are desired. To successfully investigate the adsorption process, a profusion of experimental endeavors has been done and lot of isotherm models have been developed. Nevertheless, a range of drawbacks have prevented both of experimental works and theoretical approaches to successfully be used. Firstly, the designing and conducting experiments are time consuming and costly. Additionally, a wide range of various assumptions must be made to simplify the problem when the currently available adsorption isotherms are going to be utilized. Resultantly, not only a big inventory of data is needed, but also modeling process encounters errors from first the steps of the investigation. These errors will be exacerbated if a set of needed data is not accessible[31]. Therefore, application of cost-effective, time-saving, robust, and simple model is vital in the investigation and prediction of CO_2 adsorption capacity by MOFs. The presented investigation has dealt with the adsorption of CO_2, as one of the main causes of global warming issue[5], on MOFs, which are known as very promising materials for CO_2 removal, by applying some well-known soft computing techniques. The most outstanding feature of these models was this fact that no limitation for the development of these comprehensive models were needed to be applied. Therefore, being capable of handling a wide range of data, the implemented models have estimated the adsorption capacity at various pressures and temperatures accurately. However, the models' performance and reliability are highly dependent on the selection of appropriate input parameters[1]. Additionally, successful application of smart models requires a large number of data points and excellent skills of programming.

Based on the calculated Hat values, an area limited by standardized residuals of ± 3 and $0 \le H \le H^*$ is detected as the acceptable region. H^* is determined for every model. The model is valid if majority of the data points are being located in the determined region. Figure 9 shows the Williams plot relating to the proposed XGBoost model as the best approach in this study. According to this figure, almost all data points were inside the acceptable region
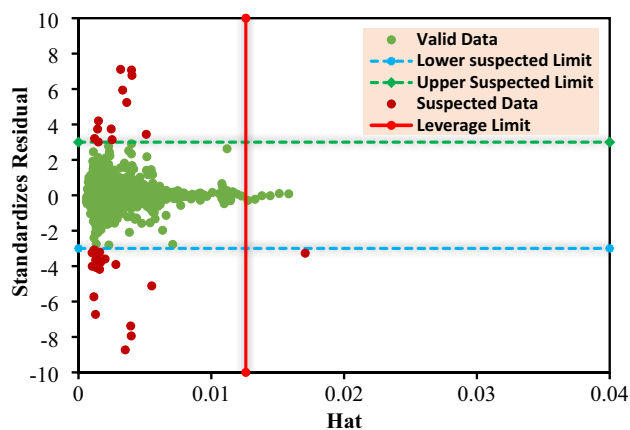
Figure 9. Williams plot for outlier detection of the proposed XGBoost model.
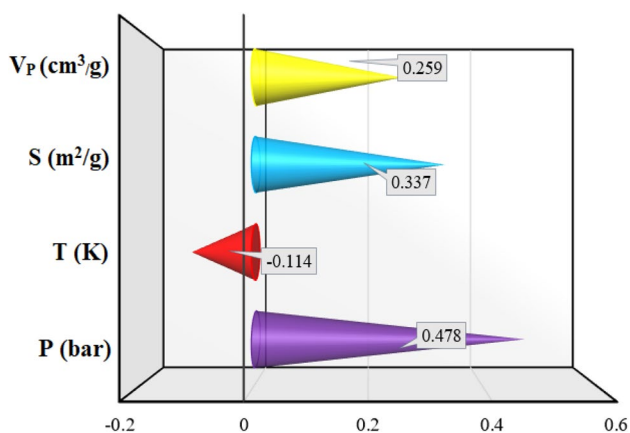
Figure 10. Evaluation of the input parameters' impact.

and just a few data points are located outside the desirable area. Thus, both developed model and experimental data are statistically valid.

Sensitivity analysis. To investigate the impact of various parameters on the adsorption capacity of MOFs, a relevancy factor was defined as follows[80–82]:

$$r = \frac{\sum_{i=1}^{N} \left(X_{k,i} - \overline{X}_k \right) \left(y_i - \bar{y} \right)}{\sqrt{\sum_{i=1}^{N} \left(X_{k,i} - \overline{X}_k \right)^2 \sum_{i=1}^{N} \left(y_i - \bar{y} \right)^2}} \tag{15}$$

in which N, $X_{k,i}$, Y_i, $\overline{X}_k$, $\overline{Y}$ stand for the number of assembled data points, the kth parameter of ith input value, i^{th} output data, mean of the kth input parameter, and the average of the outputs, respectively.

Laying between ± 1, the relevancy factor depicts how much an input parameter affects the CO_2 removal capacity of the MOFs. As the absolute value of the relevancy factor corresponding to a specified parameter increases, the carbon dioxide adsorption capacity will change more dramatically with any change in that parameter. The calculated relevancy factor for each affecting parameter on the adsorption of CO_2 on various MOFs is illustrated in Fig. 10. According to the calculated r values, among all investigated parameters, only temperature negatively impacts the adsorption capacity having an $r = -0.114$. Furthermore, it was found that pressure and surface area of MOFs have the most and the second most significant effects on the CO_2 adsorption capacities possessing "r" factors of 0.478 and 0.337, respectively. Pore volume also was found to have a positive impact on the CO_2 adsorption capacity of MOFs as provides more adoption sites for CO_2 molecules to be captured by them.

Conclusions

The adsorption capacity of various MOFs for CO_2 capture was modeled by development of robust models using 1191 data points. The data was used to train and test the XGBoost, LightGBM, CatBoost, and RF approaches and it was found that the XGBoost as the best fitting model and the Catboost as the second most trustworthy model

predicted the CO_2 adoptions more precisely than other methods. The R^2 and RMSE values of 0.9955 and 0.5682, 0.9659 and 1.5712, 0.9837 and 1.0853, and 0.9466 and 1.9667 were obtained for XGBoost, CatBoost, LightGBM, and RF approaches, respectively. Additionally, it was found that all parameters except temperature have positive impact on the CO_2 capture by MOFs and temperature had the smallest influence on the CO_2 uptake capacity. It has also been discovered that the amount of CO_2 adsorption on the MOF structure is directly related to the surface area and the polarity of the surface. As these features are higher, more CO_2 will be absorbed and even at low pressures, a high adsorption could be achieved using some MOFs like Cu-BTTri and Mg_2(dobdc). Langmuir isotherm model showed well-fitting with correlation coefficient of 0.999 and 0.980 for PCN-11 and Mg-MOF-174, respectively, but totally the XGBoost models represented better fitting correlation with R^2 values of 0.998 for both MOFs. The obtained results of the current work could be beneficially utilized in the environmental studies such as carbon capture and gas separation and purification.

References

1. Dashti, A. *et al.* Efficient hybrid modeling of CO_2 absorption in aqueous solution of piperazine: Applications to energy and environment. *Chem. Eng. Res. Des.* **144**, 405–417 (2019).
2. Amirkhani, F., Mosadegh, M., Asghari, M. & Parnian, M. J. The beneficial impacts of functional groups of CNT on structure and gas separation properties of PEBA mixed matrix membranes. *Polym. Test.* **82**, 106285 (2020).
3. Mosadegh, M., Amirkhani, F., Harami, H. R., Asghari, M. & Parnian, M. J. Effect of Nafion and APTEOS functionalization on mixed gas separation of PEBA-FAU membranes: Experimental study and MD and GCMC simulations. *Sep. Purif. Technol.* **247**, 116981 (2020).
4. Pachauri, R. K. *et al.*, *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Ipcc, 2014).
5. Hadipoor, M., Keivanimehr, F., Baghban, A., Ganjali, M. R. & Habibzadeh, S. *Carbon Dioxide as a Main Source of Air Pollution: Prospective and Current Trends to Control* 623–688 (Elsevier, New York, 2021).
6. Ciferno, J. P., Fout, T. E., Jones, A. P. & Murphy, J. T. Capturing carbon from existing coal-fired power plants. *Chem. Eng. Prog.* **105**(4), 33 (2009).
7. Hadipoor, M., Taghavi, H. & Taghavi, H. Experimental investigation of CO_2 injection performance in heterogeneous reservoirs: Parametric analysis. *Pet. Sci. Technol.* **38**(16), 837–848 (2020).
8. Dargahi-Zarandi, A., Hemmati-Sarapardeh, A., Shateri, M., Menad, N. A. & Ahmadi, M. Modeling minimum miscibility pressure of pure/impure CO2-crude oil systems using adaptive boosting support vector regression: Application to gas injection processes. *J. Pet. Sci. Eng.* **184**, 106499 (2020).
9. Mahdaviara, M. *et al.* Toward smart schemes for modeling CO_2 solubility in crude oil: Application to carbon dioxide enhanced oil recovery. *Fuel* **285**, 119147 (2021).
10. Aghaie, M., Rezaei, N. & Zendehboudi, S. A systematic review on CO_2 capture with ionic liquids: Current status and future prospects. *Renew. Sustain. Energy Rev.* **96**, 502–525 (2018).
11. Aghaie, M., Rezaei, N. & Zendehboudi, S. Assessment of carbon dioxide solubility in ionic liquid/toluene/water systems by extended PR and PC-SAFT EOSs: Carbon capture implication. *J. Mol. Liq.* **275**, 323–337 (2019).
12. Jiang, Y. *et al.* Metal–organic frameworks with target-specific active sites switched by photoresponsive motifs: Efficient adsorbents for tailorable CO_2 capture. *Angew. Chem. Int. Ed.* **58**(20), 6600–6604 (2019).
13. Yong, Z., Mata, V. & Rodrigues, A. R. E. Adsorption of carbon dioxide at high temperature—a review. *Sep. Purif. Technol.* **26**(2–3), 195–205 (2002).
14. Harami, H. R. *et al.* Mass transfer through PDMS/zeolite 4A MMMs for hydrogen separation: Molecular dynamics and grand canonical Monte Carlo simulations. *Int. Commun. Heat Mass Transf.* **108**, 104259 (2019).
15. Liu, J., Thallapally, P. K., McGrail, B. P., Brown, D. R. & Liu, J. Progress in adsorption-based CO_2 capture by metal–organic frameworks. *Chem. Soc. Rev.* **41**(6), 2308–2322 (2012).
16. Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **295**(5554), 469–472 (2002).
17. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**(6759), 276–279 (1999).
18. Ockwig, N. W., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Reticular chemistry: Occurrence and taxonomy of nets and grammar for the design of frameworks. *Acc. Chem. Res.* **38**(3), 176–182 (2005).
19. Abdi, J., Banisharif, F. & Khataee, A. Amine-functionalized Zr-MOF/CNTs nanocomposite as an efficient and reusable photocatalyst for removing organic contaminants. *J. Mol. Liquids* **334**, 116971 (2021).
20. Abdi, J. & Abedini, H. MOF-based polymeric nanocomposite beads as an efficient adsorbent for wastewater treatment in batch and continuous systems: Modelling and experiment. *Chem. Eng. J.* **400**, 125862 (2020).
21. Sun, D., Ma, S., Ke, Y., Petersen, T. M. & Zhou, H.-C. Synthesis, characterization, and photoluminescence of isostructural Mn Co, and Zn MOFs having a diamondoid structure with large tetrahedral cages and high thermal stability. *Chem. Commun.* **21**, 2663–2665 (2005).
22. Abdi, J., Vossoughi, M., Mahmoodi, N. M. & Alemzadeh, I. Synthesis of metal-organic framework hybrid nanocomposites based on GO and CNT with high adsorption capacity for dye removal. *Chem. Eng. J.* **326**, 1145–1158 (2017).
23. Liu, J. *et al.* Rational synthesis of a novel 3, 3, 5-c polyhedral metal-organic framework with high thermal stability and hydrogen storage capability. *J. Mater. Chem. A* **4**(30), 11630–11634 (2016).
24. Abdi, J., Vossoughi, M., Mahmoodi, N. M. & Alemzadeh, I. Synthesis of amine-modified zeolitic imidazolate framework-8, ultrasound-assisted dye removal and modeling. *Ultrason. Sonochem.* **39**, 550–564 (2017).
25. Yeh, C.-T. *et al.* Microwave synthesis and gas sorption of calcium and strontium metal–organic frameworks with high thermal stability. *CrystEngComm* **14**(4), 1219–1222 (2012).
26. Kitagawa, S., Noro, S.-I. & Nakamura, T. Pore surface engineering of microporous coordination polymers. *Chem. Commun.* **7**, 701–707 (2006).
27. Wang, Z. & Cohen, S. M. Postsynthetic modification of metal–organic frameworks. *Chem. Soc. Rev.* **38**(5), 1315–1329 (2009).
28. Rowsell, J. L., Spencer, E. C., Eckert, J., Howard, J. A. & Yaghi, O. M. Gas adsorption sites in a large-pore metal-organic framework. *Science* **309**(5739), 1350–1354 (2005).
29. Kitagawa, S., Kitaura, R. & Noro, S. I. Functional porous coordination polymers. *Angew. Chem. Int. Edn.* **43**(18), 2334–2375 (2004).

30. Babarao, R. & Jiang, J. Diffusion and separation of CO_2 and CH_4 in silicalite, C168 schwarzite, and IRMOF-1: A comparative study from molecular dynamics simulation. *Langmuir* **24**(10), 5474–5484 (2008).

31. Langmuir, I. The adsorption of gases on plane surfaces of glass, mica and platinum. *J. Am. Chem. Soc.* **40**(9), 1361–1403 (1918).

32. Yang, Q., Zhong, C. & Chen, J.-F. Computational study of CO_2 storage in metal-organic frameworks. *J. Phys. Chem. C* **112**(5), 1562–1569 (2008).

33. Zhao, Z., Li, Z. & Lin, Y. Adsorption and diffusion of carbon dioxide on metal-organic framework (MOF-5). *Ind. Eng. Chem. Res.* **48**(22), 10015–10020 (2009).

34. Akazawa, M., Hashimoto, K., Katsuhiko, N. & Kaname, Y. Machine learning approach for the prediction of postpartum hemorrhage in vaginal birth. *Sci. Rep.* **11**(1), 22620 (2021).

35. Belmabkhout, Y. *et al.* Natural gas upgrading using a fluorinated MOF with tuned H_2S and CO_2 adsorption selectivity. *Nat. Energy* **3**(12), 1059–1066 (2018).

36. Liang, L. *et al.* Carbon dioxide capture and conversion by an acid-base resistant metal-organic framework. *Nat. Commun.* **8**(1), 1233 (2017).

37. Abdi, J. *et al.* Assessment of competitive dye removal using a reliable method. *J. Environ. Chem. Eng.* **2**(3), 1672–1683 (2014).

38. Mohammadi, F. *et al.* Modelling and optimizing pyrene removal from the soil by phytoremediation using response surface methodology, artificial neural networks, and genetic algorithm. *Chemosphere* **237**, 124486 (2019).

39. Fan, M., Hu, J., Cao, R., Ruan, W. & Wei, X. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. *Chemosphere* **200**, 330–343 (2018).

40. Bac, B. H. *et al.* Performance evaluation of nanotubular halloysites from weathered pegmatites in removing heavy metals from water through novel artificial intelligence-based models and human-based optimization algorithm. *Chemosphere* **282**, 131012 (2021).

41. Davis, L., Handbook of genetic algorithms (1991)

42. Abdi, J., Hadipoor, M., Hadavimoghaddam, F. & Hemmati-Sarapardeh, A. Estimation of tetracycline antibiotic photodegradation from wastewater by heterogeneous metal-organic frameworks photocatalysts. *Chemosphere* **287**, 132135 (2022).

43. Arabloo, M. *et al.* A novel modeling approach to optimize oxygen–steam ratios in coal gasification process. *Fuel* **153**, 1–5 (2015).

44. Chamkalani, A. *et al.* Utilization of support vector machine to calculate gas compressibility factor. *Fluid Phase Equilib.* **358**, 189–202 (2013).

45. Kamari, A., Bahadori, A., Mohammadi, A. H. & Zendehboudi, S. New tools predict monoethylene glycol injection rate for natural gas hydrate inhibition. *J. Loss Prev. Process Ind.* **33**, 222–231 (2015).

46. Kamari, A., Mohammadi, A. H., Bahadori, A. & Zendehboudi, S. Prediction of air specific heat ratios at elevated pressures using a novel modeling approach. *Chem. Eng. Technol.* **37**(12), 2047–2055 (2014).

47. Yildirim, Y. & Bayramoglu, M. Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak. *Chemosphere* **63**(9), 1575–1582 (2006).

48. Dorogush, A. V., Ershov, V., Gulin, A. CatBoost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363 (2018).

49. Breiman, L. Random Forests. *Mach. Learn.* **45**(1), 5–32 (2001).

50. Dashti, A., Bahrololoomi, A., Amirkhani, F. & Mohammadi, A. H. Estimation of CO_2 adsorption in high capacity metal-organic frameworks: Applications to greenhouse gas control. *J. CO2 Util.* **41**, 101256 (2020).

51. Millward, A. R. & Yaghi, O. M. Metal-organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature. *J. Am. Chem. Soc.* **127**(51), 17998–17999 (2005).

52. Herm, Z. R., Swisher, J. A., Smit, B., Krishna, R. & Long, J. R. Metal-organic frameworks as adsorbents for hydrogen purification and precombustion carbon dioxide capture. *J. Am. Chem. Soc.* **133**(15), 5664–5667 (2011).

53. Simmons, J. M., Wu, H., Zhou, W. & Yildirim, T. Carbon capture in metal–organic frameworks—a comparative study. *Energy Environ. Sci.* **4**(6), 2177–2185 (2011).

54. Chen, T. & Guestrin, C. In *Xgboost: A Scalable Tree Boosting System*, *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).

55. Zhang, J. *et al.* A unified intelligent model for estimating the (gas+ n-alkane) interfacial tension based on the eXtreme gradient boosting (XGBoost) trees. *Fuel* **282**, 118783 (2020).

56. Dev, V. A. & Eden, M. R. Gradient boosted decision trees for lithology classification. In *Computer Aided Chemical Engineering* Vol. 47 113-118 (Elsevier, 2019).

57. Ke, G. *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural. Inf. Process. Syst.* **30**, 3146–3154 (2017).

58. Yang, X., Dindoruk, B. & Lu, L. A comparative analysis of bubble point pressure prediction using advanced machine learning algorithms and classical correlations. *J. Pet. Sci. Eng.* **185**, 106598 (2020).

59. Sun, X., Liu, M. & Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Res. Lett.* **32**, 101084 (2020).

60. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. CatBoost: Unbiased boosting with categorical features. arXiv preprint arXiv:1706.09516 (2017).

61. Meng, Q., Ke, G., Wang, T., Chen, W., Ye, Q., Ma, Z.-M. & Liu, T.-Y. A communication-efficient parallel algorithm for decision tree. arXiv preprint arXiv:1611.01276 (2016).

62. Wu, Y. & Misra, S. Intelligent image segmentation for organic-rich shales using random forest, wavelet transform, and hessian matrix. *IEEE Geosci. Remote Sens. Lett.* **17**(7), 1144–1147 (2019).

63. Shaikhina, T. *et al.* Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomed. Signal Process. Control* **52**, 456–462 (2019).

64. Yang, L. *et al.* Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci. Rep.* **10**(1), 1–8 (2020).

65. Shamshirband, S. *et al.* Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases. *Mathematics* **7**(10), 965 (2019).

66. Zhou, Z., Davoudi, E. & Vaferi, B. Monitoring the effect of surface functionalization on the CO_2 capture by graphene oxide/methyl diethanolamine nanofluids. *J. Environ. Chem. Eng.* **9**(5), 106202 (2021).

67. Daryayehsalameh, B., Nabavi, M. & Vaferi, B. Modeling of CO_2 capture ability of [Bmim][BF4] ionic liquid using connectionist smart paradigms. *Environ. Technol. Innov.* **22**, 101484 (2021).

68. Rousseeuw, P. J. & Leroy, A. M. *Robust Regression and Outlier Detection* Vol. 589 (Wiley, New York, 2005).

69. Burgess, C. G., Everett, D. H. & Nuttall, S. Adsorption hysteresis in porous materials. *Pure Appl. Chem.* **61**(11), 1845–1852 (1989).

70. Bourrelly, S. *et al.* Different adsorption behaviors of methane and carbon dioxide in the isotypic nanoporous metal terephthalates MIL-53 and MIL-47. *J. Am. Chem. Soc.* **127**(39), 13519–13521 (2005).

71. Mu, B., Li, F. & Walton, K. S. A novel metal–organic coordination polymer for selective adsorption of CO_2 over CH_4. *Chem. Commun.* **18**, 2493–2495 (2009).

72. Salles, F. *et al.* Multistep N2 breathing in the metal–organic framework Co (1, 4-benzenedipyrazolate). *J. Am. Chem. Soc.* **132**(39), 13782–13788 (2010).

73. Furukawa, H. *et al.* Ultrahigh porosity in metal-organic frameworks. *Science* **329**(5990), 424–428 (2010).

74. Gupta, R. B. *Hydrogen Fuel: Production, Transport, and Storage* (CRC Press, London, 2008).

75. Sircar, S. & Golden, T. Purification of hydrogen by pressure swing adsorption. *Sep. Sci. Technol.* **35**(5), 667–687 (2000).
76. Descamps, C., Bouallou, C. & Kanniche, M. Efficiency of an integrated gasification combined cycle (IGCC) power plant including CO_2 removal. *Energy* **33**(6), 874–881 (2008).
77. Liang, Z., Marshall, M. & Chaffee, A. L. CO_2 adsorption-based separation by metal organic framework (Cu-BTC) versus zeolite (13X). *Energy Fuels* **23**(5), 2785–2789 (2009).
78. Bastin, L. *et al.* A microporous metal—organic framework for separation of CO_2/N_2 and CO_2/CH_4 by fixed-bed adsorption. *J. Phys. Chem. C* **112**(5), 1575–1581 (2008).
79. Murray, L. J. *et al.* Highly-selective and reversible O_2 binding in Cr3 (1, 3, 5-benzenetricarboxylate) 2. *J. Am. Chem. Soc.* **132**(23), 7856–7857 (2010).
80. Ahmadi, M. H. *et al.* An insight into the prediction of TiO_2/water nanofluid viscosity through intelligence schemes. *J. Therm. Anal. Calorim.* **139**(3), 2381–2394 (2020).
81. Lashkarbolooki, M., Vaferi, B. & Mowla, D. Using artificial neural network to predict the pressure drop in a rotating packed bed. *Sep. Sci. Technol.* **47**(16), 2450–2459 (2012).
82. Bemani, A., Baghban, A. & Mohammadi, A. H. An insight into the modeling of sulfur content of sour gases in supercritical region. *J. Pet. Sci. Eng.* **184**, 106459 (2020).

Author contributions

J.A. Writing-Original Draft, Data curation; Formal analysis, F.H. Writing-Review & Editing, Validation, Methodology, M.H. Writing-Review & Editing, Validation, Data curation, A.H.-S.: Writing-Review & Editing, Methodology, Validation, Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-04168-w.

Correspondence and requests for materials should be addressed to A.H.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.