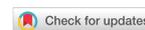


Research Article

Open Access



# Machine learning for prediction of CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O selective adsorption and separation in metal-zeolites

Ya-Ting Gu, Yu-Ming Gu\*, Qiantu Tao, Xinzhu Wang, Qin Zhu, Jing Ma\*

Key Laboratory of Mesoscopic Chemistry of Ministry of Education, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, Jiangsu, China.

\*Correspondence to: Dr. Yuming Gu, Prof. Jing Ma, Key Laboratory of Mesoscopic Chemistry of Ministry of Education, School of Chemistry and Chemical Engineering, Nanjing University, Xianlin Avenue, Nanjing 210023, Jiangsu, China. E-mail: guyuming19930701@163.com; majing@nju.edu.cn

**How to cite this article:** Gu YT, Gu YM, Tao Q, Wang X, Zhu Q, Ma J. Machine learning for prediction of CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O selective adsorption and separation in metal-zeolites. *J Mater Inf* 2023;3:19. <https://dx.doi.org/10.20517/jmi.2023.25>

**Received:** 23 Jun 2023 **First Decision:** 28 Jul 2023 **Revised:** 19 Aug 2023 **Accepted:** 1 Sep 2023 **Published:** 6 Sep 2023

**Academic Editors:** Xingjun Liu, Liming Yang **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

## Abstract

Carbon dioxide (CO<sub>2</sub>) capture, utilization, and storage technologies are crucial in reducing global warming and producing various high-value chemicals and fuels. It is challenging to effectively separate the molecules of CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O, whose kinetic diameters are close to each other. Although zeolites have garnered considerable attention in gas separation, the huge chemical space of metal-doped zeolites (metal-zeolites) coming from the combination of different metal active sites, topology, and Si/Al ratios poses a difficulty in finding an optimal material for selectively trapping CO<sub>2</sub>. In this study, we build machine learning (ML) models to predict the selective adsorption of CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O on metal-zeolites through the regulation of electrostatic polarization interaction. The stability of 208 metal-zeolites encompassing five distinct topological structures is estimated through the formation energy ( $E_f$ ), indicating the potential accessibility in the experiment for most of the studied systems, especially for Sc-, Y-, and Zr-zeolites. Adsorption of CO<sub>2</sub> on metal sites has two possible configurations: linear vs. bent CO<sub>2</sub>, depending on different embedded metals. The concerted binding of CO<sub>2</sub> with both carbon and oxygen atoms on the metal center leads to the bent geometry and larger binding energies on metal-zeolites (Zr-, Nb-, Mo-zeolites). Accessible descriptors associated with the zeolites, adsorbates, and metals are selected to train the adsorption strength index ( $I$ ), showing good performance [mean absolute error (MAE) = 0.04,  $R^2$  = 0.88]. The predicted adsorption selectivity is in agreement with the experimental systems (Co-, Zn-, Cu-, Fe-SSZ-13). It is found that medium-pore-sized zeolites [pore limiting diameter (PLD) = ~7 Å] anchored with Zr, Nb, or Mo are promising materials for the CO<sub>2</sub> adsorption and separation. The proposed ML scheme may also be applicable to give a fast prediction of CO<sub>2</sub> adsorption and separation ability in other porous metal-organic frameworks or amorphous materials.

**Keywords:** Metal-zeolites, machine learning, separation, selective adsorption



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



## INTRODUCTION

The global climate is undergoing rapid and significant changes as a result of the increasing emissions of carbon dioxide (CO<sub>2</sub>) into the atmosphere<sup>[1]</sup>. To control and balance CO<sub>2</sub> emissions to meet global energy demand, scientists have devised several strategies for carbon capture and separation, including membrane separation, post-combustion, and amine-containing solvent absorption<sup>[2]</sup>. Alternatively, there has been a notable upsurge in both academic and industrial interest in CO<sub>2</sub> selective separation that relies on solid materials<sup>[3,4]</sup>. Nonetheless, due to the similar kinetic diameter (*d*) values, separation of CO<sub>2</sub> (*d* = 3.3 Å) from nitrogen (*d* = 3.6 Å) and water (*d* = 2.7 Å) is a significant challenge. In general, CO<sub>2</sub> exhibits a higher affinity for adsorption on metal atoms compared to N<sub>2</sub> due to the larger electric quadrupole moment<sup>[5]</sup>. The transition metal can interact with CO<sub>2</sub> through  $\pi$ -complexation by forming both  $\sigma$  bonds and  $\pi$  backdonation. The adsorption of polar water molecules is unavoidable in some cases since the empty or partially occupied *d* orbitals of transition metals can accept electron density from lone-pair electrons of H<sub>2</sub>O to form  $\sigma$  bonds. The effects of moisture on CO<sub>2</sub> adsorption are detrimental because of the competition of adsorption sites<sup>[6]</sup>. In this work, we will characterize these three molecules by their different polarizability values (CO<sub>2</sub>:  $26.3 \times 10^{-25}$  cm<sup>3</sup>; N<sub>2</sub>:  $17.7 \times 10^{-25}$  cm<sup>3</sup>; H<sub>2</sub>O:  $14.8 \times 10^{-25}$  cm<sup>3</sup>, [Supplementary Table 1](#)).

A wide array of materials, including activated carbons, metal-organic frameworks (MOFs), shale rocks, and zeolites, have been applied for the selective adsorption capabilities towards CO<sub>2</sub><sup>[7-9]</sup>. Among them, zeolites, a kind of inorganic porous materials composed of TO<sub>4</sub> (T = Si, Al, *etc.*) tetrahedra, have been widely used in the fields of catalysis, and adsorption/separation owing to large surface areas, various channel sizes, low costs, and diverse active sites<sup>[10-12]</sup>. It has been reported that N<sub>2</sub> molecules are favorable to be adsorbed in zeolites with medium-sized pores (4~8 Å) due to the electrostatic interactions between N<sub>2</sub> and local Si-O-Si or Al-O-Si units<sup>[13]</sup>. The introduction of metal cations into the framework also provides a good strategy for zeolite modification to regulate the electrostatic polarization interaction<sup>[14]</sup>. For example, the inert N<sub>2</sub> molecules can be activated in Ti-, Co-, and Nb-zeolites to produce the NH<sub>3</sub> during the nitrogen reduction reaction<sup>[15]</sup>. The high electronegativity of metal atoms in zeolite facilitates the strong attraction of electrons from cations, leading to CO<sub>2</sub> conversion to methanol or methane<sup>[12,16,17]</sup>. Density functional theory (DFT) and *ab initio* molecular dynamics (MD) simulations have been employed to investigate the adsorption and separation behaviors of CO<sub>2</sub>, N<sub>2</sub>, H<sub>2</sub>O, CH<sub>4</sub>, C<sub>1</sub>-C<sub>4</sub> alcohols, and alkenes in zeolites<sup>[18-20]</sup>. For example, the competitive adsorption of H<sub>2</sub>O and CO<sub>2</sub> in zeolite 13X (FAU) was simulated through DFT calculations<sup>[20]</sup>.

The combination of metal cations and zeolite topologies gives rise to an extensive chemical space, posing challenges for exploring the promising material candidates effectively. Recently, an efficient high-throughput screening strategy has gained prominence for identifying potential materials for CO<sub>2</sub> adsorption separation from a vast array of zeolite structures<sup>[18,21]</sup>. Some experimental parameters, such as formation pressure, temperature, salinity, *etc.*, have been selected for predicting the shale wettability for CO<sub>2</sub> capture and storage<sup>[22]</sup>. It was reported by our group that N<sub>2</sub> adsorption isotherms can be simulated from binding energies (*E<sub>b</sub>*) and the Langmuir model, where *E<sub>b</sub>* can be predicted through three features, i.e., the geometry of pore limiting diameter (PLD), the size of adsorption cavities (*V<sub>eff</sub>*), and the local geometric distortion (*R<sub>DLS</sub>*)<sup>[13]</sup>. The N<sub>2</sub> isotherm has also been utilized as a feature descriptor for porosity by training convolutional neural networks to predict the CO<sub>2</sub>/N<sub>2</sub> separation performance in porous carbons<sup>[23]</sup>. Textural properties, such as void fraction (VF), density, and PLD, were chosen as MOF fingerprints to train the model and then predict the adsorption uptake<sup>[24]</sup>. When introducing metal cations into the zeolite framework, descriptors related to the metal properties, such as average ionization energy difference ( $\delta$ IE) and the global electronegativity difference ( $\delta\chi$ ) between the metal-zeolites and intermediates, have been devised as descriptors for charge transfer (CT) prediction without relying on DFT calculations<sup>[16]</sup>.

The goal of this work is to identify metal zeolites that exhibit high selectivity for CO<sub>2</sub> adsorption among the similar-sized N<sub>2</sub> and H<sub>2</sub>O by examining the distinct adsorption strength behaviors of different adsorbates. The formation energies of 208 metal-zeolites with different Si/Al ratios (X), encompassing five distinct topological structures (FER, MFI, MOR, CHA, and FAU), are calculated by DFT [Figure 1]. Zeolite adsorption datasets are further constructed by calculating the  $E_b$  between adsorbates (CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O) and metal-zeolites. Bent configuration of the absorbed CO<sub>2</sub> is found in metal-zeolites consisting of Sc, Y, and Zr atoms, leading to a larger binding capability for CO<sub>2</sub>. With these DFT-derived training sets of 624 binding energies, the feature learning (FL) method is employed to predict the adsorption selectivity of adsorbates in metal-zeolites with easily accessible features. The zeolites with Zr, Nb, and Mo atoms anchored on frameworks with medium PLD, such as CHA, are predicted to be good candidates for the separation of CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O mixtures. The proposed machine learning (ML) scheme holds the potential to give significant aid in the systematic design and discovery of materials for CO<sub>2</sub> capture and separation.

## MATERIALS AND METHODS

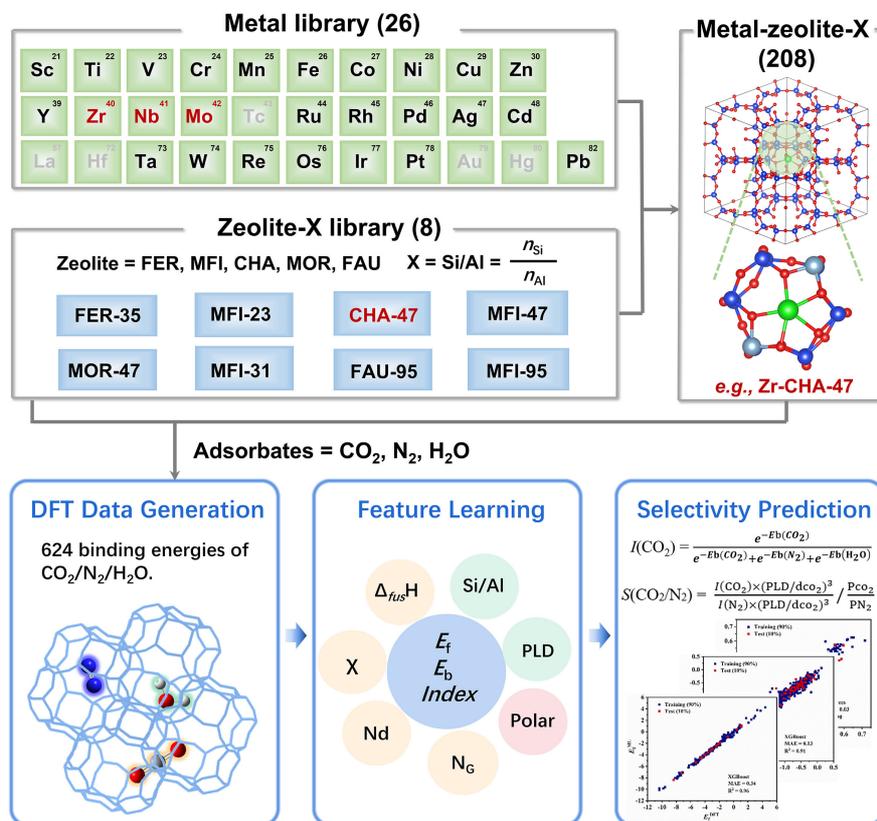
### Structural model

In the present work, the selected zeolites are CHA (2 × 2 × 2 supercell), MFI (1 × 1 × 1 primitive cell), MOR (1 × 1 × 2 supercell), FER (1 × 1 × 2 supercell), and FAU (1 × 1 × 1 primitive cell) in Figure 2A<sup>[25]</sup>. CHA is characterized by pores connected by 8-membered rings (MR), with a pore size measuring 3.8 × 3.8 Å<sup>2</sup> [Supplementary Figure 1A]. MFI possesses sinusoidal 10-membered channels and straight 10-membered channels, with dimensions of 5.3 × 5.6 Å<sup>2</sup> and 5.1 × 5.5 Å<sup>2</sup>, respectively [Supplementary Figure 1B]. MOR exhibits three distinct types of cavities: the main channel formed by a 12-MR (6.5 × 7.0 Å<sup>2</sup>) with the largest void [Supplementary Figure 1C], the side pocket formed by an 8-MR (3.4 × 4.8 Å<sup>2</sup>), and the side channel (2.4 × 5.7 Å<sup>2</sup>). FER is a zeolite with a medium-pore-sized structure [Supplementary Figure 1D] consisting of two perpendicular intersecting channels: an 8-MR channel (3.4 × 5.4 Å<sup>2</sup>) and a 10-MR channel (4.3 × 5.5 Å<sup>2</sup>). FAU possesses 12-MR channel openings (7.4 × 7.4 Å<sup>2</sup>) and a cavity with a diameter of approximately 12 Å [Supplementary Figure 1E]. To study the influence of the Si/Al ratios, models were established by replacing one, two, three, and four silicon atoms with aluminum atoms, respectively [Figure 2B]. The incorporation of transition metal atoms into the zeolites is necessary to maintain electroneutrality after the substitution of Si atoms with Al atoms.

### Construction of datasets

In this study, the present zeolite adsorption dataset contains DFT data, being grouped as a training set and test set randomly. The experimental data collected from the literature are chosen as an external test set. Datasets are composed of 26 kinds of transition metal atoms (3d, 4d, 5d transition metals and a main group atom Pb) sitting in the zeolites with five kinds of topological structures (FER, MFI, MOR, CHA, and FAU), whose PLDs are in the range from 6.31 to 11.24 Å. CHA, FER, MOR, and MFI are zeolites with medium PLDs (6.31~7.37 Å). The pore in FAU is much larger, possessing a cavity with a diameter of approximately 12 Å. So far, the number of International Zeolite Association (IZA) structures has reached ~260. The selected five kinds of topological structures attracted intensive interest [Supplementary Figure 2]. Among the pieces of literature, reports of MOR take up 11.7%, followed by MFI, accounting for 9.8%. The publications of CHA, FER, and FAU have proportions of 9.0%, 6.9%, and 1.8%, respectively. Three adsorbates, namely N<sub>2</sub>, CO<sub>2</sub>, and H<sub>2</sub>O, in 208 metal-zeolites encompassing five distinct topological structures, constitute 624 binding energies as datasets, which are divided into a training set and a test set in a ratio of 90:10.

To estimate the relative stability of the metal-zeolites, the  $E_f$  of metal-zeolite is calculated by the following equation.



**Figure 1.** Flowchart illustrating the machine learning approach for predicting formation energies and adsorption selectivity of CO<sub>2</sub> in metal-zeolites. CO<sub>2</sub>: Carbon dioxide; DFT: density functional theory; PLD: pore limiting diameter.

$$E_f = E_{M-Z} - E_Z - u_M \quad (1)$$

where  $E_{M-Z}$  is the total energy of metal-zeolite,  $E_Z$  is the energy of zeolite, and  $u_M$  is the energy of the metal atom in its most stable bulk structure. Generally, the  $E_f$  with more negative values indicates a higher thermodynamical stability.

The  $E_b$  of adsorbates adsorbed on zeolite is defined as follows.

$$E_b = E_{\text{adsorbates/M-Z}} - E_{M-Z} - E_{\text{adsorbates}} \quad (2)$$

where  $E_{\text{adsorbates/M-Z}}$  is the total energy of the zeolite complex with adsorbates adsorbed on the zeolite, and  $E_{M-Z}$  and  $E_{\text{adsorbates}}$  stand for the energies of metal-zeolite and free adsorbates guest, respectively.

The Ideal Adsorbed Solution Theory (IAST) selectivity is calculated by using the pyIAST code, which is widely used to predict estimated selectivity in experiments<sup>[26]</sup>. Herein, to predict the selectivity of adsorbates, we quantify the adsorption selectivity from the calculated adsorption strength. The adsorption strength index ( $I$ ) of adsorbates in zeolites is then defined in equations (3)-(5).

$$I(\text{CO}_2) = \frac{e^{-E_b(\text{CO}_2)}}{e^{-E_b(\text{CO}_2)} + e^{-E_b(\text{N}_2)} + e^{-E_b(\text{H}_2\text{O})}} \quad (3)$$

$$I(\text{N}_2) = \frac{e^{-E_b(\text{N}_2)}}{e^{-E_b(\text{CO}_2)} + e^{-E_b(\text{N}_2)} + e^{-E_b(\text{H}_2\text{O})}} \quad (4)$$

$$I(\text{H}_2\text{O}) = \frac{e^{-E_b(\text{H}_2\text{O})}}{e^{-E_b(\text{CO}_2)} + e^{-E_b(\text{N}_2)} + e^{-E_b(\text{H}_2\text{O})}} \quad (5)$$

The adsorption selectivity ( $S$ ) of adsorbates in zeolites is then defined as equations (6) and (7).

$$S(\text{CO}_2/\text{N}_2) = \frac{I(\text{CO}_2) \times (\text{PLD}/d_{\text{CO}_2})^3}{I(\text{N}_2) \times (\text{PLD}/d_{\text{N}_2})^3} \frac{P_{\text{CO}_2}}{P_{\text{N}_2}} \quad (6)$$

$$S(\text{CO}_2/\text{H}_2\text{O}) = \frac{I(\text{CO}_2) \times (\text{PLD}/d_{\text{CO}_2})^3}{I(\text{H}_2\text{O}) \times (\text{PLD}/d_{\text{H}_2\text{O}})^3} \frac{P_{\text{CO}_2}}{P_{\text{H}_2\text{O}}} \quad (7)$$

where  $\text{PLD}/d_{\text{CO}_2}$  represents the freedom degree of the adsorbates in the cage, and  $P_{\text{CO}_2}$ ,  $P_{\text{N}_2}$ , and  $P_{\text{H}_2\text{O}}$  mean the partial pressure of components  $\text{CO}_2$ ,  $\text{N}_2$ , and  $\text{H}_2\text{O}$ , respectively.

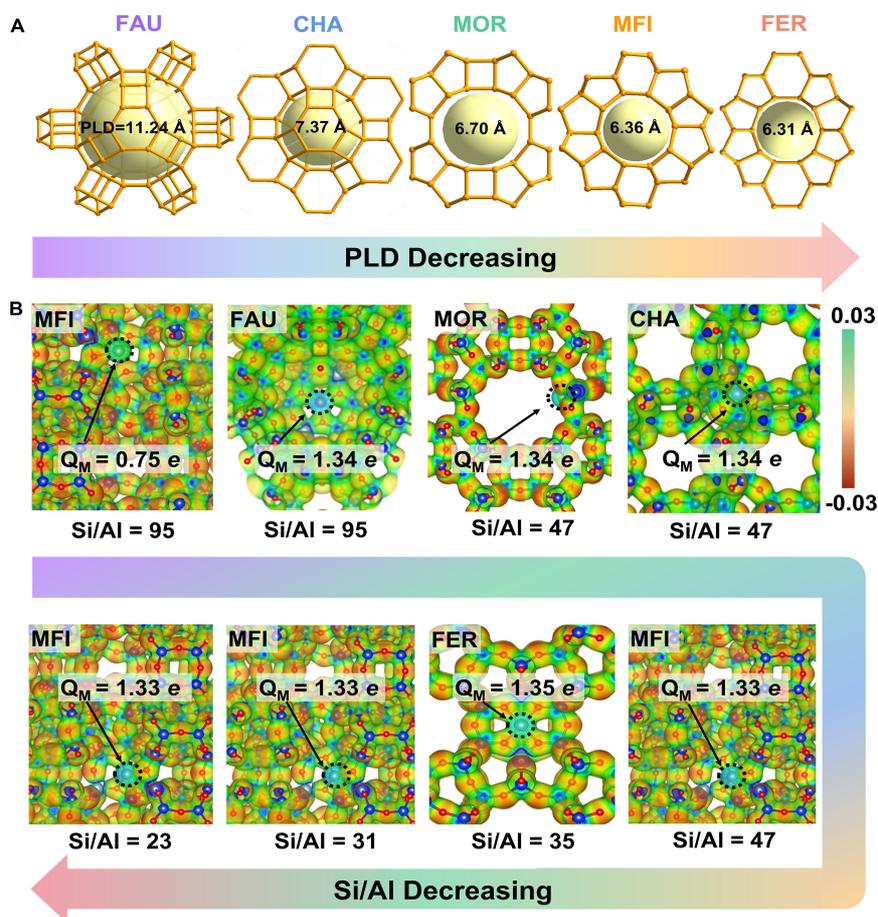
### DFT calculations

For DFT calculations, the CP2K package<sup>[27]</sup> within the Quickstep module was utilized. The Perdew-Burke-Ernzerhof (PBE) functional<sup>[28]</sup> was selected as the exchange-correlation functional. Grimme's D3 corrections were used to account for van der Waals interactions. The core electron interaction was treated using the analytical dual-space-type pseudopotential described by Geodecker-Teter-Hutter (GTH)<sup>[29,30]</sup>. A molecularly optimized (MOLOPT) double z valence plus polarization (DZVP) basis set<sup>[31]</sup> was utilized to treat the valence electrons. The Broyden-Fletcher-Goldarb-Shanno (BFGS) minimization algorithm<sup>[32,33]</sup> with hessian modeling was performed. The optimization process continued until the forces converged to a threshold of  $6 \times 10^{-4}$  Hartree per bohr. The self-consistent field (SCF) energy convergence criteria were set to  $1 \times 10^{-6}$  Hartree. A plane wave expansion for the charge density was utilized with an energy cutoff of 600 Ry.

In order to compare the results from different computational codes,  $E_b$  of adsorbates on MFI-31 zeolites was also performed using the Vienna *Ab initio* Simulation Package (VASP)<sup>[34]</sup>, which gave similar results to CP2K [Supplementary Figure 3]. The projector-augmented wave (PAW) method<sup>[35]</sup> was adopted to describe the core-valence electron interaction. Perdew, Burke, and Ernzerhof (PBE) form of the generalized gradient approximation (GGA) was utilized to account for the exchange-correlation among electrons<sup>[36]</sup>. A DFT-D3 method with Grimme with zero-damping was applied to treat the van der Waals interaction<sup>[37]</sup>. The plane-wave cutoff energy of 450 eV was chosen to ensure accuracy. The ions were relaxed until the maximum force on each atom was below  $4 \times 10^{-2}$  eV·Å<sup>-1</sup>, and the total energy was considered converged when it reached  $10^{-5}$  eV within the electronic self-consistent loop. The first Brillouin zones were sampled with  $2 \times 2 \times 3$   $\Gamma$ -centered k-point meshes.

The charge density difference (called CT) was illustrated with Visualization for Electronic and Structural Analysis (VESTA) software by the following equation:

$$\text{CT} = q_{\text{adsorbates}} - q_{\text{adsorbates}/\text{M-Z}} \quad (8)$$



**Figure 2.** (A) Five topological structures of zeolites; (B) Electrostatic potential diagram of metal-zeolites with different Si/Al ratios. PLD: Pore limiting diameter.

Where  $q_{\text{adsorbates}}$  represents the charge of the free adsorbates guest, and  $q_{\text{adsorbates/M-Z}}$  represents the charge of adsorbates in metal-zeolite.

### Machine learning

All ML algorithms were implemented using the open-source code Scikit-learn<sup>[38]</sup> and PyTorch package in the Python3 environment. A total of ten algorithms were employed for the prediction of  $E_p$ ,  $E_b$ , and  $I$ . The selected ML models include Extreme gradient boosting regression (XGBoost)<sup>[39]</sup>, ExtraTrees<sup>[40]</sup>, Gradient boosting regression (GBR)<sup>[41]</sup>, Decision tree (DT)<sup>[42]</sup>, k-nearest neighbor (kNN)<sup>[43]</sup>, Least absolute shrinkage and selection operator (LASSO)<sup>[44]</sup>, Linear ridge<sup>[45]</sup>, Multiple Linear regression (MLR)<sup>[46]</sup>, Artificial neural network (ANN)<sup>[47]</sup>, and Support vector regression (SVR)<sup>[48]</sup>. As mentioned above, the collected data obtained from DFT calculations were randomly shuffled and divided into a training set and a test set in a ratio of 90:10. A normalization pre-processing step was applied to the ML models in training and prediction. Three evaluation metrics were selected to assess the prediction errors: the mean absolute error (MAE), the root-mean-square error (RMSE), and the coefficient of determination values ( $R^2$ ), as described in equations (9)-(11).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (11)$$

In the above equation, the  $y_i$ ,  $\hat{y}_i$ ,  $\bar{y}_i$  means  $i$ -th predicted value by ML, the DFT calculation results, and the average value of DFT calculated results, respectively.

The DFT or experimental data are important to train a ML model in a proper way<sup>[49]</sup>. In our previous works, nitrogen and CO<sub>2</sub> adsorption and activation data have been collected and compiled into the ZA dataset<sup>[50]</sup>, which is expanded to have a total of 234 formation energies, 201,053 binding energies, and 1,082 reaction energies in this work.

## RESULTS AND DISCUSSION

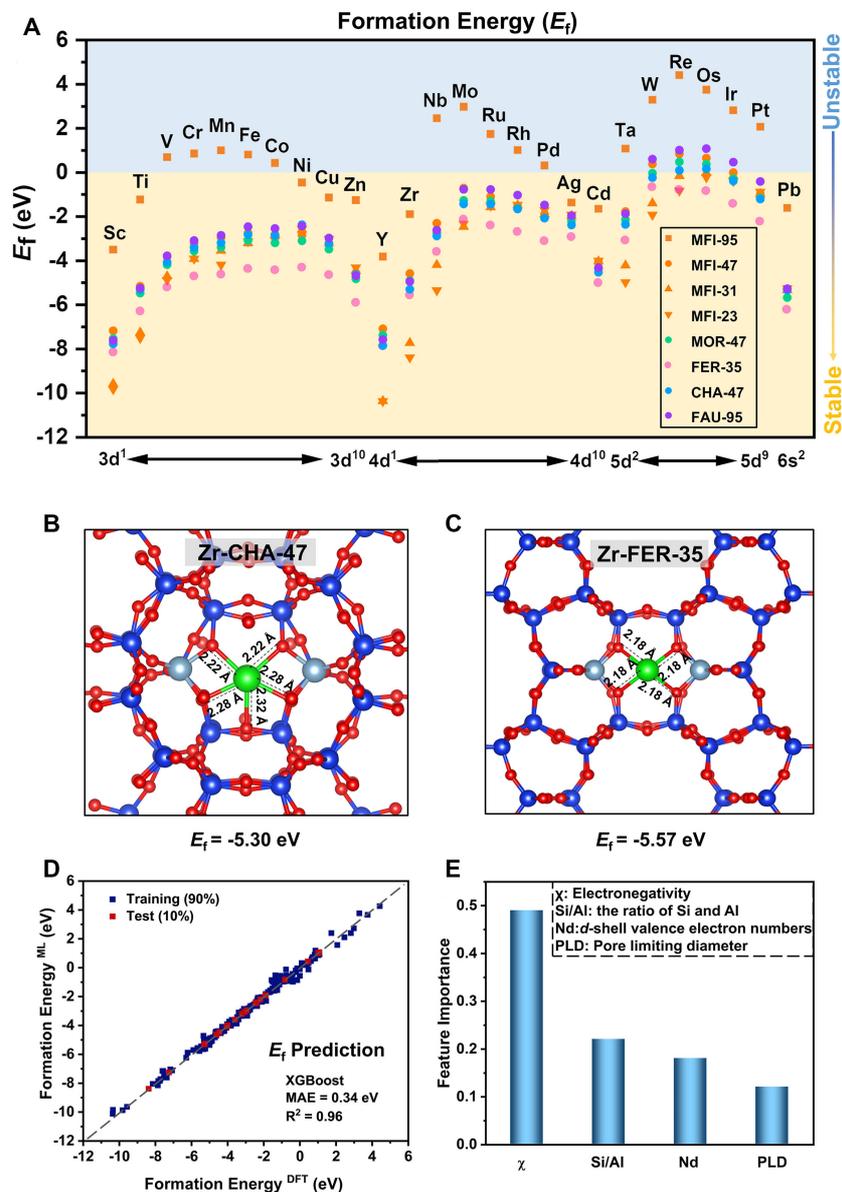
### Formation energy prediction

The pore size of zeolites has a great influence on the diffusion and separation of adsorbates. In this study, 26 transition metal atoms are involved in the zeolites with five kinds of topological structures (FER, MFI, MOR, CHA, FAU), whose PLDs are in a range from 6.31 to 11.24 Å [Figure 2A].

It is crucial to predict whether the constructed metal-zeolites could be synthesized in future experiments. The  $E_f$  of the metal-zeolite framework is applied to evaluate the thermodynamic stability of the structure. A criterion that  $E_f < 0$  eV is set for evaluating the stability of a candidate. Here, for MFI-type structures, different Si/Al ratios (Si/Al = 95, 47, 31, 23) are systematically investigated to test the effect of various Si/Al ratios. By comparing four different Si/Al ratios in the MFI structure [Figure 3A], the increase in the Al sites (2Al/3Al/4Al) leads to higher thermodynamic stability (with  $E_f$  less than 0 eV) than the MFI containing only one Al site in simulation cell with periodic boundary condition (PBC). In the MFI-95 structure (Si/Al ratio = 95), the metal cation interacts with only one negatively charged site, making it difficult to achieve efficient electrostatic stabilization, resulting in unstable configurations. It should be noted that better stabilization occurs when the metal cation directly contacts the four oxygen atoms connected with the Al atoms. To save computational cost, two Al sites in each PBC cell are selected for metal-zeolites with the other four topological structures in the following study.

The distribution of Al atoms can be affected by the conditions of synthesis in the presence of monovalent cations such as Na<sup>[51,52]</sup>. Experimental investigations demonstrated the fact that Al atoms are predominantly paired in 5-MR and 6-MR in high-silica zeolites such as MOR and MFI<sup>[10,53]</sup>. Less than 30% of Al atoms could exist in an isolated state within a ring. Statistical analysis of Al distribution in FER, MOR, and MFI zeolites with typical Si/Al ratios also showed the prevalence of short-range Al pairs in 5-MR and 6-MR<sup>[11,18]</sup>. In such an energetically favorable configuration, the metal cation coordinated to four adjacent oxygen atoms.

The embedded 26 kinds of metal atoms include transition metals from  $3d$  (Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, and Zn),  $4d$  (Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, and Cd), to  $5d$  (Ta, W, Re, Os, Ir, and Pt) and a main group metal atom (Pb). The larger size of transition metal atoms in the third period (metal atomic radius,  $R = 135\sim 175$  pm) compared to the size of the zeolite ring results in a less efficient contact between the metal



**Figure 3.** (A) Calculated  $E_f$  of metal atom embedded in the zeolite; (B) The optimized geometric configuration of (B) Zr-CHA-47; (C) Zr-FER-35; (D) ML model construction for  $E_f$  prediction by XGBoost algorithm *via* four features; (E) Feature importance analysis. DFT: Density functional theory; MAE: mean absolute error; ML: machine learning.

cations and the oxygen atoms connected to the Al atoms, which contributes to a relatively unfavorable  $E_f$ . As depicted in Figure 3A, there is a volcano-type curve relationship in the thermodynamic stability of transition metal atoms in the same period. This relationship is characterized by an initial increase and then a decrease with the number of  $d$ -shell valence electrons. Transition metals with either partially or fully filled  $d$  orbitals exhibit more negative  $E_f$  values, reaching as low as -10 eV for Sc-MFI-23 and Y-MFI-23. It means that the  $d^1$  metal-zeolites ( $M = \text{Sc}, \text{Y}$ ),  $d^2$  metal-zeolites ( $M = \text{Ti}, \text{Zr}$ ), and  $d^{10}$  metal-zeolites ( $M = \text{Cu}, \text{Zn}, \text{Pd}, \text{Ag}, \text{and Cd}$ ) exhibit stronger stability. From left to right in the same period, the oxidation state of the metal increases with the number of valence electrons in the  $d$  orbitals increases. When the number of electrons in

the  $d$  orbitals reaches 5, the highest oxidation state becomes increasingly unstable. Subsequently, the oxidation state gradually decreases, leading to an increase in the stability of metal-zeolites, especially for 2Al substituted zeolites. Thermodynamically stable Zr-CHA-47 [Figure 3B] and Zr-FER-35 [Figure 3C] have relatively short average coordination distances between Zr and O atoms of 2.26 and 2.18 Å, respectively. It is worth noting that FER exhibits relatively higher stability compared to other zeolites with Si/Al = 95, 47, 31, and 23.

ML models are further built to predict  $E_f$  to screen out the synthesizable zeolites. Various descriptors of zeolite structures offer detailed insights into their complex nature<sup>[54,55]</sup>. For feature selection, some available structural descriptors, such as the PLD and  $R_{DLS}$ , can be obtained from zeolite structure databases<sup>[22]</sup>. Features related to metal atoms, such as atomic radius (R), atomic number (AN), enthalpy of fusion ( $\Delta_{fus}H$ ),  $d$ -shell valence electron number (Nd), number of the group in the periodic table ( $N_G$ ),  $\chi$ , and first ionization energy (IE), are also utilized to predict  $E_f$ . Some experimental parameters, such as temperature and pressure, were chosen by other groups<sup>[22]</sup>. Here, we used  $\Delta_{fus}H$ , which is correlated with temperature and pressure. The  $\Delta_{fus}H$  refers to the amount of heat energy that is absorbed at constant pressure to convert a specific mass of solid metal from a solid state to a liquid state, which can be found in the CRC Handbook of Chemistry and Physics<sup>[5]</sup>. A list of selected features is presented in Supplementary Table 2. Some selected data are illustrated in Supplementary Table 3. As shown in Supplementary Table 4, the descriptors selected in this paper have a wide range of distribution.

Upon reviewing all the descriptors discussed above, which encompass the structural channels and electronic properties of zeolites and the physical and chemical properties of metals, the Pearson correlation coefficient matrix of the ten descriptors is depicted in Supplementary Figure 4.  $E_f$  exhibits positive correlations with descriptors such as Nd,  $\Delta_{fus}H$ ,  $\chi$ ,  $N_G$ , AN, and IE. It can be inferred that an increase of the value of the aforementioned descriptors implies a higher stability. Some of these features are redundant since some descriptors are correlated with each other based on the Pearson correlation coefficient matrix. For instance, Nd shows significant correlations with  $N_G$  and IE. To eliminate redundancy and enhance the prediction performance of the model, we remove strongly correlated features, resulting in a non-redundant subset of descriptors. The number of characteristic descriptors for prediction of  $E_f$  is gradually reduced from the initial ten features to four. Ultimately, the descriptors of  $\chi$  and Nd are employed to describe the electronic structure features of the metal sites. The Si/Al ratio and PLD are utilized to describe the local electrostatic environment and the pore characteristics of zeolites, respectively, for the prediction of  $E_f$ .

A training set comprising 208 formation energies is utilized to construct ten different ML models, namely XGBoost, ExtraTrees, GBR, kNN, DT, ANN, MLR, SVR, LASSO, and linear ridge. The performance of these models is evaluated using the RMSE, MAE, and  $R^2$  metrics, as described in equations (9)-(11). Grid search is employed to improve the appropriate parameters for each model through 10-fold cross-validation. Both the XGBoost and GBR models exhibit impressive performance, with MAE of 0.34 eV,  $R^2$  of 0.96 [Figure 3D], and MAE of 0.39 eV and  $R^2$  of 0.95 [Supplementary Figure 5], respectively. It should be mentioned that ANN was applied to predict CO<sub>2</sub> adsorption separation on shales<sup>[22]</sup>. In the present work, the XGBoost algorithm shows better predictive performance. XGBoost belongs to an ensemble algorithm that constructs an immensely scalable tree-boosting framework by optimizing loss functions with a minimal amount of resources<sup>[13]</sup>. The parameters of the XGBoost algorithm for predicting the  $E_f$  are listed in Supplementary Table 5. The importance of the four features,  $\chi$ , Si/Al, Nd, and PLD, is determined to be 0.49, 0.22, 0.18, and 0.12, respectively [Figure 3E]. This indicates that the  $\chi$  of the metal significantly influences the stability prediction. When the four features are further reduced, there is an obvious increase in the fitting error and a decrease in the coefficient of determination, which suggests that the combination

of these four features forms the minimum set required to achieve accurate predictions for  $E_r$ .

### Adsorption strength of adsorbates

Zeolites with different topologies and active sites exhibit different adsorption capacities. CHA has been demonstrated to display a strong affinity towards adsorbates due to the presence of a potent local electric field<sup>[53]</sup>. The adsorption capacity of zeolites with adsorbates is influenced by several factors, including (i) topology structures; (ii) the Si/Al ratio; (iii) the type and distribution of extraframework cations; and (iv) the properties of adsorbates.

Firstly, the size of the pores affects the diffusion properties of guest molecules. Zeolites with excessively small pore sizes cannot accommodate adsorbate molecules, while those with excessively large pore sizes may result in weak interactions between the framework and adsorbates. Zeolites with a PLD of 4~8 Å exhibits strong N<sub>2</sub> adsorption<sup>[22]</sup>. Again, the PLD of zeolite is selected in FL of adsorbate-zeolite binding energies and adsorption selectivity. As mentioned above, the similar size of CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O make it difficult to separate the adsorbates only by pore size, so the impact of electrostatic polarization interaction needs to be further addressed.

Secondly, the adsorption ability increases as the Si/Al ratio decreases. Zeolites with lower Si/Al ratios exhibit stronger adsorption abilities, particularly at low pressures<sup>[56-58]</sup>. The role of Si/Al ratios is rationalized by modulation of the local electric field around the binding sites. The electrostatic potential of the typical Zn-zeolites is shown in [Figure 2B](#). The redder the region, the lower the electrostatic potential is. Metals (green areas) have a higher electrostatic potential, indicating that the metal could become the active site for adsorption. As the ratio of silicon to aluminum decreases [[Figure 2B](#)], the charge of the metal center,  $q_M$ , shows an increasing trend (from 0.75 to 1.35  $e$ ). Hence, Si/Al is expected to be an important feature in describing the electrostatic interaction between the metal-zeolite and adsorbates.

Thirdly, the incorporation of metal cations into zeolites can generate exposed positive charges, which could serve as strong adsorption sites for CO<sub>2</sub>. In numerous zeolites, exposed positive charges coming from extra-framework cations are in balance of the negative charge arising from the substitution of Si<sup>4+</sup> cations with Al<sup>3+</sup>. The presence of extra-framework cations in zeolites significantly affects adsorbate adsorption. Zeolites with larger cations, such as K cations, often exhibit enhanced adsorption affinity. The adsorption heat of CO<sub>2</sub> on alkali metal-FAU zeolites follows the sequence: Li-FAU < Na-FAU < K-FAU<sup>[59]</sup>. Similar trends have been observed for X zeolites<sup>[60]</sup>. It was found that the group number ( $N_G$ ) is directly associated with  $\chi$ , where the adsorption energy aligns with the trend of  $\chi$ <sup>[61]</sup>. To describe the active metal site,  $N_d$ ,  $N_G$ , and  $\Delta_{fus}H$  are chosen as critical features.

Lastly, the different adsorbates, such as CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O, exhibit different adsorption strengths in zeolites. The different surface species and locations of adsorbates in zeolites have attracted intensive interest<sup>[62]</sup>. The adsorption can be physisorption or chemisorption. The chemisorbed species include bidentate carbonates, monodentate carbonates and carboxylates, *etc.*<sup>[62,63]</sup>. Electrostatic parameters, such as polarizability, dipole moments, and quadrupole moments, indicate the sensitiveness of adsorbates in electrostatic interaction. In other words, the local electrostatic field or the polarity is crucial in zeolite adsorbents<sup>[64]</sup>. Given a larger quadrupole moment and greater polarizability of CO<sub>2</sub>, it exhibits a stronger electrostatic interaction than N<sub>2</sub> and H<sub>2</sub>O. Thus, the polarizability (called Polar in short) of adsorbates is considered as an important feature descriptor. Additionally, the size effect (kinetic diameter,  $d$ ) and electronic properties (number of electrons in the outermost shell,  $N_p$ ) of the adsorbates are also used to differentiate their characteristics.

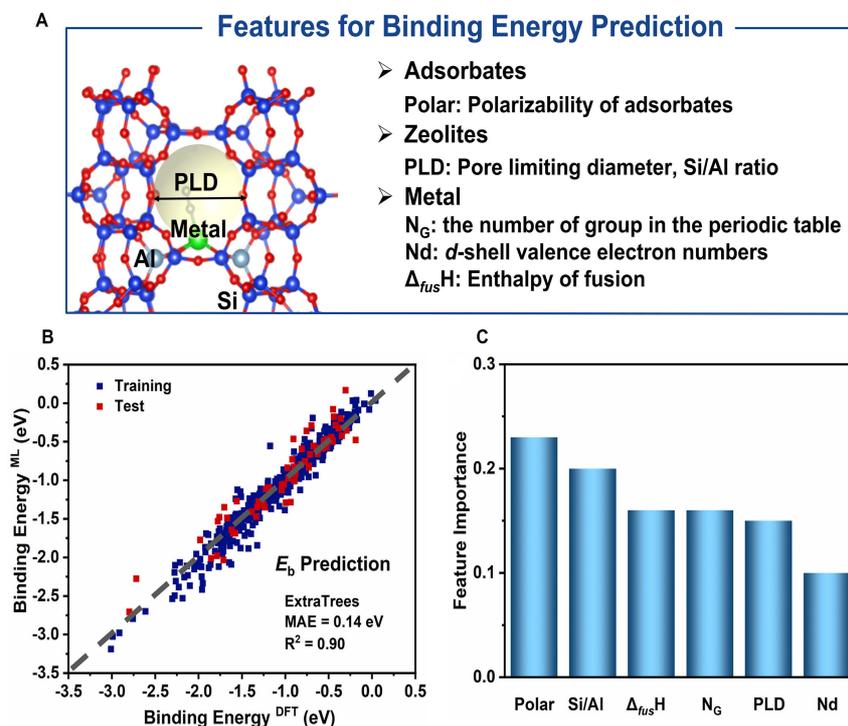
To sum up, a total of 13 characteristics are employed to describe  $E_b$  between the adsorbates and metal-zeolites [Figure 4A]. The Pearson's correlation analysis between zeolite properties, metal properties, and adsorbate properties is presented in Supplementary Figure 4B. The negative value of the  $E_b$  between metal-zeolite and adsorbates indicates a strong adsorption strength. Supplementary Figures 6-8 illustrate the  $E_b$  of  $\text{CO}_2$ ,  $\text{N}_2$ , and  $\text{H}_2\text{O}$  in various zeolite structures, including MFI with 1 Al site (MFI-95), 2 Al sites (MFI-47), 3 Al sites (MFI-31), and 4 Al sites (MFI-23) and MOR, FER, CHA, and FAU topological structures with 2 Al sites (MOR-47, FER-35, CHA-47, FAU-95, respectively).

For MFI zeolites, as the Si/Al ratio decreases from 95 to 23, the  $E_b$  of adsorbates decreases by approximately 1 eV overall, indicating the significant impact of electrostatic polarization interaction between the adsorbates and the zeolites. The features of Polar, Np, and d correlate with each other. Therefore, only Polar is retained as the feature descriptor for adsorbates. Si/Al and PLD are still kept as features. For metal atoms,  $N_G$ ,  $\Delta_{fus}H$ , and Nd are selected as descriptors for predicting the  $E_b$  between adsorbates and zeolites. Based on these six descriptors (Nd,  $N_G$ ,  $\Delta_{fus}H$ , Polar, Si/Al, and PLD), the  $E_b$  can be effectively predicted [Figure 4]. To improve accuracy, a 10-fold cross-validation is performed using grid search to choose appropriate hyperparameters for ten different algorithms. Among these algorithms, ExtraTrees achieves the best prediction performance with MAE of 0.14 eV and  $R^2$  of 0.90 [Figure 4B]. The GBR (MAE = 0.15 eV,  $R^2$  = 0.88) and XGBoost (MAE = 0.16 eV,  $R^2$  = 0.86) algorithms also provide reliable prediction results [Supplementary Figure 5B]. However, linear models, such as MLR, Linear ridge, and LASSO, give poor prediction performance. The importance of the six features is shown in Figure 4C. It is evident that the Polar of the adsorbates plays a pivotal role, as indicated by its feature importance.

The adsorption and separation of adsorbates can be achieved by utilizing the shape-selective effects of zeolites and the unique properties of different metal sites. As shown in Figure 5A,  $\text{CO}_2$  molecules display the bent or linear configurations, depending on different metal sites. The pre-transition metals (Sc, Ti, Y, Zr, etc.) have more empty  $d$  orbitals, which are easy to form the  $\pi$ -backbonding between metal and CO. The C and O atoms can be adsorbed on metal (M) sites side by side with the synergistic mode, leading to stronger adsorption strengths for the bent  $\text{CO}_2$ , particularly in the case of CHA zeolites [Figure 5B]. Linear  $\text{CO}_2$  can be adsorbed on the  $d^8 \sim d^{10}$  metals via the head-to-head coordination between the metal and O atoms in physisorption mode. The distance between adsorbates ( $\text{CO}_2$ ,  $\text{N}_2$ , and  $\text{H}_2\text{O}$ ) and metal-zeolites are summarized in Figure 5C. Compared to the linear configuration of  $\text{CO}_2$  adsorbed in metal-zeolites, the distance between metal and O atoms of bent  $\text{CO}_2$  ( $d_{M-O}$ ) is shorter, indicating a stronger electrostatic polarization interaction.  $\text{CO}_2$  shows the bent configuration when the charge of metal is larger than 1.6  $e$ . The correlation between the charge of the metal and the binding energy indicates that electrostatic interaction is the dominant factor of the adsorption [Figure 5D]. The  $\text{CO}_2$  molecule with a bent configuration will also be a starting point for further  $\text{CO}_2$  activation in metal-zeolites.

### Selective separation for adsorbates

In order to quantify adsorption selectivity of  $\text{CO}_2$ ,  $\text{N}_2$ , and  $\text{H}_2\text{O}$ , we define the adsorption selectivity of adsorbates,  $S$ , in equations (6) and (7). The selectivity,  $S$ , is determined by the adsorption strength index ( $I$ ) with the adsorption energy written in the exponential function. Higher values of the  $I$  indicate stronger adsorption strength. The ML models are further constructed for the  $I$  prediction of  $\text{CO}_2$ ,  $\text{N}_2$ , and  $\text{H}_2\text{O}$  in 208 metal-zeolite systems. The Pearson's correlation analysis between zeolite, metal, and adsorbate properties is presented in Supplementary Figure 4C. The above-mentioned six accessible features (Polar,  $\Delta_{fus}H$ , Si/Al, PLD,  $N_G$ , and Nd) are also applicable in selectivity prediction. The ExtraTrees model demonstrates good performance in predicting the  $I$ , with  $R^2$  = 0.88 and MAE = 0.04 [Figure 6A]. The feature importance is shown in Figure 6B. The parameters of the ExtraTrees algorithm for predicting the adsorption strength index are listed in Supplementary Table 6. The GBR (MAE = 0.04,  $R^2$  = 0.83) model exhibited comparable



**Figure 4.** (A) Six-feature scheme for  $E_b$  prediction on metal-zeolites; (B) ML model construction for  $E_b$  prediction by ExtraTrees algorithm via six features; (C) Feature importance analysis.  $E_b$ : Binding energies; MAE: mean absolute error; ML: machine learning.

prediction performance to the Extratrees model [Supplementary Figure 5C].

In order to further validate the applicability of our prediction model, some external tests of separation selectivity for experimental reported systems<sup>[12]</sup> are selected. The IAST selectivity is usually calculated by using the pyIAST code<sup>[26]</sup>. In this work, we adopted the adsorption strength and amount of the adsorbates to represent the adsorption quantity. The separation quantity is determined by multiplying  $I$  with the volume ratio of the zeolite cage diameter to the adsorbate diameter  $(PLD/d)^3$ . The ratio of the separation quantity to the ratio of the partial pressures of the adsorbates is defined as the separation selectivity [equations (6) and (7)]. It was reported that transition metal cation-exchanged SSZ-13 (CHA) zeolites have been employed for CO<sub>2</sub> capture and separation from N<sub>2</sub><sup>[12]</sup>. As listed in Table 1, the selectivity of the reported metal-zeolite exhibits a trend of Co-SSZ-13 > Zn-SSZ-13 > Cu-SSZ-13 > Fe-SSZ-13, which is qualitatively reproduced by our predicted selectivity parameters<sup>[10]</sup>.

Subsequently, the ML model is extended to zeolites with other topologies in order to identify highly selective adsorbents. The selectivity heat map is presented in Supplementary Figure 9. FER zeolites are highly selective for CO<sub>2</sub> with low N<sub>2</sub> uptake, probably being a potential adsorption separation porous material for CO<sub>2</sub>/N<sub>2</sub>. When CO<sub>2</sub> competes with N<sub>2</sub> for adsorption, the FER zeolite with PLD = 6.31 Å has a stronger binding strength with CO<sub>2</sub> than N<sub>2</sub> among the selected zeolites. FER is characterized by a medium-pore-sized zeolite that features two perpendicular intersecting channels: an 8-MR channel (3.4 × 5.4 Å<sup>2</sup>) and a 10-MR channel (4.3 × 5.5 Å<sup>2</sup>). The accessibility of adsorbates to the narrow-pore zeolite cages depends on the extent of window occupation. Hence, adsorbates enter and diffuse through the 10-MR window, while the adsorption occurs in the 8-MR cage easily, which matches the size of the adsorbates well. This observation aligns with our chemical intuitions since a pore with a small PLD restricts the accessibility of

**Table 1. Application of ML models for experimentally reported systems (CO<sub>2</sub>/N<sub>2</sub>: 15/85, 298 K, and 101.3 kPa)**

Metal-zeolite	Si/Al	Gas mixture	Separation selectivity (ML in this work)	IAST selectivity (Exp. in Ref. <sup>[12]</sup> )
Co-SSZ-13	9.92	CO <sub>2</sub> /N <sub>2</sub>	8.5	40
Zn-SSZ-13	10.62	CO <sub>2</sub> /N <sub>2</sub>	8.3	35
Cu-SSZ-13	10.31	CO <sub>2</sub> /N <sub>2</sub>	8.0	30
Fe-SSZ-13	17.58	CO <sub>2</sub> /N <sub>2</sub>	7.6	27

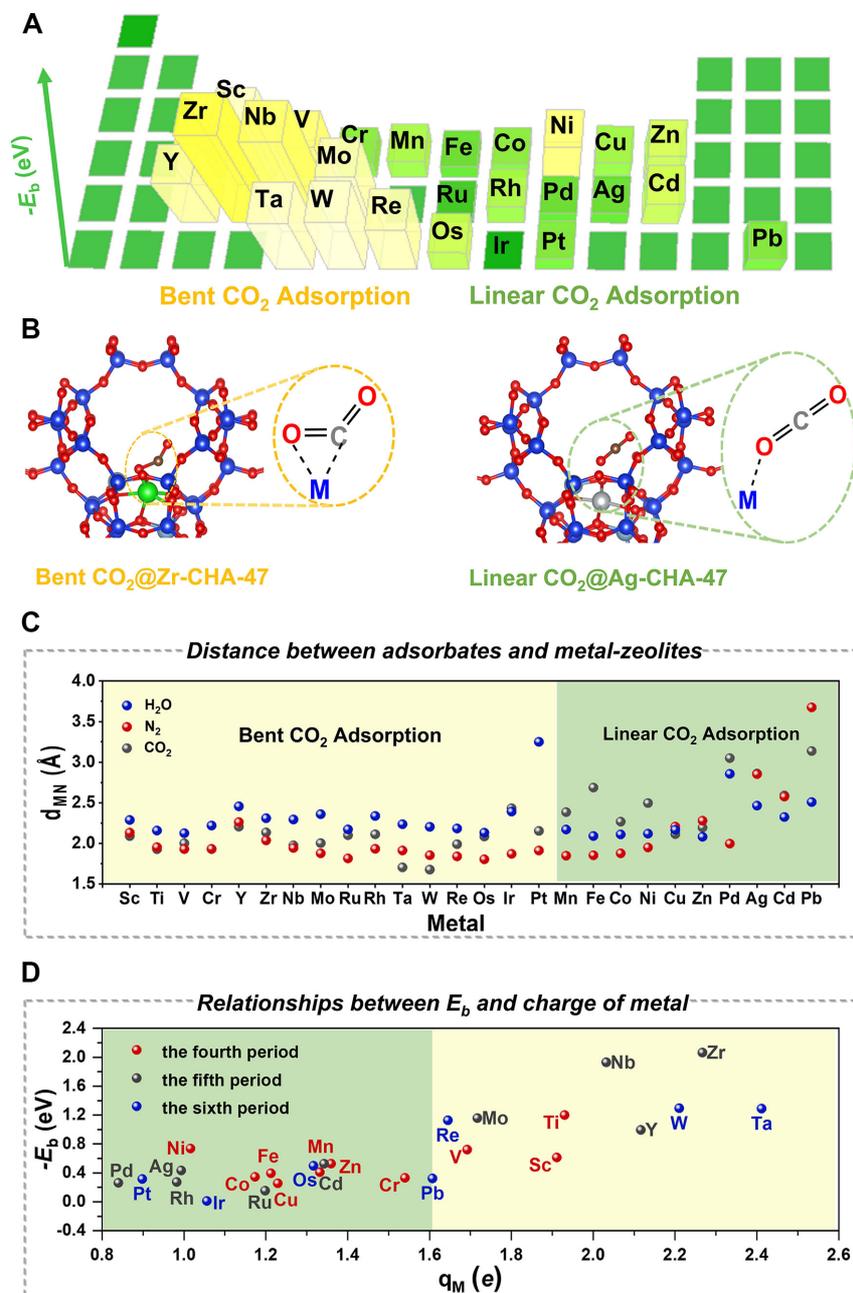
CO<sub>2</sub>: Carbon dioxide; IAST: Ideal Adsorbed Solution Theory; ML: machine learning.

CO<sub>2</sub> molecules, and excessively large pores lead to decreased binding strength. Thus, the optimal PLD of the zeolite is about 7 Å for achieving high CO<sub>2</sub> adsorption selectivity relative to N<sub>2</sub> [Figure 6C] and H<sub>2</sub>O [Figure 6D].

The adverse effects of moisture on CO<sub>2</sub> adsorption, resulting from competition for adsorption sites, have been recognized widely<sup>[65,66]</sup>. Due to the formation of hydrogen bonds between water and the oxygen atoms in the framework of zeolites, the adsorption of H<sub>2</sub>O is usually stronger than that of N<sub>2</sub> and CO<sub>2</sub> in metal-zeolites. The separation between CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O can be influenced by both kinetics and thermodynamics. Kinetically, H<sub>2</sub>O possesses a slightly smaller effective kinetic diameter ( $d = 2.7$  Å) compared to N<sub>2</sub> ( $d = 3.6$  Å) and CO<sub>2</sub> ( $d = 3.3$  Å), facilitating H<sub>2</sub>O diffusion [Supplementary Table 1]. The diffusion of adsorbates into and out of zeolites is constrained by the dimensions of the zeolite apertures. These kinetic effects are particularly prominent for zeolites with effective small and medium pore openings. However, it is difficult for silicon zeolites to achieve good separation of CO<sub>2</sub> from the mixed gases with similar kinetic diameters. Thermodynamically, metal-zeolites may be a good choice to regulate the electrostatic interaction, given the different polarizability of the adsorbates. In some cases, H<sub>2</sub>O displays a higher affinity for adsorption relative to N<sub>2</sub> and CO<sub>2</sub> due to its dipole moment, resulting in stronger interactions with the adsorbents<sup>[13,14]</sup>. DFT calculations have revealed that H<sub>2</sub>O, with its pronounced dipole moment, exhibits a much higher affinity for adsorption sites involving exposed metal cations compared to CO<sub>2</sub><sup>[67]</sup>. Experimental studies using binary adsorption measurements for CO<sub>2</sub> and H<sub>2</sub>O mixtures in LTA zeolites have also demonstrated a substantial decrease in CO<sub>2</sub> adsorption<sup>[68]</sup>.

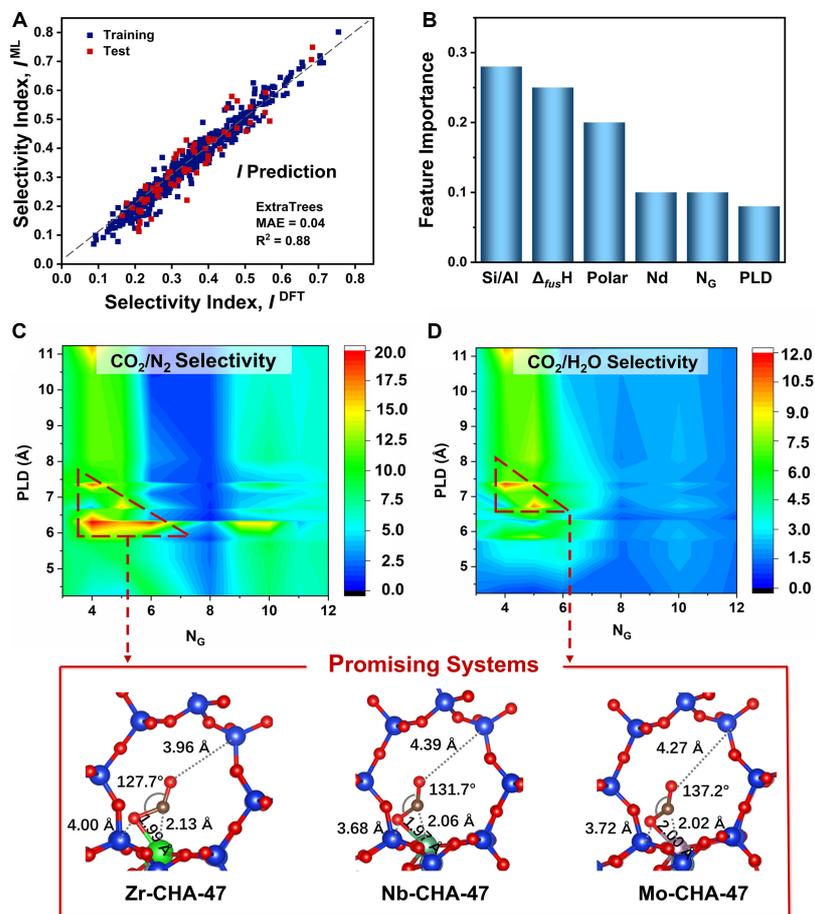
Minimizing H<sub>2</sub>O adsorption is crucial for reducing the energy required for regeneration. The design of adsorbents with strong CO<sub>2</sub> binding sites and hydrophobic surfaces presents a huge challenge in the development of the next generation of CO<sub>2</sub> adsorption separation materials. It has been suggested that the Cu-SSZ-13 zeolite may retain its CO<sub>2</sub> binding energy even in the presence of H<sub>2</sub>O, showing stronger CO<sub>2</sub> adsorption<sup>[53]</sup>. SSZ-13 is a kind of stable zeolite with chabazite topology consisting of 4-, 6-, and 8-MR. The corner-sharing Al/SiO<sub>4</sub> tetrahedra form double 6-MR prisms, which are further connected to create a supercage ( $6.7 \times 10$  Å<sup>2</sup>) with 8-MR windows (approximately 3.8 Å). The small pore aperture size of 8MR allows for size exclusion of adsorbates, and the pore size can be further adjusted with the incorporation of different metal cations at the aperture<sup>[16]</sup>.

Interestingly, Zr-, Nb-, and Mo-CHA-47 exhibit both large  $E_f$  and strong adsorption of CO<sub>2</sub> but lower N<sub>2</sub> and H<sub>2</sub>O uptake [Figure 7], ingeniously addressing the long-standing issue of strong water adsorption interference. CO<sub>2</sub> displayed the bent configuration on Zr-, Nb-, and Mo-CHA-47 with the short coordination distance ( $d_{M-O} = \sim 2.0$  Å,  $d_{M-C} = \sim 2.1$  Å). The Mo-CHA-47 exhibits the best CO<sub>2</sub> selectivity and the highest CT value in the Radar chart. The CT between adsorbates and metal-zeolites shows a similar trend with the binding strength, indicating that the regulation of electrostatic polarization plays an important role in the separation selectivity [Supplementary Figure 10]. In general, the magnitude of the

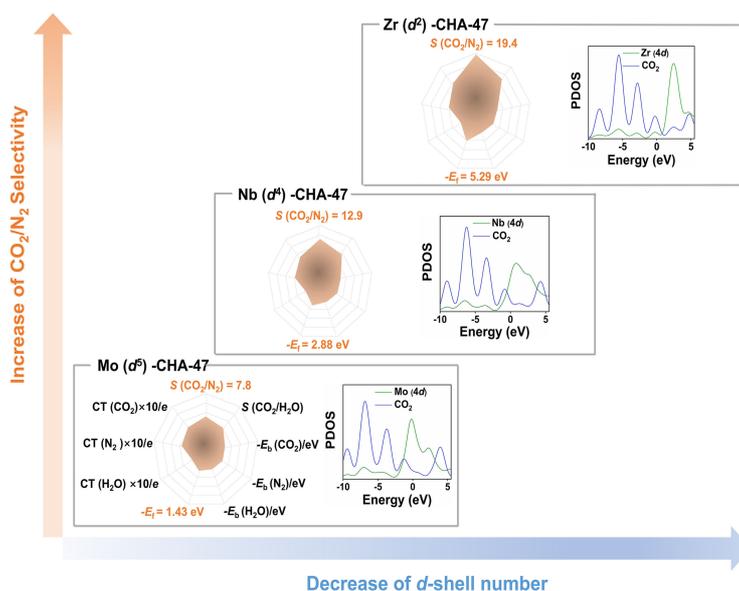


**Figure 5.** (A) The distribution of CO<sub>2</sub> adsorption with bent and linear configurations in the periodic table; (B) the structures of bent and linear CO<sub>2</sub> on metal-zeolites; (C) distance between adsorbates and metal-zeolites; (D) the relationships between the binding energy and charge of metals. CO<sub>2</sub>: Carbon dioxide;  $E_b$ : binding energies.

isosteric enthalpy is within the range of 30–60 kJ/mol, indicating a moderate CO<sub>2</sub> binding strength compared to physisorption and chemisorption of CO<sub>2</sub> [69,70], which suggests that the zeolite adsorbents can be easily regenerated. CO<sub>2</sub>, with its bent configuration, is strongly adsorbed and exhibits difficulty in diffusion. The projected density of states (PDOS) in Figure 7 also indicates the significant contribution of CO<sub>2</sub> molecules and *d* orbitals of metals around the Fermi surface. In summary, Zr, Nb, and Mo, anchored on medium-pore-size zeolites (PLD = ~7 Å), show promise for the separation of CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O, making them potential candidates for adsorption and separation candidates.



**Figure 6.** (A) ML model construction for *I* prediction by ExtraTrees algorithm via six features; (B) Feature importance analysis. The adsorption selectivity heat map of (C) CO<sub>2</sub>/N<sub>2</sub>; (D) CO<sub>2</sub>/H<sub>2</sub>O and the promising structures of Zr-CHA-47, Nb-CHA-47, and Mo-CHA-47, respectively. CO<sub>2</sub>: Carbon dioxide; DFT: density functional theory; ML: machine learning; PLD: pore limiting diameter.



**Figure 7.** The Radar chart for comprehensive evaluation of Zr-CHA-47, Nb-CHA-47, and Mo-CHA-47 on CO<sub>2</sub>/N<sub>2</sub>/H<sub>2</sub>O separation and PDOS of CO<sub>2</sub> adsorbed on Zr-CHA-47, Nb-CHA-47, and Mo-CHA-47, respectively. CO<sub>2</sub>: Carbon dioxide; CT: charge transfer; E<sub>b</sub>: binding energies; PDOS: projected density of states.

## CONCLUSIONS

In conclusion, we present a ML model to find suitable candidate zeolites that can effectively separate the CO<sub>2</sub> among the mixed gas (N<sub>2</sub> and H<sub>2</sub>O). The regulation of electrostatic polarization interaction could improve the selectivity of CO<sub>2</sub> in zeolites despite the similar kinetic diameter of CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O. It is found that medium-pore-sized zeolites (PLD = ~7 Å) containing transition metals (Zr, Nb, or Mo) are predicted to possess high CO<sub>2</sub> adsorption separation selectivity based on the binding strength between the framework and adsorbates. More CT occurs between metals (Zr, Nb, or Mo) and CO<sub>2</sub> than N<sub>2</sub> and H<sub>2</sub>O, leading to stronger adsorption affinity. The ML models of XGBoost and ExtraTrees algorithms provide reliable predictions for  $E_p$ ,  $E_b$ , and  $I$  with easily available descriptors, including metal center, electronic, and geometry descriptors. The ML models can be transferred to predict experimentally reported systems (Co-, Zn-, Cu-, Fe-SSZ-13). The present work provides an efficient tool to give a quick prediction of CO<sub>2</sub> selectivity to accelerate the discovery of separation materials. The generalization ability of current ML models could be further improved by enlarging the dataset. It is also expected that the model could be extended to predict adsorption selectivity of other adsorbates in other porous MOFs and solid materials. The chemisorption of CO<sub>2</sub> on metal-zeolites will be explored further to screen out the high-performing catalysts in future work. A comprehensive atomistic view on adsorption of the adsorbates on various materials provides a useful strategy for designing target-oriented catalysts for capture, separation, and even catalysis with the combination of theoretical calculations and ML.

## DECLARATIONS

### Authors' contributions

Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Gu YT, Gu YM, Ma J

Performed machine learning modeling and quantum chemical calculations: Tao Q

Performed data acquisition and provided technical support: Wang X, Zhu Q

### Availability of data and materials

Some data are given in Supporting Information for illustration. The data are available on the web at <http://106.15.196.160:5656/> upon the request of a license.

### Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (grant No. 22033004) and the Jiangsu Provincial Excellent Postdoctoral Program (grant No. 2023ZB655). We are grateful to the High Performance Computing Centre of Nanjing University for providing the IBM Blade cluster system.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2023.

## REFERENCES

1. Veltý A, Corma A. Advanced zeolite and ordered mesoporous silica-based catalysts for the conversion of CO<sub>2</sub> to chemicals and fuels. *Chem Soc Rev* 2023;52:1773-946. DOI PubMed
2. Sifat NS, Haseli Y. A critical review of CO<sub>2</sub> capture technologies and prospects for clean power generation. *Energies* 2019;12:4143. DOI
3. Usman M, Iqbal N, Noor T, et al. Advanced strategies in metal-organic frameworks for CO<sub>2</sub> capture and separation. *Chem Rec* 2022;22:e202100230. DOI
4. Madejski P, Chmiel K, Subramanian N, Kuš T. Methods and techniques for CO<sub>2</sub> capture: review of potential solutions and applications in modern energy technologies. *Energies* 2022;15:887. DOI
5. Tomaszewski R. Citations to chemical resources in scholarly articles: CRC handbook of chemistry and physics and the merck index. *Scientometrics* 2017;112:1865-79. DOI
6. Wang H, Yin Y, Bai J, Wang S. Multi-factor study of the effects of a trace amount of water vapor on low concentration CO<sub>2</sub> capture by 5A zeolite particles†. *RSC Adv* 2020;10:6503-11. DOI PubMed PMC
7. Jia T, Gu Y, Li F. Progress and potential of metal-organic frameworks (MOFs) for gas storage and separation: a review. *J Environ Chem Eng* 2022;10:108300. DOI
8. Zhao M, Yang Y, Gu XS. MOF based CO<sub>2</sub> capture: adsorption and membrane separation. *Inorg Chem Commun* 2023;152:110722. DOI
9. Osman AI, Hefny M, Abdel Maksoud MIA, Elgarahy AM, Rooney DW. Recent advances in carbon capture storage and utilisation technologies: a review. *Environ Chem Lett* 2021;19:797-849. DOI
10. Cheung O, Liu Q, Bacsik Z, Hedin N. Silicoaluminophosphates as CO<sub>2</sub> sorbents. *Microporous Mesoporous Mater* 2012;156:90-6. DOI
11. Li Y, Chen H, Wang C, et al. Achieving highly selective CO<sub>2</sub> adsorption on SAPO-35 zeolites by template-modulating the framework silicon content†. *Chem Sci* 2022;13:5687-92. DOI PubMed PMC
12. Sun M, Gu Q, Hanif A, Wang T, Shang J. Transition metal cation-exchanged SSZ-13 zeolites for CO<sub>2</sub> capture and separation from N<sub>2</sub>. *J Chem Eng* 2019;370:1450-8. DOI
13. Gu Y, Liu Z, Yu C, et al. Zeolite adsorption isotherms predicted by pore channel and local environment descriptors: feature learning on DFT binding strength. *J Phys Chem C* 2020;124:9314-28. DOI
14. Fu D, Park Y, Davis ME. Zinc containing small-pore zeolites for capture of low concentration carbon dioxide. *Angew Chem Int Ed Engl* 2022;61:e202112916. DOI PubMed
15. Gu Y, Zhu Q, Liu Z, et al. Nitrogen reduction reaction energy and pathways in metal-zeolites: deep learning and explainable machine learning with local acidity and hydrogen bonding features. *J Mater Chem A* 2022;10:14976-88. DOI
16. Zhu Q, Gu Y, Liang X, Wang X, Ma J. A machine learning model to predict CO<sub>2</sub> reduction reactivity and products transferred from metal-zeolites. *ACS Catal* 2022;12:12336-48. DOI
17. Choi HJ, Min JG, Ahn SH, et al. Framework flexibility-driven CO<sub>2</sub> adsorption on a zeolite†. *Mater Horiz* 2020;7:1528-32. DOI
18. Moliner M, Román-Leshkov Y, Corma A. Machine learning applied to zeolite synthesis: the missing link for realizing high-throughput discovery. *Acc Chem Res* 2019;52:2971-80. DOI PubMed
19. Wu LM, Li XY, Xie FM, Zhong DL, Englezos P, Yan J. Minireview of hydrate-based CO<sub>2</sub> separation from a CO<sub>2</sub>/CH<sub>4</sub> gas mixture: progress and outlook. *Energy Fuels* 2022;36:10478-88. DOI
20. Yuan Y, You H, Ricardez-sandoval L. Recent advances on first-principles modeling for the design of materials in CO<sub>2</sub> capture technologies. *Chin J Chem Eng* 2019;27:1554-65. DOI
21. Schwalbe-Koda D, Kwon S, Paris C, et al. A priori control of zeolite phase competition and intergrowth with high-throughput simulations. *Science* 2021;374:308-15. DOI
22. Zhang H, Thanh HV, Rahimi M, et al. Improving predictions of shale wettability using advanced machine learning techniques and nature-inspired methods: implications for carbon capture utilization and storage. *Sci Total Environ* 2023;877:162944. DOI
23. Wang S, Li Y, Dai S, Jiang DE. Prediction by convolutional neural networks of CO<sub>2</sub>/N<sub>2</sub> selectivity in porous carbons from N<sub>2</sub> adsorption isotherm at 77 K. *Angew Chem Int Ed Engl* 2020;59:19645-8. DOI PubMed
24. Srinivasu K, Natarajan S, Ghanty TK. Confinement directed adsorption of noble gases (Xe/Kr) in MFM-300(M)-based metal-organic framework materials. *J Phys Chem C* 2019;123:27531-41. DOI
25. Baerlocher C, McCusker LB. Database of zeolite structures. Available from: <http://www.iza-structure.org/databases/>. [Last accessed on

- 4 Sep 2023].
26. Cai X, Gharagheizi F, Bingel LW, Shade D, Walton KS, Sholl DS. A collection of more than 900 gas mixture adsorption experiments in porous materials from literature meta-analysis. *Ind Eng Chem Res* 2021;60:639-51. DOI
  27. Vandevondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput Phys Commun* 2005;167:103-28. DOI
  28. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010;132:154104. DOI PubMed
  29. Goedecker S, Teter M, Hutter J. Separable dual-space Gaussian pseudopotentials. *Phys Rev B Condens Matter* 1996;54:1703-10. DOI PubMed
  30. Hartwigsen C, Goedecker S, Hutter J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys Rev B* 1998;58:3641-62. DOI
  31. Vandevondele J, Hutter J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J Chem Phys* 2007;127:114105. DOI PubMed
  32. Broyden CG. The convergence of single-rank quasi-Newton methods. *Math Comput* 1970;24:365-82. Available from: <https://www.ams.org/journals/mcom/1970-24-110/S0025-5718-1970-0279993-0/S0025-5718-1970-0279993-0.pdf>. [Last accessed on 4 Sep 2023]
  33. Shanno DF. Conditioning of quasi-Newton methods for function minimization. *Math Comput* 1970;24:647-56. Available from: <https://www.ams.org/journals/mcom/1970-24-111/S0025-5718-1970-0274029-X/S0025-5718-1970-0274029-X.pdf>. [Last accessed on 4 Sep 2023]
  34. Kresse G, Furthmüller J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater Sci* 1996;6:15-50. DOI
  35. Blöchl PE. Projector augmented-wave method. *Phys Rev B Condens Matter* 1994;50:17953-79. DOI PubMed
  36. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett* 1996;77:3865-8. DOI PubMed
  37. Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 2011;32:1456-65. DOI PubMed
  38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825-30. Available from: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. [Last accessed on 4 Sep 2023]
  39. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 1998;10:1895-923. DOI PubMed
  40. Guo Y, Du L, Wei D, Li C. Robust SAR automatic target recognition via adversarial learning. *IEEE J Sel Top Appl Earth Observations Remote Sensing* 2021;14:716-29. DOI
  41. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38:367-78. DOI
  42. Rodríguez JJ, Quintana G, Bustillo A, Ciurana J. A decision-making tool based on decision trees for roughness prediction in face milling. *Int J Comput Integr Manuf* 2017;30:943-57. DOI
  43. Patrick EA, Fischer FP. A generalized k-nearest neighbor rule. *Inf Control* 1970;16:128-52. DOI
  44. D'Angelo GM, Rao D, Gu CC. Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies. *BMC Proc* 2009;3 Suppl 7:S62. DOI PubMed PMC
  45. Hoerl RW. Ridge regression: a historical context. *Technometrics* 2020;62:420-5. DOI
  46. Schumann TEW. XXVI. The principles of a mechanical method for calculating regression equations and multiple correlation coefficients and for the solution of simultaneous linear equations. *Lond, Edinb, Dublin Philos Mag J Sci* 1940;29:258-73. DOI
  47. Abiodun OI, Jantan A, Omolara AE, et al. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access* 2019;7:158820-46. DOI
  48. Seko A, Maekawa T, Tsuda K, Tanaka I. Machine learning with systematic density-functional theory calculations: application to melting temperatures of single- and binary-component solids. *Phys Rev B* 2014;89:054303. DOI
  49. Li Y, Yu J. New stories of zeolite structures: their descriptions, determinations, predictions, and evaluations. *Chem Rev* 2014;114:7268-316. DOI
  50. Gu Y, Liu Z, Yu X, et al. Zeolite adsorption database. *J Phys Chem C* 2020;124:9314-28. Available from: <http://106.15.196.160:5656/index/en>. [Last accessed on 4 Sep 2023]
  51. Jeffroy M, Nieto-draghi C, Boutin A. New molecular simulation method to determine both aluminum and cation location in cationic zeolites. *Chem Mater* 2017;29:513-23. DOI
  52. Li Y, Guo W, Fan W, et al. A DFT study on the distributions of Al and Brønsted acid sites in zeolite MCM-22. *J Mol Catal A Chem* 2011;338:24-32. DOI
  53. Hudson MR, Queen WL, Mason JA, Fickel DW, Lobo RF, Brown CM. Unconventional, highly selective CO<sub>2</sub> adsorption in zeolite SSZ-13. *J Am Chem Soc* 2012;134:1970-3. DOI PubMed
  54. Anderson R, Biong A, Gómez-Gualdrón DA. Adsorption isotherm predictions for multiple molecules in MOFs using the same deep learning model. *J Chem Theory Comput* 2020;16:1271-83. DOI
  55. Caro-Ortiz S, Zuidema E, Rigutto M, Dubbeldam D, Vlucht TJH. Competitive adsorption of Xylenes at chemical equilibrium in zeolites. *J Phys Chem C Nanomater Interfaces* 2021;125:4155-74. DOI PubMed PMC
  56. Wilson SMW, Tezel FH. CO<sub>2</sub> and CO adsorption equilibrium on ZSM-5 for different SiO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> ratios. *Sep Sci Technol* 2019;54:722-

30. DOI
57. Proding S, Derewinski MA. Recent progress to understand and improve zeolite stability in the aqueous medium. *Pet Chem* 2020;60:420-36. DOI
58. Heard CJ, Grajciar L, Rice CM, et al. Fast room temperature lability of aluminosilicate zeolites. *Nat Commun* 2019;10:4690. DOI PubMed PMC
59. Feng J, Hu Y, Bao Q, Liang D, Xu Y. Carbon monoxide and carbon dioxide adsorption on alkali metal cation-exchanged SSZ-13 zeolites. *Micro Nano Letters* 2020;15:529-34. DOI
60. Chen S, Zhu M, Tang Y, Fu Y, Li W, Xiao B. Molecular simulation and experimental investigation of CO<sub>2</sub> capture in a polymetallic cation-exchanged 13X zeolite. *J Mater Chem A* 2018;6:19570-83. DOI
61. Crawford P, Hu P. Importance of electronegativity differences and surface structure in molecular dissociation reactions at transition metal surfaces. *J Phys Chem B* 2006;110:24929-35. DOI PubMed
62. Wang L, Chen W, Zhang D, et al. Surface strategies for catalytic CO<sub>2</sub> reduction: from two-dimensional materials to nanoclusters to single atoms†. *Chem Soc Rev* 2019;48:5310-49. DOI
63. Yue B, Liu S, Chai Y, Wu G, Guan N, Li L. Zeolites for separation: fundamental and application. *J Energy Chem* 2022;71:288-303. DOI
64. Fu D, Davis ME. Carbon dioxide capture with zeotype materials. *Chem Soc Rev* 2022;51:9340-70. DOI PubMed
65. Joos L, Swisher JA, Smit B. Molecular simulation study of the competitive adsorption of H<sub>2</sub>O and CO<sub>2</sub> in zeolite 13X. *Langmuir* 2013;29:15936-42. DOI PubMed
66. Lee KM, Lim YH, Jo YM. Evaluation of moisture effect on low-level CO<sub>2</sub> adsorption by ion-exchanged zeolite. *Environ Technol* 2012;33:77-84. DOI PubMed
67. Mason JA. Metal-organic frameworks for gas storage and separation. Available from: <https://escholarship.org/uc/item/07w7s2d2>. [Last accessed on 4 Sep 2023].
68. Brandani F, Ruthven DM. The effect of water on the adsorption of CO<sub>2</sub> and C<sub>3</sub>H<sub>8</sub> on type X zeolites. *Ind Eng Chem Res* 2004;43:8339-44. DOI
69. Shafeeyan MS, Daud WMAW, Shamiri A, Aghamohammadi N. Adsorption equilibrium of carbon dioxide on ammonia-modified activated carbon. *Chem Eng Res Des* 2015;104:42-52. DOI
70. Huong PT, Lee BK. Improvement of selective separation of CO<sub>2</sub> over N<sub>2</sub> by transition metal-exchanged nano-zeolite. *Microporous Mesoporous Mater* 2017;241:155-64. DOI