ACS MATERIALS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/materialsau
Article

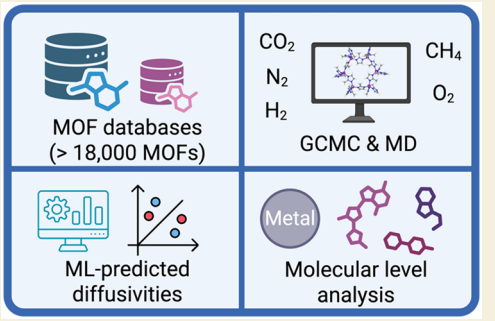
Molecular Modeling-Based Machine Learning for Accurate Prediction of Gas Diffusivity and Permeability in Metal−Organic Frameworks

Pelin Sezgin,[†] Feride Neva Yüngül,[†] Beste Naz Karaca,[†] Hasan Can Gulbalkan,[†] and Seda Keskin*

Cite This: *ACS Mater. Au* 2026, 6, 140−153
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Gas diffusion determines the performance of metal−organic frameworks (MOFs) in various practical applications, including membrane-based separations, yet its experimental measurement is challenging. We presented an efficient computational framework that integrates high-fidelity molecular dynamics (MD) simulations with machine learning (ML) to predict the diffusivities of CO_2, N_2, O_2, CH_4, and H_2 in >18,000 synthesized and hypothetical MOFs. ML models trained on MD data accurately predicted gas diffusivities of any given MOF within minutes using only easily accessible structural and guest-related properties. We provided an interactive, user-friendly web interface for predicting diffusivities of MOFs to facilitate material selection. Leveraging ML-predicted diffusivities, we evaluated membrane-based gas separation performances of all MOFs for seven industrially important separations: CO_2/N_2, CO_2/CH_4, N_2/CH_4, H_2/CO_2, H_2/CH_4, H_2/N_2, and O_2/N_2. The best MOF membranes offering high selectivity and permeability were identified and analyzed by using molecular fingerprinting to reveal the critical chemical properties for designing next-generation MOFs.



KEYWORDS: *metal−organic framework, molecular simulation, machine learning, diffusion, gas separation*

Downloaded via 212.2.212.152 on March 22, 2026 at 14:37:21 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

Metal−organic frameworks (MOFs) have emerged as highly promising materials for various chemical processes such as gas storage, gas separation, and catalysis, thanks to their tunable pore structures, high surface areas, and chemical diversities.[1,2] Central to many of these applications is a comprehensive understanding of gas diffusion within MOFs. For example, in gas storage systems, high diffusivity ensures rapid uptake and release of gases, whereas in gas separation systems, diffusion differences between gas species directly determine the selectivity. A MOF membrane's gas separation performance can only be evaluated if the gas diffusivities are known. Without a precise knowledge of diffusion behavior, the design and selection of the most appropriate MOFs for real-world applications would be inefficient.

The challenge here is that both experimental quantification and computational modeling of gas diffusivity in porous materials are inherently difficult. Experimentally, measuring gas diffusivity in MOFs is possible through techniques such as pulsed-field gradient nuclear magnetic resonance or quasi-elastic neutron scattering, but these methods suffer from issues such as the requirement of large sample volumes and sensitivity to crystal size distributions, which can introduce significant uncertainties into the gas diffusion measurement in MOFs.[3] Furthermore, as the chemical and structural diversity of MOFs continues to expand, it becomes increasingly impractical to evaluate every candidate material using experimental methods.[4] From a computational standpoint, simulating gas diffusivity in MOFs through molecular dynamics (MD) simulations has been shown to be highly accurate.[5] But performing MD simulations to compute the gas diffusivities requires very large computational time and resources, even for a single MOF. Therefore, high-throughput computational screening (HTCS) of large numbers of MOFs using MD simulations has been very limited in the literature: For example, self-diffusivities of CO_2, N_2, CH_4, and H_2 gases in more than 4000 synthesized MOFs were computed at infinite dilution, representing a condition where gas molecules do not interact with each other using MD simulations.[6−10] The diffusivities of CO_2, CH_4, and N_2 molecules at infinite dilution were also calculated for 17,257 hypothetical MOFs by MD simulations.[11] Diffusion coefficients of CO_2, CH_4, and N_2 molecules in 4764 synthesized MOFs were reported at practical operating conditions, 10 bar and 298 K.[12] Given the rapidly growing number of synthesized and computer-generated MOFs, it is practically impossible to perform MD

Received: July 10, 2025
Revised: November 4, 2025
Accepted: November 4, 2025
Published: November 12, 2025



© 2025 The Authors. Published by American Chemical Society

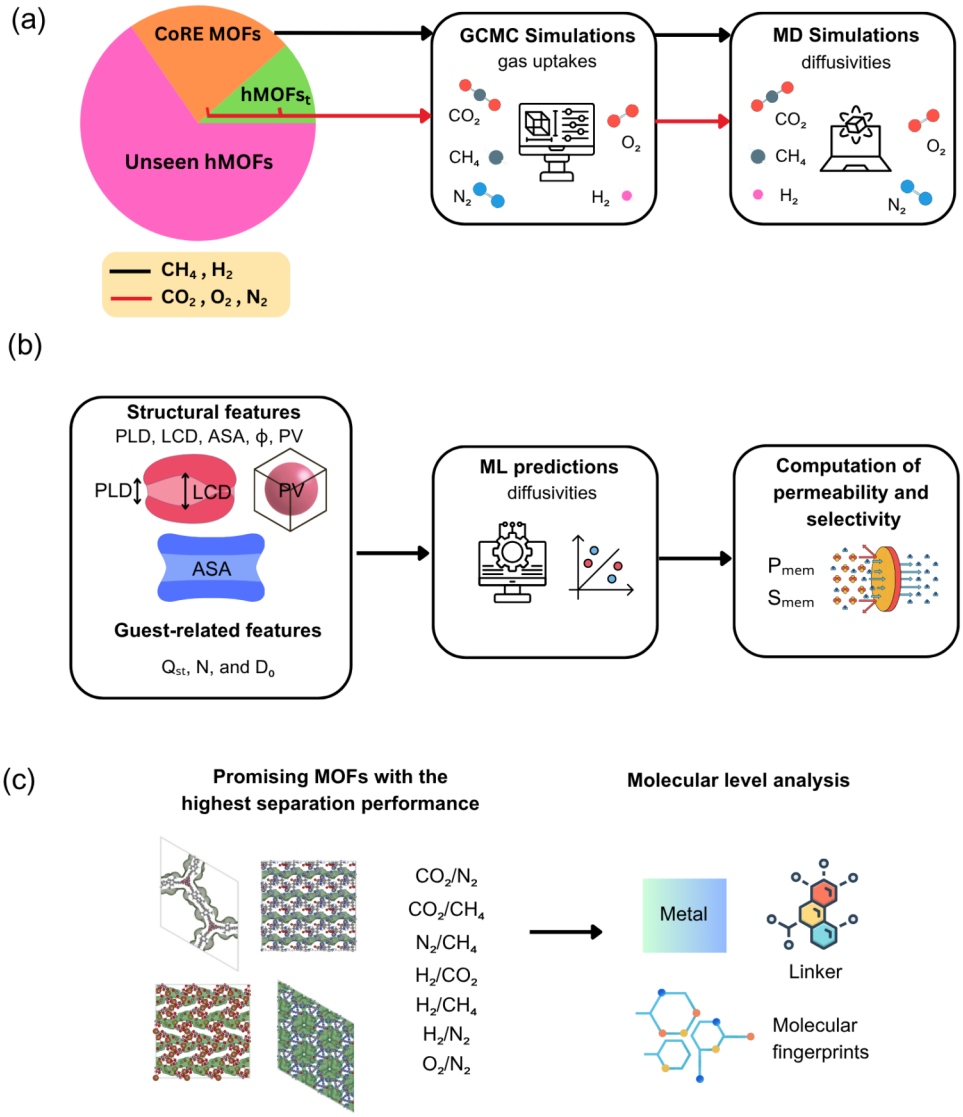


Figure 1. Molecular modeling-based ML approach: (a) CoRE MOF and hMOF databases were filtered based on the kinetic diameter of the guest molecules. GCMC and MD simulations were performed to compute the adsorbed gas loadings and the corresponding self-diffusivities of CO_2, O_2, N_2, CH_4, and H_2 gases. (b) Structural and guest-related features of CoRE MOFs and hMOFs were computed. The most important features affecting the gas diffusivities were extracted, and ML models were developed to predict diffusivities in seen and unseen MOF data sets. ML-predicted adsorption and diffusion data were combined to predict gas permeabilities and selectivities for CO_2/N_2, CO_2/CH_4, N_2/CH_4, H_2/CO_2, H_2/CH_4, H_2/N_2, and O_2/N_2 separations. (c) The most useful MOF membrane candidates were identified and analyzed using molecular fingerprinting.

simulations for a wide range of gas species across all available MOFs under various pressure conditions.

Machine learning (ML) has recently emerged as a powerful complementary approach in the field of MOF research.[13] Notable progress has been made in using ML to facilitate the discovery of optimal synthesis conditions of MOFs for enhancing the crystallinity,[14] to accelerate quantum chemistry calculations for the development of MOFs for air capture,[15] and to predict the gas adsorption capacities of MOFs.[16] All of these have significantly reduced the computational cost associated with screening large MOF databases. Despite this growing momentum, the application of ML to gas transport properties of MOFs, particularly diffusivity and permeability, is very limited: The diffusivity of CH_4 in 1500 synthesized MOFs was predicted using transfer learning by utilizing gas uptakes obtained from Grand Canonical Monte Carlo (GCMC) simulations, geometric properties of MOFs such as pore sizes, and energetic descriptors such as energy grids.[17] O_2 and N_2 diffusivities of 5000 synthesized MOFs were calculated using MD simulations at 1 bar and infinite dilution, and the resulting simulation data were used to develop ML models to predict gas diffusivities.[18] The impact of the flexibility of MOF structures on the computed gas diffusivities was studied for the synthesized MOFs by developing ML models. Results showed that the effect of framework flexibility may be safely neglected in many MOF−gas pairs to save computational time.[19] An ML model was developed to predict the diffusivities of different gases including Xe, Kr, and H_2S in 6013 synthesized MOFs using physical descriptors of the MOFs (such as pore sizes and surface areas) and the gas molecules (such as kinetic diameter and polarizability).[20] Our group[21] combined MD simulations with ML to assess the permeation of He, CH_4, H_2, and N_2 gases through 5249 MOF membranes and showed that ML-predicted gas permeation data of MOF membranes and MOF/

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

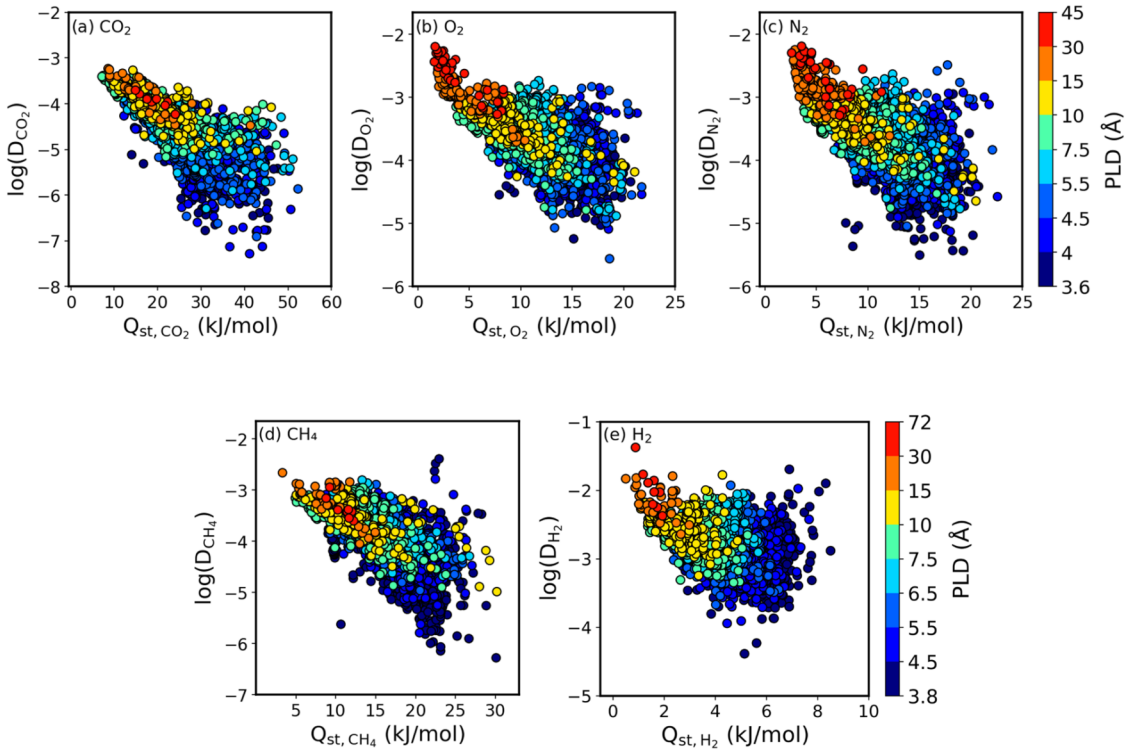


Figure 2. Molecular interplay between gas diffusivities and pore properties of MOFs: Diffusivities of (a) CO_2, (b) O_2, (c) N_2, (d) CH_4, and (e) H_2 gases in 3983, 3987, 3986, 3983, and 3983 CoRE MOFs, respectively, were analyzed as functions of their heat of adsorption values and the pore sizes of the frameworks. Diffusivities were computed by MD simulations.

polymer mixed matrix membranes are in good agreement with the permeation data computed from molecular simulations.

As this literature review highlights, diffusion data are currently available for only a limited variety of guest molecules and a limited number of MOFs, even though the MOF design space comprises several thousands of distinct materials. To fully harness the potential of both synthesized and hypothetical MOFs for diffusion-based applications, there is a strong need for robust computational approaches that can serve as efficient alternatives to traditional molecular simulations, which are often time-consuming and computationally expensive. By the development of ML-based approaches to predict the diffusion behavior of various gas species across a wide range of MOFs, high-throughput screening of large numbers of materials for membrane-based gas separation applications can be enabled, and the rational design of materials can be facilitated.

Motivated from this need, we proposed an efficient computational approach that combines high-fidelity molecular modeling methods with ML for predicting the diffusivities of five industrially important molecules: CO_2, N_2, O_2, CH_4, and H_2 in >18,000 distinct types of synthesized and hypothetical MOFs. The main contributions of our work are as following:

(a) A comprehensive molecular diffusivity data for CO_2, O_2, N_2, CH_4, and H_2 gases in synthesized and hypothetical MOFs was generated both at infinite dilution and at 1 bar using MD simulations, which represents the most extensive diffusion data set produced for MOFs to date.

(b) ML models were trained on this high-quality MD data to accurately predict the self-diffusivities of gases within 10 min on a computer with 8 CPUs, as an efficient alternative to MD simulations. Our ML models are publicly available and depend only on properties of MOFs and guest molecules that are either easily computable or experimentally accessible.

(c) Initially trained on experimental MOFs, our ML models were tested on unseen materials, a very large number of hypothetical MOFs, yielding accurate results that demonstrate the models' strong generalization capability. We designed an interactive, user-friendly, publicly available web interface (https://mofgasdiffusionpredictor.streamlit.app/) for predicting gas diffusivities of any given MOF to facilitate material selection and enable broader applicability in materials discovery.

(d) Using ML-predicted gas diffusivities, we assessed membrane-based gas separation performance of the entire MOF space for seven industrially, socially, and economically important gas separations, namely, CO_2/N_2, CO_2/CH_4, N_2/CH_4, H_2/CO_2, H_2/CH_4, H_2/N_2, and O_2/N_2. The most promising MOF membranes that can exceed the upper bound were identified, and their atomic structures were analyzed in detail using a molecular fingerprinting approach to describe the most useful metal sites, functional groups, and linkers of MOFs.

2. RESULTS AND DISCUSSION

The computational workflow of our study, combining molecular modeling and ML is shown in Figure 1. We examined the experimentally synthesized MOF database (CoRE MOF)[22] and a hypothetical MOF database (hMOF),[23] which include 11,793 and 23,891 materials, respectively. We only used the computation-ready structures from the recently updated version of the CoRE MOF

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

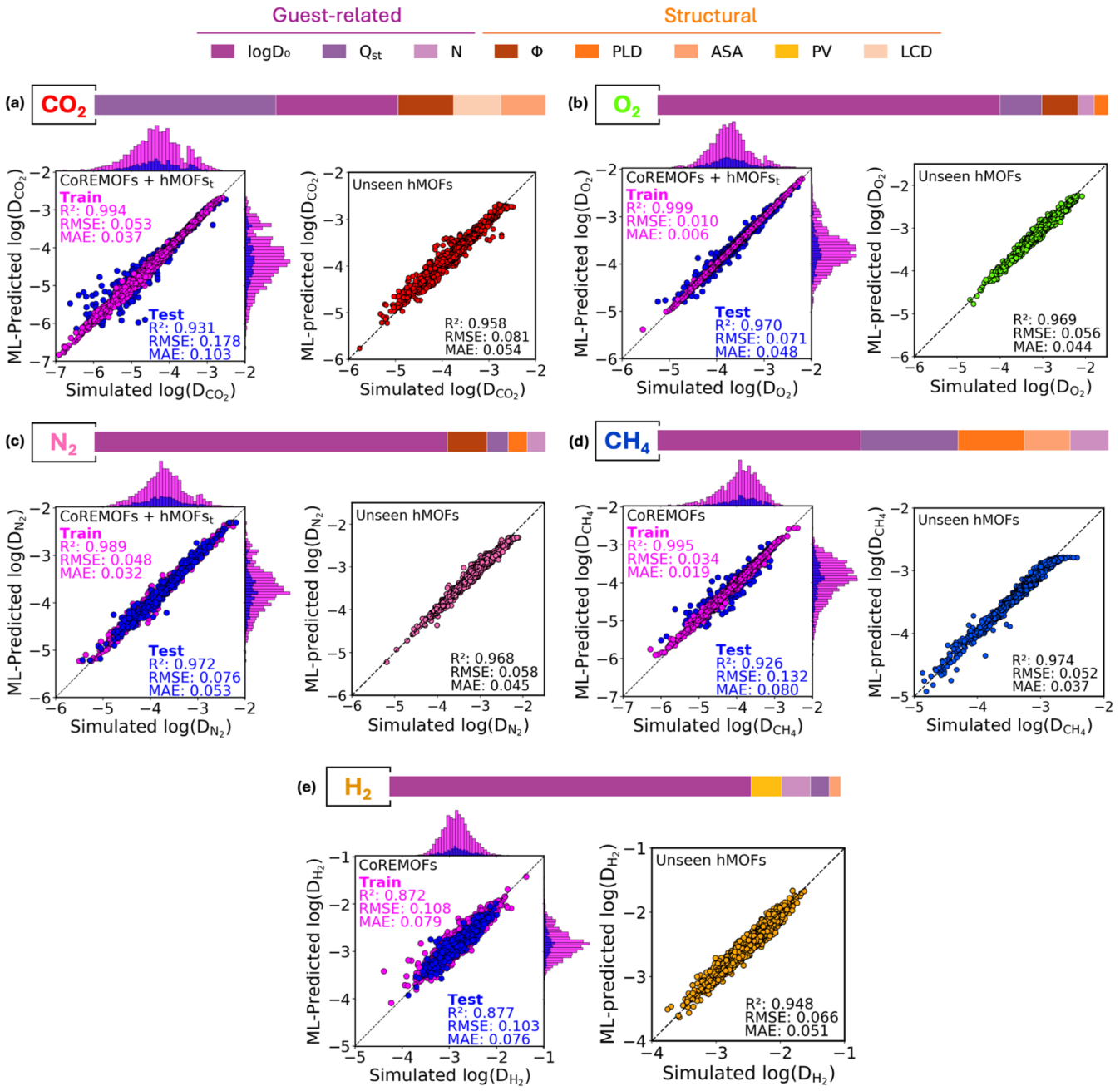


Figure 3. ML models for predicting the gas diffusivities of CoRE MOFs and hMOFs: ML models were trained using MD data generated for seen MOFs and tested on unseen hMOFs for (a) CO_2, (b) O_2, (c) N_2, (d) CH_4, and (e) H_2. The bar figures represent the importance of the five most important features used in ML models, the structural features are PLD, LCD, ASA, PV, and ϕ, and the guest-related features are log D_0, N, and Q_{st}. The number of seen and unseen MOFs used in developing and testing the ML models are given in Table S1.

database.[24,25] Zeo++ software (version 0.3)[26] was used to calculate the structural properties of each MOF, which include the pore limiting diameter (PLD), the largest cavity diameter (LCD), N_2 accessible surface area (ASA), pore volume (PV), density (ρ), and porosity (ϕ).

To ensure transport of all gas molecules through the pores, we filtered MOFs based on their pore sizes, requiring PLD > 3.64 Å for CO_2 (3.30 Å), O_2 (3.46 Å), and N_2 (3.64 Å), and PLD > 3.80 Å for CH_4 (3.80 Å) and H_2 (2.96 Å), where the values in parentheses represent the kinetic diameters of the gases. As a result, we ended up with 4663 CoRE MOFs for CO_2, O_2, and N_2, and 4331 CoRE MOFs for CH_4 and H_2. Before performing molecular simulations, the partial charges of MOF atoms were assigned using the charge equilibration method[27,28] to compute the electrostatic interactions between the gas molecules that have quadrupole moments: CO_2, N_2, O_2, and MOF atoms. The structures for which partial point charges were computed to be larger or smaller than ±5 were eliminated. We also eliminated the MOFs for which gas uptakes in the pores were computed to be very low at 1 bar, less than one molecule per unit cell. As a result, we studied 3989 CoRE MOFs for CO_2, O_2, and N_2 diffusion and 4193 CoRE MOFs for H_2 and CH_4 diffusion.

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

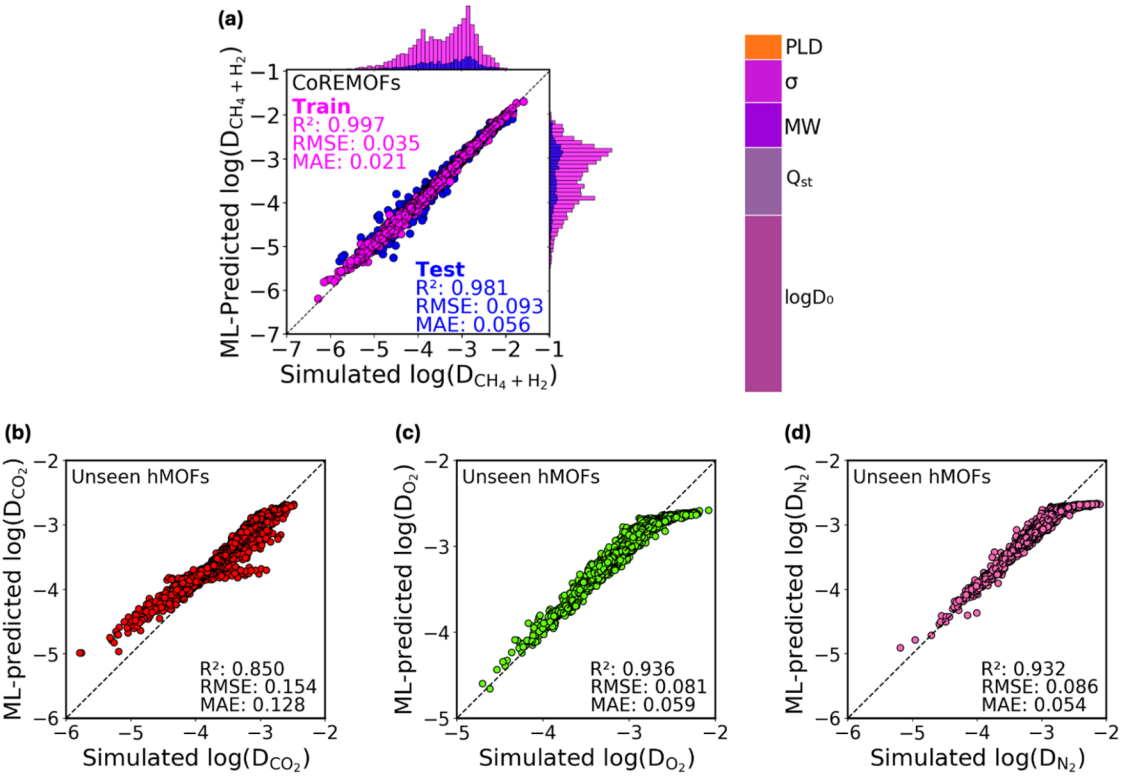


Figure 4. Universal ML model for predicting the diffusivities of unseen gases: ML model was trained on MD data generated for (a) CH_4 and H_2 gases across 3983 CoRE MOFs and tested to predict the diffusivities of (b) CO_2, (c) O_2, and (d) N_2 in 4080, 4476, and 4444 hMOFs, respectively. The bar figures represent the importance of the features used in ML models. The numbers of CoRE MOFs and hMOFs used in training and transfer are provided in Table S1.

2.1. Molecular Diffusion Characteristics of MOFs

We first examined the correlations among structural features, guest-related features, and the simulated gas diffusivities of CoRE MOFs in Figures S1 and S2 and then investigated how gas diffusivities vary with the strongly correlated features of MOFs: heat of adsorption values of each gas at infinite dilution (Q_{st}), PLD, and D_0 in Figure 2. After excluding the MOFs that have diffusivities < 10^{-8} cm^2/s, the limit of MD to accurately characterize molecular diffusion, we reported diffusivities for CO_2, CH_4, H_2, N_2, and O_2 in 3983, 3983, 3983, 3986, and 3987 CoRE MOFs, respectively. Diffusivities of CO_2, O_2, N_2, CH_4, and H_2 vary in the ranges of 5.2×10^{-8}–5.8×10^{-4}, 2.8×10^{-6}–1.8×10^{-3}, 3.2×10^{-6}–3.3×10^{-3}, 5.3×10^{-7}–4.1×10^{-3}, and 4.1×10^{-5}–4.2×10^{-2} cm^2/s, respectively. The self-diffusivities decrease as Q_{st} values increase since Q_{st} describes the strength of interactions between the gas molecules and the MOF structures. CO_2 is the slowest molecule due to its strong adsorption as indicated by the high Q_{st} values (Figure 2(a)). O_2 and N_2 diffusivities are comparable because of the similar physicochemical properties of these gases (Figure 2(b,c)). CH_4 diffusivity range is concentrated within moderate values compared to other gases (Figure 2(d)). H_2 is the fastest diffusing species because of its smallest kinetic diameter and lowest molecular weight, thus H_2 diffusivity has the lowest dependency on the pore geometry of MOFs (Figure 2(e)). The general trend shows that diffusivities increase as the PLDs increase. The linear relation between diffusivities computed at infinite dilution and 1 bar is also given in Figure S3.

ML models were developed to predict gas diffusivities in MOFs based on structural and guest-related descriptors, enabling accurate and rapid estimations while bypassing computationally intensive MD simulations. Figure 3 shows the performance of ML models to predict the logarithmic diffusivities (log D) of five gases in MOFs, along with the feature importance charts and transferability to unseen hMOFs that were not included in training the models. A logarithmic scale was used for gas diffusivities because they span several orders of magnitude across different MOFs, making logarithmic representation more appropriate for analysis. Figure 3(a–e) presents that ML-predicted diffusivities of CO_2, O_2, N_2, CH_4, and H_2 demonstrate a strong agreement with the simulated diffusivities, yielding the coefficient of determination (R^2) values of 0.931, 0.970, 0.972, 0.926, and 0.877, respectively, thereby confirming the accuracy of the ML models.

The horizontal bars shown above the models of each gas visualize the importance of features obtained from the Shapley Additive Explanations (SHAP) feature importance analysis based on mean values. Across all models, $\log D_0$ consistently ranks as the most dominant descriptor, except for CO_2, reaffirming the importance of guest-related features that influence the diffusivity behavior, such as Q_{st} and N. The most important structural features are PLD, ϕ, and ASA, capturing the guest confinement effects within pores. For CO_2, Q_{st} is the most important feature, and ϕ, LCD, and ASA have the greatest influence on the output of the model. For CH_4, the importance of D_0 and Q_{st} is followed by PLD and ASA, indicating a balance between guest-related features and pore geometry. In the case of O_2 and N_2, feature importance is inclined toward guest-related features, and structural features such as PLD and ϕ have relatively minor contributions. For H_2, PV plays a notable role in addition to D_0. Figure S4(a–e)

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140–153

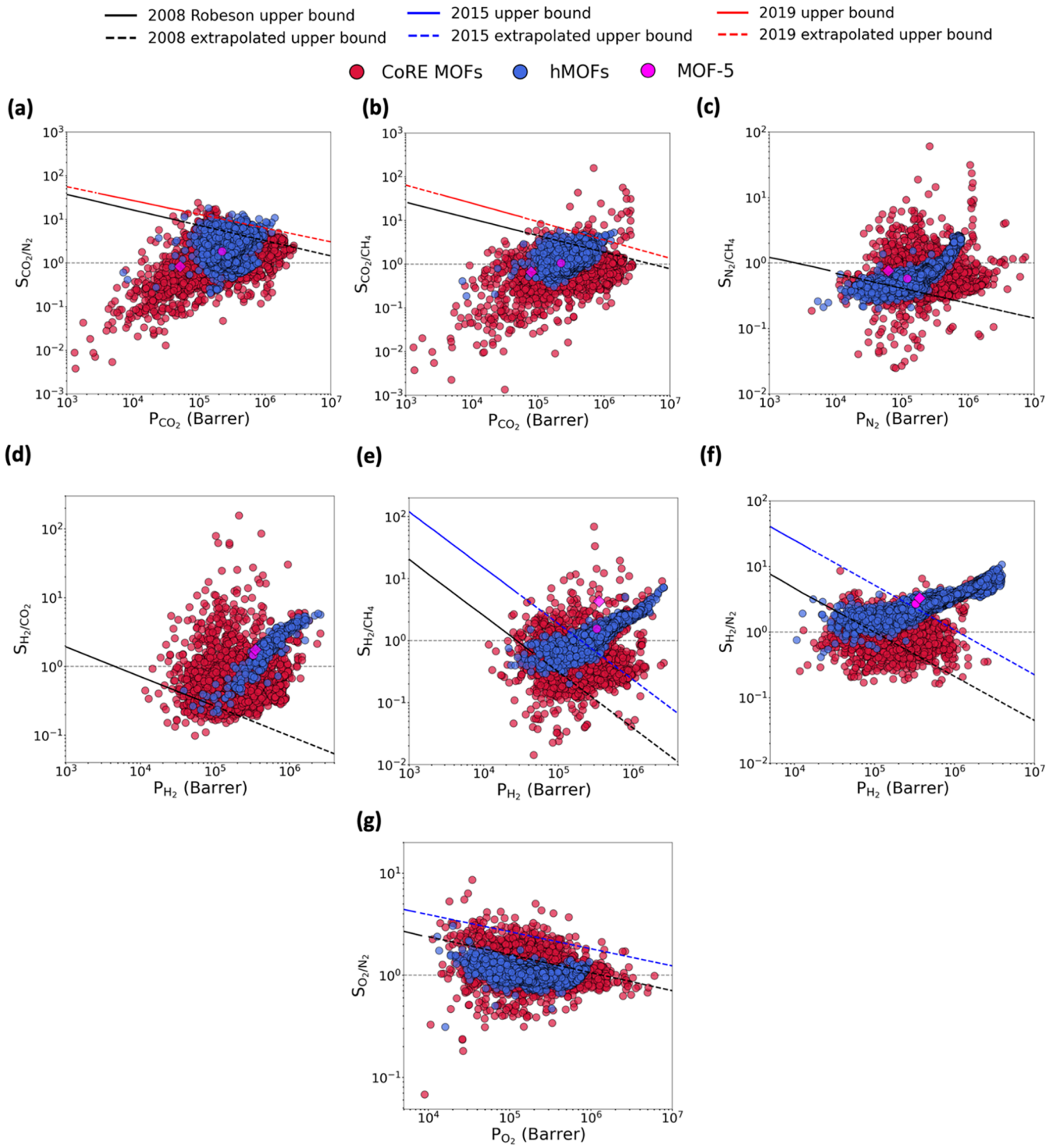


Figure 5. Membrane-based gas separation performance of the MOF space: ML-predicted gas permeabilities and selectivities for CoRE MOFs and hMOFs for (a) CO_2/N_2, (b) CO_2/CH_4, (c) N_2/CH_4, (d) H_2/CO_2, (e) H_2/CH_4, (f) H_2/N_2, and (g) O_2/N_2 separations. Experimentally reported gas permeabilities and selectivities of the MOF-5 membrane were also included for comparison with diamond symbols. Table S2 presents the number of MOFs studied for each separation.

shows the SHAP feature distribution plots of the ML models. For all models, higher D_0 values and lower Q_{st} values lead to an increase in the diffusivities.

We then evaluated the model transferability to a large unseen hMOF data set: 4080 hMOFs for CO_2, 4476 hMOFs for O_2, and 4444 hMOFs for N_2. The results showed that the ML-predicted values are well aligned with the simulated data with R^2 values of 0.958 (CO_2), 0.969 (O_2), and 0.968 (N_2), confirming strong generalization across the hMOF data set. Similarly, testing for 2788 hMOFs exhibits high R^2 values of 0.974 and 0.948 for CH_4 and H_2, respectively. For CO_2, a slight underestimation at the high-diffusivity end is observed in Figure 3(a). As Q_{st} is the most important feature to predict diffusivity, the deviation arises from the differences in the Q_{st}

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

ranges of the training set extending up to 52.3 kJ/mol and the hMOF subset up to 53.8 kJ/mol. Similarly, as shown in Figure 3(d), a deviation between ML predictions and simulation results for CH_4 diffusivities is observed during the model transfer to unseen hMOFs, and diffusivities > 1.5×10^{-3} cm^2/s were generally underpredicted by the model. This is again due to the Q_{st} range of the CoRE MOFs used to train the model, which extends up to 30.2 kJ/mol, whereas in the hMOF subset, Q_{st} values have a lower limit of 22.9 kJ/mol (Figure S5). Since lower Q_{st} values correspond to higher diffusivities as discussed in the SHAP analysis, CH_4 diffusivities tend to saturate in this region. For CO_2 and CH_4, only 1.3 and 3% of MOFs, respectively, fall into these high-diffusivity regions. Thus, the plateau is a result of the limited number of high-diffusivity examples in the training data rather than a systematic error of the ML models.

After developing gas-specific ML models, we built a universal model capable of generalizing diffusivity predictions across multiple gases, thereby reducing the need to use separate models for each gas species. We used simulated diffusivities of CH_4 and H_2 obtained for CoRE MOFs and incorporated the kinetic diameter (σ) and molecular weight (MW) of each gas into the feature set as identifiers of the gas type. CH_4 and H_2 were selected for training because they span the range of kinetic diameters among the gases we considered, effectively establishing the lower and upper bounds. In addition, these gases do not have electrostatic interactions with the framework; therefore, the time requirement to obtain D_0 values is significantly less. Figure 4(a) shows the high prediction power of the universal ML model, resulting in a high R^2 value (0.98). According to the SHAP analysis given in Figure S4, the five most important features of the universal model are D_0, Q_{st}, MW, σ, and PLD, showing the critical role of gas identity.

Figure 4(b–d) demonstrates the universal model's transferability to both unseen gases and unseen materials: Trained on CH_4 and H_2, modeled as single-sphere nonpolar molecules, the model successfully generalizes to CO_2, O_2, and N_2, which are represented as three-site molecules with quadrupole moments. The model also performs well on unseen hMOFs. The unseen diffusivity predictions for CO_2 exhibit the largest deviations from the simulated data. This can be attributed to Q_{st}, which is the second most important feature to predict diffusivities. For CO_2, Q_{st} values of unseen hMOFs are in the range of 4.6–53.8 kJ/mol, whereas Q_{st} values of CH_4 and H_2 used for the train set cover the range of 0.5–30.2 kJ/mol. This leads to overprediction at the low diffusivity range, as shown in Figure 4(b). The saturation observed at the highest diffusivity values for O_2 and N_2 in Figure 4(c,d) is due to the tree-based nature of the model (ExtraTreesRegressor),[29] which cannot extrapolate beyond the range of the training data. Despite these limitations, the universal model provides reliable predictions across a wide range of unseen gas–unseen MOF combinations.

2.2. Gas Permeation Analysis of MOFs

After demonstrating the accuracy of our ML models, we used the ML-predicted diffusivities to assess gas permeabilities and membrane selectivities of 3982, 3444, 3442, 3444, 3983, 3445, and 3986 CoRE MOFs and 14,180 hMOFs for CO_2/N_2, CO_2/CH_4, N_2/CH_4, H_2/CO_2, H_2/CH_4, H_2/N_2, and O_2/N_2 separations, respectively, in Figure 5. A comparison between ML-predicted and simulated gas permeabilities is presented in Figure S6, showing good agreement. To better assess the potential of MOF membranes for each separation, we included the upper bounds[30,31] that define the trade-off relationship between permeability and selectivity of polymer membranes. Since the permeabilities of both CoRE MOFs and hMOFs are significantly higher than those of the polymers, we extrapolated the upper bounds using dashed lines for all separations. The gas separation performance of membranes is governed by the complex interplay between adsorption and diffusion dynamics; therefore, some MOF membranes preferentially allow the permeation of one species, while others favor the opposite component. We draw a dotted line at a selectivity of unity to better represent which component is preferentially permeated.

Figure 5(a,b) represents CO_2/N_2 and CO_2/CH_4 separation performances of CoRE MOFs and hMOFs, respectively. The CO_2 permeabilities were computed to be between 1.3×10^3 and 2.7×10^6 Barrer for CoRE MOFs and between 7.2×10^3 and 1.4×10^6 Barrer for hMOFs. CO_2/N_2 selectivity values were computed to be in the range of 0.004–23.9 for CoRE MOFs and 0.2–18 for hMOFs. 3067 out of 3982 CoRE MOFs and 12,027 out of 14,180 hMOFs were identified as CO_2-selective for CO_2/N_2 separation, representing the potential of MOF membranes for the postcombustion carbon capture process. The CO_2/CH_4 selectivities of CoRE MOFs and hMOFs were computed to be 0.0014–159 and 0.19–5.8, respectively. 1975 CoRE MOFs out of 3444 and 5088 hMOFs out of 14,180 were identified as CO_2-selective for CO_2/CH_4 separation, showing the promise of MOFs for natural gas purification. A total of 426 (142) CoRE MOFs and 614 (293) hMOFs surpassed the 2008 Robeson upper bound, while only 19 (30) CoRE MOFs and 24 (10) hMOFs exceeded the 2019 upper bound for CO_2/N_2 (CO_2/CH_4) separations. We studied 3442 CoRE MOFs and 14,180 hMOFs for N_2/CH_4 separation, a process important for natural gas upgrading, as shown in Figure 5(c). N_2 permeabilities were calculated to be between 4.3×10^3 and 7.3×10^6 Barrer and between 5.4×10^3 and 8.0×10^5 Barrer for CoRE MOFs and hMOFs, respectively. N_2/CH_4 selectivity values ranged from 0.025 to 60.6 for CoRE MOFs and from 0.2 to 2.7 for hMOFs. 349 CoRE MOFs and 539 hMOFs were found to be N_2-selective. 2804 CoRE MOFs and 13,356 hMOFs were located above the 2008 Robeson upper bound.

Figure 5 (d–f) shows the H_2 purification performance of MOFs from other gases. We studied 3444, 3983, and 3445 CoRE MOFs for H_2/CO_2, H_2/CH_4, and H_2/N_2 separations, respectively, in addition to 13,971 hMOFs. CoRE MOFs showed H_2 permeabilities ranging from 1.2×10^4 to 2.6×10^6 Barrer with selectivities of 0.1–156 (H_2/CO_2), 0.014–68.7 (H_2/CH_4), and 0.16–8.6 (H_2/N_2). hMOFs were found to offer similar H_2 permeabilities but generally lower selectivities of 0.1–21.2 (H_2/CO_2), 0.2–12 (H_2/CH_4), and 0.4–10.7 (H_2/N_2). 659, 574, and 1699 CoRE MOFs and 10,360, 11,079, and 13,945 hMOFs were identified as H_2-selective for the respective separations. A total of 3172, 3491, and 2400 CoRE MOFs (13,664, 13,868, and 13,850 hMOFs) surpassed the 2008 Robeson upper bound for H_2/CO_2, H_2/CH_4, and H_2/N_2 separations, respectively. However, 702 and 223 CoRE MOFs (11,734 and 10,079 hMOFs) exceeded the 2015 upper bound for H_2/CH_4 and H_2/N_2 separations, respectively.

Figure 5(g) shows air separation performances of MOF membranes, where a total of 3986 CoRE MOFs and 14,180 hMOFs were evaluated for O_2/N_2 separation. O_2 permeabilities ranged from 9.0×10^3 to 6.0×10^6 Barrer for CoRE

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140–153

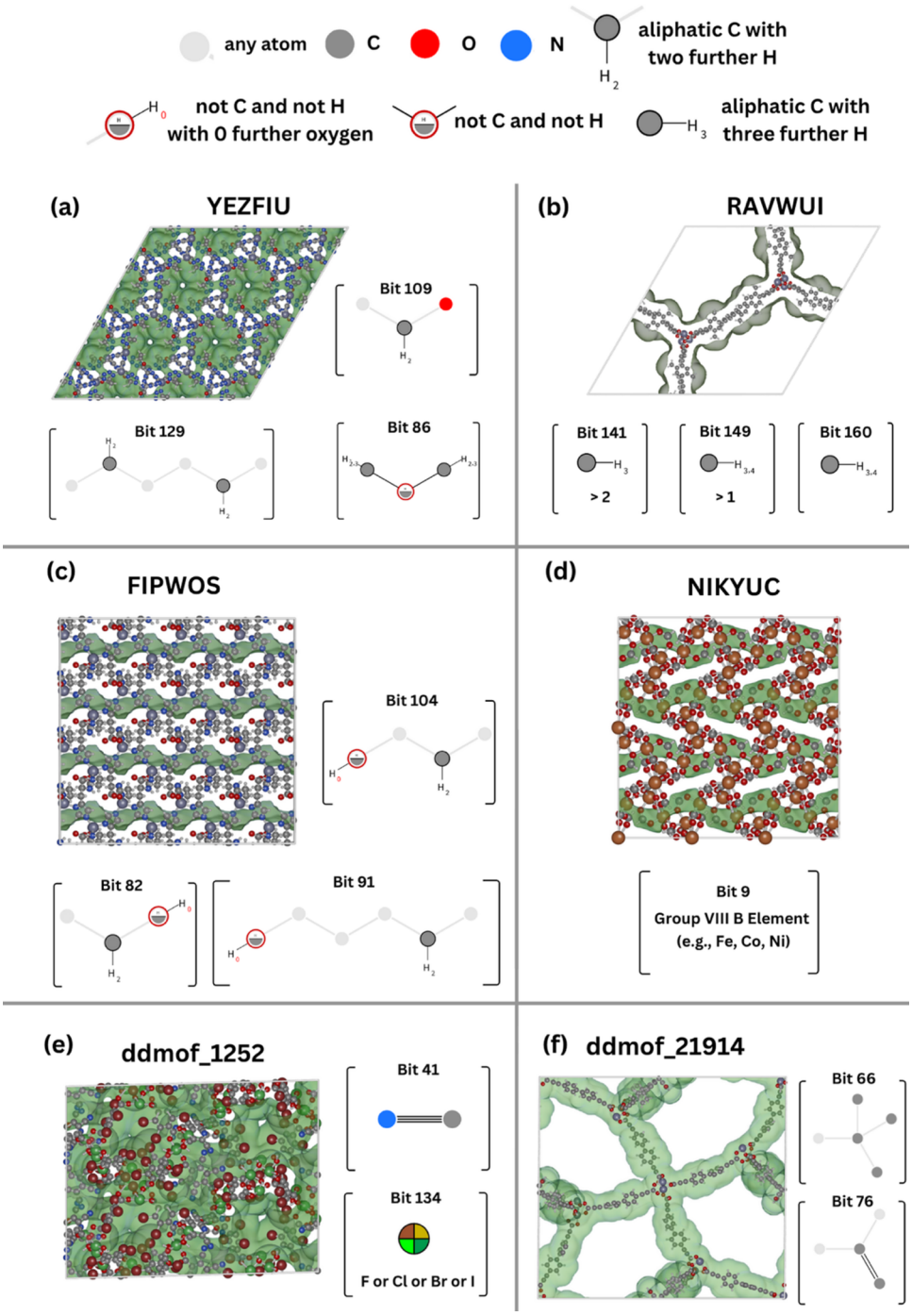


Figure 6. Key chemical properties of high-performing MOF membranes: Molecular fingerprint analysis for (a–d) CoRE MOFs and (e, f) hMOFs. Carbon, hydrogen, oxygen, nitrogen, bromine, fluorine, zinc, iron, and nickel atoms in the framework are represented with gray, white, red, blue, dark red, light green, light purple, orange, and green colors, respectively. The visualizations of MOFs were generated using the iRASPA software,[41] and isosurfaces (green regions generated based on the N_2 probe) were presented.

MOFs and from 1.3×10^4 to 9.5×10^5 Barrer for hMOFs. Like many other porous materials, O_2/N_2 selectivity values are low, between 0.07 and 8.6 and between 0.3 and 3 for CoRE MOFs and hMOFs, respectively. 2649 CoRE MOFs and 6318 hMOFs were identified as O_2-selective. Only 58 CoRE MOFs exceeded the 2015 bound, and no hMOFs reached this threshold. These results highlight the challenging nature of the O_2/N_2 separation due to similar kinetic diameters and diffusivities of the two gases, ultimately limiting membrane selectivity.

We then analyzed the pore sizes of the MOF membranes that can exceed the upper bound for each gas separation: In the CO_2/N_2 (CO_2/CH_4) separation, the CoRE MOFs that surpassed both upper bounds featured PLDs of 3.6–4.2 Å (3.8–11.2 Å) and LCDs of 4.6–11.5 Å (4.4–11.0 Å), indicating that narrow pores enhance membrane selectivity while maintaining high CO_2 permeability. The top-performing hMOFs that surpassed both upper bounds exhibited larger pore sizes: PLDs of 4.1–24.5 Å (6.3–30.8 Å) and LCDs of 7.5–28.6 Å (11.7–32.8 Å) for CO_2/N_2 (CO_2/CH_4) separa-

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140–153

tion. For N_2/CH_4 separation, CoRE MOFs that surpassed the upper bound have PLD and LCD values ranging from 3.8 to 38.1 Å and from 4.00 to 38.4 Å, respectively. For three different H_2 separations, high-performing CoRE MOFs and hMOFs have large pore sizes, reaching up to 72 and 45 Å, respectively. These results highlight the potential of large-pored MOFs favoring high diffusivity of H_2. Further insights into the structural distributions of PLD, LCD, ASA, porosity, and PV of CoRE MOFs and hMOFs that surpassed the upper bounds are provided in Figure S7.

To further confirm the accuracy of ML in predicting gas separation performances of MOFs, we included the experimentally reported gas permeability and selectivity data of the MOF-5 membrane for different gas separations in Figure 5. MOF-5 was chosen as a reference membrane material because it was the first MOF synthesized in the form of a thin film membrane[32] used as the MOF membrane in the first atomistic modeling study,[33] and it has reported permeability data for most of the gases that we studied in this work. ML predictions for the MOF-5 membrane closely match experimental data across most separations, demonstrating the model's accuracy. Both ML-predicted and experimental values for MOF-5 fall below the upper bounds for CO_2/N_2 and CO_2/CH_4 separations, indicating moderate permeability but low selectivity. For N_2/CH_4 separation, MOF-5 is nonselective and exhibits relatively low N_2 permeability, despite surpassing the 2008 upper bound. For H_2/CO_2 separation, it shows high H_2 permeability but low selectivity, located above the 2008 upper bound. Moreover, for H_2/CH_4 and H_2/N_2 separations, MOF-5 exceeds the 2015 upper bound, highlighting a better performance for hydrogen-based separations. These results are consistent across both the ML predictions and experimental data.

For other MOF membranes that were in our train set but have permeability data available for only certain gas separations, such as CPO-27-Ni and Cu-BTC, the agreement between the ML predictions and experimental values is also good. For example, ML-predicted CO_2 permeability and CO_2/CH_4 selectivity of the CPO-27-Ni membrane[34] (3.9×10^5 Barrer and 0.2, respectively) closely aligned with experiments (1.6×10^5 Barrer and 0.7, respectively). Similarly, ML-predicted H_2 and N_2 permeabilities for CPO-27-Ni (9.6×10^5 and 7.8×10^5 Barrer, respectively) agree reasonably well with the experimental results (3.6×10^5 and 2.1×10^6 Barrer, respectively). ML-predicted CO_2, N_2, CH_4, and H_2 permeabilities for the Cu-BTC membrane (6.3×10^5, 1.4×10^5, 3.4×10^5, and 2.3×10^5 Barrer) generally tend to overestimate the experimental values[35] (5.0×10^4, 5.0×10^4, 2.9×10^4, and 2.3×10^5 Barrer, respectively). This is consistent with previous discussions, highlighting that MOFs are modeled as perfect, defect-free crystals in molecular simulations, leading to higher predicted permeabilities compared to fabricated membranes, which may contain structural defects.[36,37]

Finally, we identified the most promising MOF membranes offering the highest TSP values, the multiplication of membrane selectivity and natural logarithm of permeability, from both the CoRE MOF and hMOF data sets. To identify the top-performing MOFs common to at least two different gas separations, we examined the top 60 CoRE MOFs and 30 hMOFs for each separation, resulting in 297 unique CoRE MOFs and 163 unique hMOFs.

Figure 6(a–f) shows atomic representations of the promising MOFs: YEZFIU for CO_2 separations (CO_2/CH_4 and CO_2/N_2), RAVWUI for H_2 separations (H_2/CH_4, H_2/CO_2, and H_2/N_2), FIPWOS for separation of other gases from CH_4 (H_2/CH_4 and N_2/CH_4), and NIKYUC for separation of gases from N_2 (CO_2/N_2 and O_2/N_2). For the promising hMOFs, we identified ddmof_1252 for CO_2/N_2 and CO_2/CH_4 separations, and ddmof_21914 for H_2-based separations (H_2/CH_4, H_2/CO_2, and H_2/N_2). Table S3 lists these four most promising CoRE MOFs and two hMOFs, with their confirmed structural and chemical validity as analyzed by using the recent MOSAEC algorithm,[38] Manz's tool,[39] MOFchecker package,[40] and manual checks. The table also includes our notes on the stability of CoRE MOFs based on the information obtained from their corresponding synthesis articles.

YEZFIU was identified as a promising membrane candidate because of its high TSP for the CO_2/CH_4 and CO_2/N_2 separations. Although CO_2 typically diffuses one to two orders of magnitude slower than CH_4 and N_2 in most MOFs, CO_2 diffusivity in YEZFIU was computed to be comparable to (slightly higher than) that of N_2 (CH_4), which leads to diffusion selectivities of 1.1 (CO_2/N_2) and 1.5 (CO_2/CH_4). As a result, both adsorption and diffusion favor CO_2 over other gases in this MOF. Similarly, the CO_2 diffusivity in the hypothetical ddmof_1252 was computed to be closer to N_2 and CH_4 with corresponding diffusion selectivities of 0.6 and 0.7, respectively. As a result, the membrane selectivity of this hMOF was calculated as 11.6 for CO_2/N_2 separation, locating it over both upper bounds. The high TSP value of RAVWUI , thanks to its high H_2 permeability (1.6×10^6 Barrer) and high membrane selectivity, is driven by the fast H_2 diffusion in the large pores of the MOF. Although H_2 is very weakly adsorbed in RAVWUI compared to other gases, it diffuses 43, 93, and 17 times faster than CH_4, CO_2, and N_2, respectively. As a result, high diffusion selectivity toward H_2 dominates the adsorption selectivity toward other gases and locates the membrane above the upper bounds for all three separations. Similarly, the large-pored hMOF, ddmof_21914, exhibits high diffusion selectivities of 22.6, 17.5, and 12.5 toward H_2 over CH_4, CO_2, and N_2, respectively, placing this membrane above the upper bounds. FIPWOS was identified to be promising for N_2/CH_4 separation due to its significantly higher N_2 permeability over CH_4, which can be attributed to slower diffusion of the latter in the narrow pores of this MOF. NIKYUC was identified as the top membrane candidate for O_2/N_2 separation with the highest membrane selectivity of 8.6 among all MOFs since both adsorption and diffusion favor O_2 over N_2 in this structure.

We analyzed the promising molecular fingerprints of MOFs by identifying the chemical bits (Molecular Access Keys (MACCS keys)), which represent generic chemical entities with changing elements, groups, and bonds. To define the atomic and molecular environments, MACCS keys use the following symbols: A is any valid periodic table element, Q is the heteroatom (any non-C and non-H atom), X is the halogen (F, Cl, Br or I), and $CH_3 > 2$ represents at least two methyl groups present within the structure. We found that there are 176 CoRE MOFs and 79 hMOFs that share all of the common bits found within the most promising four CoRE MOFs and two hMOFs in their respective groups out of 297 CoRE and 163 hMOFs. We specifically focused on the bits unique to only one of the four CoRE MOFs and two of the hMOFs: In YEZFIU, 86th (CH_2–Q–CH_2), 109th (A–CH_2–O–), and 129th (A–CH_2–A–A–CH_2–A) bits were unique and the presence of O and N atoms in the bits is known to enhance the

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140–153

electrostatic interactions with CO_2.[42] In RAVWUI, 141st (CH_3 > 2), 149th (CH_3 > 1), and 160th (CH_3) bits were found to be material-specific, which are methyl-rich, nonpolar fractions. FIPWOS has bits of 82nd (A−CH_2−QH), 91st (QH−A−A−A−CH_2−A), and 104th (QH−A−CH_2−A) that represent bulky molecular fragments. NIKYUC has the unique ninth bit that represents group 8B metals (e.g., Fe, Co, and Ni). Since this bit expresses the presence of open d-orbital sites, it favors CO_2 over N_2 in addition to O_2 over N_2.[43] In ddmof_1252, 41st (CTN (carbon triple bonded nitrogen), cyanide) and 134th (#X (halogen)) bits were unique. These functional groups are known to provide strong electrostatic interactions with CO_2, thus contributing to CO_2 selectivity over CH_4 and N_2.[44] The 66th (CC(C)(C)A) and 76th (C═C(A)A) bits were identified in ddmof_21914. The presence of acetylene and carbon-rich fragments favors CO_2 adsorption.[45]

3. CONCLUSION

We developed a molecular modeling-based ML approach for accurate prediction of gas diffusivity and permeability in MOFs. We trained our ML models on high-fidelity MD simulation data to enable accurate and rapid prediction of CO_2, N_2, O_2, CH_4, and H_2 diffusivities in any given MOF using only structural and guest-related descriptors. To facilitate high-throughput screening, we generated an interactive, user-friendly web interface that accurately predicts the gas diffusivities based on our approach and made the interface and all our ML models publicly available for the benefit of the broader MOF research community. To demonstrate the value of having gas diffusivity data for over 18,000 distinct types of synthesized and hypothetical MOFs across a wide range of gas molecules, we assessed the membrane performances of all MOFs for CO_2/N_2, CO_2/CH_4, N_2/CH_4, H_2/CO_2, H_2/CH_4, H_2/N_2, and O_2/N_2 separations. Our results revealed that many MOFs, particularly hMOFs, exhibit permeabilities several orders of magnitude higher than those of polymer membranes thanks to their large porosities and wide pores (PLD values up to 45 Å), with high selectivities reaching up to 159 for CO_2/CH_4 and 156 for H_2/CO_2 in CoRE MOFs, and up to 21.2 for H_2/CO_2 and 18.0 for CO_2/N_2 in hMOFs. These results also underline that many high-performance membrane candidates have yet to be synthesized and numerous existing MOFs have not been explored for membrane applications. Molecular fingerprinting of the most promising membrane materials identified the key chemical features, such as the presence of polar linker fragments and electronegative functional groups (cyanide and halogen) for enhanced CO_2 separations. These insights into structure−property relations will provide valuable guidance for the rational design of next-generation MOF membranes. Overall, our study demonstrates the power of a molecular modeling-based ML approach for screening the vast MOF design space and provides a very large amount of molecular transport data to support future experimental and computational MOF research.

4. METHODS

4.1. Molecular Simulations

To investigate the diffusion behaviors of CO_2, N_2, O_2, CH_4, and H_2 gases in CoRE MOFs, we performed MD simulations at infinite dilution and 1 bar. For infinite dilution, 30 gas molecules were inserted into a simulation box for each MOF, and gas−gas interactions were turned off. GCMC simulations were performed to compute the number of adsorbed gas molecules at 1 bar. All GCMC and MD simulations were performed using the RASPA software (version 2.0.5).[46] CH_4 and H_2 gases were treated as single and spherical atoms,[47,48] whereas CO_2 was modeled as a three-site linear rigid molecule,[49] as well as N_2 and O_2, which comprise two atoms and a dummy site representing the center of mass.[50,51] The dispersive interactions between frameworks and gases were described using the Lennard-Jones 12-6 (LJ) potential and the interaction parameters were obtained from the universal force field (UFF).[52] Electrostatic interactions were described by using the Coulomb potential. The pair potentials between different atoms were calculated via Lorentz−Berthelot mixing rules. A cut-off distance of 13 Å was used to truncate the intermolecular interactions, and the simulation cell lengths were increased to a minimum of 26 Å along each dimension, and periodic boundary conditions were applied. Ewald summation was employed for the accurate treatment of long-range electrostatic interactions.[53] Each GCMC simulation was performed using 10^4 initialization and 2 × 10^4 production cycles. The heat of adsorption values of each gas at infinite dilution (Q_{st}) were also calculated using the Widom particle insertion method.[54]

Each MD simulation consisted of 5 × 10^6 cycles, including 10^3 initialization and 10^4 equilibration cycles. With a 1 fs time step and a total simulation time of 5 ns, MD simulations were performed in the NVT ensemble at 298 K, employing a Nosê−Hoover thermostat.[55] The infinite dilution diffusivities (D_0) and diffusivities at 1 bar (D) were computed based on the MSD data via Einstein's relation in each direction.[56] MSD profiles were analyzed to identify the linear regions from which diffusivities were extracted based on the slope of the trajectories. Directional trajectories were excluded if their linear fit yielded a coefficient of determination (R^2) less than 0.9 or if diffusivities deviated by more than three orders of magnitude from the highest diffusivity observed among the three directions. The diffusion coefficients were calculated as the average of the directional values. We excluded the MOFs that have diffusivities < 10^{-8} cm^2/s, which is the limit of MD to accurately characterize molecular diffusion. Finally, we computed diffusivities for CO_2, CH_4, H_2, N_2, and O_2 in 3983, 3983, 3983, 3986, and 3987 CoRE MOFs, respectively.

Previous studies by our group have demonstrated that molecular simulations can accurately reproduce the experimental permeabilities of CO_2, CH_4, H_2, and N_2 gases in several MOFs, including Cu-BTC, Ni-MOF-74, MOF-5, MIL-53(Al), ZIF-8, ZIF-90, and ZIF-95.[7−10] We note that it is not computationally feasible to evaluate the membrane-based gas separation performance of more than 18,000 MOFs that we examined in this work under fully flexible conditions due to the high computational cost of calculating intra- and intermolecular interactions. Therefore, we used a rigid framework assumption in MD simulations. This assumption may lead to inaccuracies for certain MOFs, just as force fields may fail to perfectly capture all interactions. However, this does not invalidate our results. The scope of our work clearly states that the ML models were trained to reproduce diffusivities obtained from simulations under a rigid framework assumption, which served as the computational ground truth. The primary goal was to demonstrate how the ML can be effectively applied within this theoretical framework. Furthermore, in our previous studies,[7,57] we examined the impact of flexibility on the gas permeability of representative MOFs. For example, CO_2 and CH_4 diffusivities were computed to be lower in flexible simulations than those obtained from rigid simulations because of the decrease in the pore sizes of the MOFs.[7] As a result, the membrane permeabilities and selectivities were lower in the flexible case. In another work,[57] we showed that structural flexibility has a more pronounced impact on the separation performance of narrow-pored MOFs, especially when pore sizes are close to the kinetic diameters of the gases, whereas its effect is far less significant in large-pored MOFs. Because only 12.7% of the MOFs examined in this work have PLDs below 5.0 Å, flexibility is not expected to significantly affect the overall trends or the ranking of the MOF membranes.

4.2. Machine Learning

We developed ML models to capture the relationships between the MOF descriptors and the target gas diffusivities. We used structural

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

and guest-related descriptors (listed in Table S4) for varying numbers of CoRE MOFs and hMOFs (listed in Table S1) as the input data, while the simulated diffusion coefficients obtained from MD simulations performed at 1 bar were used as the target data to train the models. Correlation heatmaps illustrating both the interfeature relations and correlations with diffusivities are shown in Figures S1 and S2. The feature set employed in our study was designed to represent the key physical and chemical characteristics of MOFs while also incorporating energetic descriptors to represent framework−guest interactions. The correlation heatmaps presented in Figures S1 and S2 illustrate the relationships among the descriptors and their relevance to the target property. Features showing meaningful correlations with the diffusivity values were retained to strengthen the model performance. Thus, the final set of descriptors integrates both theoretical considerations and data-driven insights obtained from the correlation analysis. To minimize duplicates and prevent data leakage, we compared the feature matrices of all of the MOFs. Although some structures may appear similar based on their descriptors, their energetic characteristics, particularly the heat of adsorption values (Q_{st}), differ, ensuring that such similarities are not systematic. This approach is consistent with the recent recommendations for avoiding redundancy in MOF ML data sets.[58] To select the best ML algorithms and optimize the model parameters, we employed the Tree-Based Pipeline Optimization Tool (TPOT)[59] (version 0.12.2). The regression algorithms from the scikit-learn package[60] were used for model selection. To ensure consistent feature distributions between the train and test data sets, stratified sampling was used to split the data into 80% for the train and 20% for the test. 5-fold cross-validation was applied during model training to reduce the risk of overfitting. The performance of the ML models was assessed using three metrics: R^2, root-mean-square error (RMSE), and mean absolute error (MAE), with their formulas listed in Table S5. SHAP analysis[61] was performed to understand how each feature influenced the ML predictions in terms of the magnitude and direction.

We developed two different ML models to predict CH_4 and H_2 diffusivities in 3983 CoRE MOFs and then transferred them to an unseen set of 2788 hMOFs, 10% of the entire hMOF database, which were not included in the train set. These hMOFs were selected to represent the whole database and span the full range of structural (PLD and PV) and guest-related (CH_4 and H_2 diffusivities at infinite dilution) features of CoRE MOFs, as illustrated in Figure S5. The ML models demonstrated successful transferability to the unseen hMOF data set. We also developed three different ML models to predict CO_2, O_2, and N_2 diffusivities in CoRE MOFs, as presented in Figure S8(a−c), in which ML-predicted and simulated diffusivities were in good agreement. However, these models demonstrated limited transferability to unseen hMOF data sets because of the differences in feature distributions between CoRE MOFs and hMOFs, as shown in Figures S9−S11. Due to the absence of high diffusivities up to 6.3×10^{-4} cm^2/s in CoRE MOFs, compared to hMOFs having diffusivities up to 7.7×10^{-3} cm^2/s, 621, 225, and 257 hMOFs having CO_2, O_2, and N_2 diffusivities in the ranges of 3.2×10^{-4}−3.1×10^{-3}, 1.0×10^{-3}−6.4×10^{-3}, and 1.0×10^{-3}−7.7×10^{-3} cm^2/s were selected from the 2370 hMOF set. This set was generated by sorting PLDs and selecting every 10th structure to capture the overall feature distribution of the hMOF database. These hMOFs were added to the train set ($hMOFs_t$) to develop new models that were successfully transferred to unseen 4080, 4476, and 4444 hMOFs for predicting CO_2, O_2, and N_2 diffusivities, respectively. Figures S12−S14 present the distributions of structural and guest-related features for the new training and test sets of seen MOFs and unseen hMOFs. We note that the unseen hMOFs were selected to have high porosities (0.8−0.95) and LCDs (>20 Å) to cover materials with potentially high diffusivities. All the details and optimized hyperparameters of the final ML models are presented in Table S6.

Following the development of gas-specific ML models, we also developed a universal ML model capable of generalizing diffusivity predictions across all five gases by using the same feature set with two extra guest-related features: σ and MW of the gas species. Since the gases with the largest and smallest kinetic diameters in our work are CH_4 and H_2, we utilized their diffusivities in CoRE MOFs to develop the universal model and then tested it on unseen hMOFs to predict diffusivities of unseen gases: CO_2, O_2, and N_2. The details and optimized hyperparameters of the universal ML model are presented in Table S7.

We also analyzed the features used to develop our models. Including $\log D_0$ as a feature provided a reasonable simulation cost and allowed us to avoid additional MD simulations at higher guest loadings, which are much more computationally demanding due to the extensive calculations required for guest−guest and MOF−guest interactions. While D_0 and D are mathematically related, D_0 alone does not capture framework−guest and guest−guest interactions at finite loadings. To show this, we also trained the models without using $\log D_0$ as a feature. As shown in Figure S15, the models could still capture general trends but failed to reproduce them quantitatively for the test data, yielding much lower R^2 values of 0.88, 0.81, 0.78, 0.73, and 0.61 for CO_2, O_2, N_2, CH_4, and H_2, respectively, compared to the corresponding R^2 values of 0.93, 0.97, 0.97, 0.93, and 0.88 obtained when $\log D_0$ was included as a feature. When $\log D_0$ was excluded, the heat of adsorption and porosity emerged as the most important features for predicting the CO_2 and O_2 diffusivities. For CH_4 and N_2, the heat of adsorption was found to be the key descriptor, and the importance of pore sizes increased. For H_2, the significance of the PLD increased markedly, emerging as the dominant descriptor. These results suggest that while the model remains physically interpretable in the absence of D_0, with other features partially accounting for finite-loading effects, D_0 is a critical descriptor for achieving accurate diffusivity predictions in MOFs.

We acknowledge that MD-derived diffusivities can differ from the experimental values. Our models were trained and validated within a self-consistent computational framework, where all input features and target properties were derived from the fully atomistic simulations, GCMC and MD, using the same level of theory. Because the model has implicitly learned relationships among features that are internally consistent within this simulation domain, combining experimental and simulated inputs without retraining models may compromise predictive reliability. For this reason, we recommend employing D_0 values obtained from simulations when using the models. We note that experimentally measured D_0 data in the literature are scarce and therefore provided the ranges of D_0 values that we computed for each gas in Table S8.

Finally, we used the ML-predicted gas diffusivities to compute the permeabilities of CO_2, O_2, N_2, CH_4, and H_2 gases through 3987 CoRE MOFs and 14,180 hMOFs for assessing the membrane-based gas separation performance of the MOF space. ML-predicted permeabilities were compared with the simulated permeabilities for a total of 18,167 MOFs. Simulated permeabilities were computed using $P_i = N_i \times D_i/f_i$, where N_i, D_i, and f_i represent the adsorbed loading at 1 bar (obtained from GCMC simulations), the corresponding self-diffusivity (obtained from MD simulations), and the feed-side pressure of gas i, 1 bar in this work. ML-predicted permeabilities were computed using the same expression combining previously reported ML-predicted gas uptakes[62,63] with the ML-predicted diffusivities obtained in this study. The membrane selectivities were calculated as the ratio of gas permeabilities, $S_{i/j}^{mem} = P_i/P_j$.

To identify the most promising MOFs for gas separations, we calculated the trade-off between selectivity and permeability using $TSP = \ln P_i \times S_{i/j}^{mem}$. The MOFs offering the highest TSP values for at least two different gas separations were analyzed in detail using a molecular fingerprint. We utilized the MACCS keys[64] substructure search method available in the RDKit tool[65] to analyze these MOFs. This method relies on a predefined set of 166 structural patterns expressed in Simplified Molecular-Input Line-Entry System (SMILES)[66] Arbitrary Target Specification (SMARTS)[67] notation. For each MOF, we generated a 166-bit binary fingerprint, where each bit represents the presence ("1") or absence ("0") of a specific substructure. We extracted the fingerprints of the promising MOFs to identify their distinguishing chemical features.

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

■ ASSOCIATED CONTENT

Data Availability Statement

All structural and guest-related features of MOFs together with ML scripts developed in this work and all of the ML-predicted data are openly accessible on GitHub: https://github.com/hasancangulbalkan/MOFgasdiffusionpredictor. The web interface that we constructed is available on https://mofgasdiffusionpredictor.streamlit.app/.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsmaterialsau.5c00111.

Correlation heat map for structural and guest-related features of MOFs and self-diffusivities of CO_2, N_2, O_2, CH_4, and H_2; self-diffusivities of CO_2, O_2, N_2, CH_4, and H_2 gases in MOFs as functions of self-diffusivities computed at infinite dilution and porosity; SHAP feature importance distributions for gas-specific ML models; distributions of key structural, CO_2, O_2, N_2, CH_4, and H_2-related properties of MOFs; comparison of ML-predicted and simulated CO_2, O_2, N_2, CH_4, and H_2 permeabilities; structural properties of MOFs surpassing relevant upper bounds for each gas separation; comparison of ML-predicted and simulated CO_2, O_2, and N_2 diffusivities in MOFs and unseen hMOFs; number of MOFs utilized in gas-specific ML models; number of MOF membranes studied for each gas separation; structural and chemical check results for CoRE and hMOFs; details of the descriptors; metrics calculated to evaluate the ML models' accuracy; best ML pipelines and parameters for diffusivity predictions; and best ML pipeline and parameters for diffusivity predictions for the universal ML model (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Seda Keskin − *Department of Chemical and Biological Engineering, Koç University, 34450 Istanbul, Turkey;* orcid.org/0000-0001-5968-0336; Phone: +90 (212) 338-1362; Email: skeskin@ku.edu.tr

Authors

Pelin Sezgin − *Department of Chemical and Biological Engineering, Koç University, 34450 Istanbul, Turkey*

Feride Neva Yüngül − *Department of Chemical and Biological Engineering, Koç University, 34450 Istanbul, Turkey*

Beste Naz Karaca − *Department of Chemical and Biological Engineering, Koç University, 34450 Istanbul, Turkey*

Hasan Can Gulbalkan − *Department of Chemical and Biological Engineering, Koç University, 34450 Istanbul, Turkey;* orcid.org/0000-0001-8001-6018

Complete contact information is available at: https://pubs.acs.org/10.1021/acsmaterialsau.5c00111

Author Contributions

†P.S., F.N.Y., B.N.K., and H.C.G. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.K. acknowledges funding by the European Union (ERC, STARLET, 101124002). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

■ REFERENCES

(1) Howarth, A. J.; Liu, Y.; Li, P.; Li, Z.; Wang, T. C.; Hupp, J. T.; Farha, O. K. Chemical, thermal and mechanical stabilities of metal−organic frameworks. *Nat. Rev. Mater.* **2016**, *1* (3), No. 15018.

(2) Kim, M.; Choi, W.; Lee, C. H.; Kim, D. W. 2D MOFs and zeolites for composite membrane and gas separation applications: a brief review. *ACS Mater. Au* **2024**, *4* (2), 148−161.

(3) Qian, Q.; Asinger, P. A.; Lee, M. J.; Han, G.; Rodriguez, K. M.; Lin, S.; Benedetti, F. M.; Wu, A. X.; Chi, W. S.; Smith, Z. P. MOF-based membranes for gas separations. *Chem. Rev.* **2020**, *120* (16), 8161−8266.

(4) Ongari, D.; Talirz, L.; Smit, B. Too many materials and too many applications: an experimental problem waiting for a computational solution. *ACS Cent. Sci.* **2020**, *6* (11), 1890−1900.

(5) Formalik, F.; Shi, K.; Joodaki, F.; Wang, X.; Snurr, R. Q. Exploring the Structural, Dynamic, and Functional Properties of Metal-Organic Frameworks through Molecular Modeling. *Adv. Funct. Mater.* **2024**, *34* (43), No. 2308130.

(6) Azar, A. N. V.; Velioglu, S.; Keskin, S. Large-scale computational screening of metal organic framework (MOF) membranes and MOF-based polymer membranes for H_2/N_2 separations. *ACS Sustainable Chem. Eng.* **2019**, *7* (10), 9525−9536.

(7) Altintas, C.; Keskin, S. Molecular simulations of MOF membranes and performance predictions of MOF/polymer mixed matrix membranes for CO_2/CH_4 separations. *ACS Sustainable Chem. Eng.* **2019**, *7* (2), 2739−2750.

(8) Daglar, H.; Keskin, S. Computational screening of metal−organic frameworks for membrane-based $CO_2/N_2/H_2O$ separations: Best materials for flue gas separation. *J. Phys. Chem. C* **2018**, *122* (30), 17347−17357.

(9) Avci, G.; Velioglu, S.; Keskin, S. High-throughput screening of MOF adsorbents and membranes for H_2 purification and CO_2 capture. *ACS Appl. Mater. Interfaces* **2018**, *10* (39), 33693−33706.

(10) Altintas, C.; Avci, G.; Daglar, H.; Gulcay-Ozcan, E.; Erucar, I.; Keskin, S. Computer simulations of 4240 MOF membranes for H_2/CH_4 separations: Insights into structure−performance relations. *J. Mater. Chem. A* **2018**, *6* (14), 5836−5847.

(11) Qiao, Z.; Peng, C.; Zhou, J.; Jiang, J. High-throughput computational screening of 137953 metal−organic frameworks for membrane separation of a $CO_2/N_2/CH_4$ mixture. *J. Mater. Chem. A* **2016**, *4* (41), 15904−15912.

(12) Qiao, Z.; Xu, Q.; Jiang, J. High-throughput computational screening of metal-organic framework membranes for upgrading of natural gas. *J. Membr. Sci.* **2018**, *551*, 47−54.

(13) Chong, S.; Lee, S.; Kim, B.; Kim, J. Applications of machine learning in metal-organic frameworks. *Coord. Chem. Rev.* **2020**, *423*, No. 213487.

(14) Zheng, Z.; Zhang, O.; Nguyen, H. L.; Rampal, N.; Alawadhi, A. H.; Rong, Z.; Head-Gordon, T.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. Chatgpt research group for optimizing the crystallinity of MOFs and COFs. *ACS Cent. Sci.* **2023**, *9* (11), 2161−2170.

(15) Sriram, A.; Choi, S.; Yu, X.; Brabson, L. M.; Das, A.; Ulissi, Z.; Uyttendaele, M.; Medford, A. J.; Sholl, D. S. The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. *ACS Cent. Sci.* **2024**, *10*, 923−941.

(16) Liu, S.; Dupuis, R.; Fan, D.; Benzaria, S.; Bonneau, M.; Bhatt, P.; Eddaoudi, M.; Maurin, G. Machine learning potential for modelling H_2 adsorption/diffusion in MOFs with open metal sites. *Chem. Sci.* **2024**, *15* (14), 5294−5302.

(17) Lim, Y.; Kim, J. Application of transfer learning to predict diffusion properties in metal−organic frameworks. *Mol. Syst. Des. Eng.* **2022**, *7* (9), 1056−1064.
(18) Orhan, I. B.; Daglar, H.; Keskin, S.; Le, T. C.; Babarao, R. Prediction of O_2/N_2 selectivity in metal−organic frameworks via high-throughput computational screening and machine learning. *ACS Appl. Mater. Interfaces* **2022**, *14* (1), 736−749.
(19) Yang, Y.; Yu, Z.; Sholl, D. S. Machine learning models for predicting molecular diffusion in metal−organic frameworks accounting for the impact of framework flexibility. *Chem. Mater.* **2023**, *35* (23), 10156−10168.
(20) Guo, S.; Huang, X.; Situ, Y.; Huang, Q.; Guan, K.; Huang, J.; Wang, W.; Bai, X.; Liu, Z.; Wu, Y.; Qiao, Z. Interpretable machine-learning and big data mining to predict gas diffusivity in metal-organic frameworks. *Adv. Sci.* **2023**, *10* (21), No. 2301461.
(21) Daglar, H.; Keskin, S. Combining machine learning and molecular simulations to unlock gas separation potentials of MOF membranes and MOF/polymer MMMs. *ACS Appl. Mater. Interfaces* **2022**, *14* (28), 32134−32148.
(22) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal−Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985−5998.
(23) Majumdar, S.; Moosavi, S. M.; Jablonka, K. M.; Ongari, D.; Smit, B. Diversifying databases of metal organic frameworks for high-throughput computational screening. *ACS Appl. Mater. Interfaces* **2021**, *13* (51), 61004−61014.
(24) Zhao, G.; Brabson, L. M.; Chheda, S.; Huang, J.; Kim, H.; Liu, K.; Mochida, K.; Pham, T. D.; Terrones, G. G.; Yoon, S.; et al. CoRE MOF DB: A curated experimental metal-organic framework database with machine-learned properties for integrated material-process screening. *Matter* **2025**, *8*, No. 102140.
(25) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal−Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985−5998.
(26) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134−141.
(27) Rappe, A. K.; Goddard III, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem. A* **1991**, *95* (8), 3358−3363.
(28) Wilmer, C. E.; Snurr, R. Q. Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: Calculation of framework charges via charge equilibration. *Chem. Eng. J.* **2011**, *171* (3), 775−781.
(29) Meduri, S.; Nandanavanam, J. Prediction of hydrogen uptake of metal organic frameworks using explainable machine learning. *Energy and AI* **2023**, *12*, No. 100230.
(30) Robeson, L. M. The upper bound revisited. *J. Membr. Sci.* **2008**, *320* (1−2), 390−400.
(31) Comesaña-Gándara, B.; Chen, J.; Bezzu, C. G.; Carta, M.; Rose, I.; Ferrari, M.-C.; Esposito, E.; Fuoco, A.; Jansen, J. C.; McKeown, N. B. Redefining the Robeson upper bounds for CO_2/CH_4 and CO_2/N_2 separations using a series of ultrapermeable benzotriptycene-based polymers of intrinsic microporosity. *Energy Environ. Sci.* **2019**, *12* (9), 2733−2740.
(32) Liu, Y.; Ng, Z.; Khan, E. A.; Jeong, H.-K.; Ching, C.-b.; Lai, Z. Synthesis of continuous MOF-5 membranes on porous α-alumina substrates. *Microporous Mesoporous Mater.* **2009**, *118* (1−3), 296−301.
(33) Keskin, S.; Sholl, D. S. Screening metal− organic framework materials for membrane-based methane/carbon dioxide separations. *J. Phys. Chem. C* **2007**, *111* (38), 14055−14059.
(34) Lee, D.-J.; Li, Q.; Kim, H.; Lee, K. Preparation of Ni-MOF-74 membrane for CO_2 separation by layer-by-layer seeding technique. *Microporous Mesoporous Mater.* **2012**, *163*, 169−177.
(35) Guo, H.; Zhu, G.; Hewitt, I. J.; Qiu, S. Twin copper source" growth of metal− organic framework membrane: $Cu_3(BTC)_2$ with high permeability and selectivity for recycling H_2. *J. Am. Chem. Soc.* **2009**, *131* (5), 1646−1647.
(36) Li, S.; Han, W.; An, Q. F.; Yong, K. T.; Yin, M. J. Defect engineering of MOF-based membrane for gas separation. *Adv. Funct. Mater.* **2023**, *33* (38), No. 2303447.
(37) Demir, H.; Keskin, S. A New Era of Modeling MOF-Based Membranes: Cooperation of Theory and Data Science. *Macromol. Mater. Eng.* **2024**, *309* (1), No. 2300225.
(38) White, A. J.; Gibaldi, M.; Burner, J.; Mayo, R. A.; Woo, T. K. High Structural Error Rates in "Computation-Ready" MOF Databases Discovered by Checking Metal Oxidation States. *J. Am. Chem. Soc.* **2025**, *147*, 17579−17583.
(39) Chen, T.; Manz, T. A. Identifying misbonded atoms in the 2019 CoRE metal−organic framework database. *RSC Adv.* **2020**, *10* (45), 26944−26951.
(40) Jin, X.; Jablonka, K. M.; Moubarak, E.; Li, Y.; Smit, B. MOFChecker: a package for validating and correcting metal−organic framework (MOF) structures. *Digital Discovery* **2025**, *4*, 1560−1569.
(41) Dubbeldam, D.; Calero, S.; Vlugt, T. J. iRASPA: GPU-accelerated visualization software for materials scientists. *Mol. Simul.* **2018**, *44* (8), 653−676.
(42) Mahajan, S.; Lahtinen, M. Recent progress in metal-organic frameworks (MOFs) for CO_2 capture at different pressures. *J. Environ. Chem. Eng.* **2022**, *10* (6), No. 108930.
(43) Queen, W. L.; Hudson, M. R.; Bloch, E. D.; Mason, J. A.; Gonzalez, M. I.; Lee, J. S.; Gygi, D.; Howe, J. D.; Lee, K.; Darwish, T. A.; et al. Comprehensive study of carbon dioxide adsorption in the metal−organic frameworks M_2 (dobdc) (M= Mg, Mn, Fe, Co, Ni, Cu, Zn). *Chem. Sci.* **2014**, *5* (12), 4569−4581.
(44) Park, H.; Yan, X.; Zhu, R.; Huerta, E. A.; Chaudhuri, S.; Cooper, D.; Foster, I.; Tajkhorshid, E. A generative artificial intelligence framework based on a molecular diffusion model for the design of metal-organic frameworks for carbon capture. *Commun. Chem.* **2024**, *7* (1), No. 21.
(45) Ashworth, D. J.; Roseveare, T. M.; Schneemann, A.; Flint, M.; Bernáldes, I. D.; Vervoorts, P.; Fischer, R. A.; Brammer, L.; Foster, J. A. Increasing alkyl chain length in a series of layered metal−organic frameworks aids ultrasonic exfoliation to form nanosheets. *Inorg. Chem.* **2019**, *58* (16), 10837−10845.
(46) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42* (2), 81−101.
(47) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1.United-atom description of n-alkanes. *J. Phys. Chem. B* **1998**, *102* (14), 2569−2577.
(48) Buch, V. Path integral simulations of mixed para-D2 and ortho-D2 clusters: The orientational effects. *J. Chem. Phys.* **1994**, *100* (10), 7610−7629.
(49) Eggimann, B. L.; Sunnarborg, A. J.; Stern, H. D.; Bliss, A. P.; Siepmann, J. I. An online parameter and property database for the TraPPE force field. *Mol. Simul.* **2014**, *40* (1−3), 101−105.
(50) Murthy, C.; Singer, K.; Klein, M.; McDonald, I. Pairwise additive effective potentials for nitrogen. *Mol. Phys.* **1980**, *41* (6), 1387−1399.
(51) Mellot, C.; Lignieres, J. Monte Carlo simulations of N_2 and O_2 adsorption in silicalite and CaLSX zeolites. *Mol. Simul.* **1997**, *18* (6), 349−365.
(52) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard, W. A., III; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024−10035.
(53) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369* (3), 253−287.

https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153

(54) Frenkel, D.; Smit, B. *Understanding molecular simulation*; Academic Press: San Diego, 2002; Vol. 2, pp 1−6.
(55) Evans, D. J.; Holian, B. L. The nose-hoover thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069−4074.
(56) Einstein, A. *Investigations on the Theory of the Brownian Movement*; Courier Corporation, 1956.
(57) Daglar, H.; Erucar, I.; Keskin, S. Exploring the performance limits of MOF/polymer MMMs for O_2/N_2 separation using computational screening. *J. Membr. Sci.* **2021**, *618*, No. 118555.
(58) Jablonka, K. M.; Rosen, A. S.; Krishnapriyan, A. S.; Smit, B. An ecosystem for digital reticular chemistry. *ACS Cent. Sci.* **2023**, *9*, 563−581.
(59) Le, T. T.; Fu, W.; Moore, J. H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* **2020**, *36* (1), 250−256.
(60) Martinsson, P.-G.; Rokhlin, V.; Tygert, M. A randomized algorithm for the decomposition of matrices. *Appl. Comput. Harmon. Anal.* **2011**, *30* (1), 47−68.
(61) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2* (1), 56−67.
(62) Sezgin, P.; Keskin, S. Integrating Molecular Simulations with Machine Learning to Discover Selective MOFs for CH_4/H_2 Separation. *J. Phys. Chem. C* **2025**, *129*, 13089−13099.
(63) Gulbalkan, H. C. MOFs_CO2N2O2_ML (2025), GitHub repository, https://github.com/hasancangulbalkan/MOFs_CO2N2O2_ML.
(64) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273−1280.
(65) *RDKit: Open-Source Cheminformatics Software, 2022*, http://www.rdkit.org. (accessed May 15, 2025).
(66) Weininger, D. SMILES-A Language for Molecules and Reactions. *Handbook of Chemoinformatics: From Data to Knowledge in 4 Volumes* **2003**, 80−102.
(67) Daylight Chemical Information Systems, I. SMARTS-A Language for Describing Molecular Patterns. 2019.



https://doi.org/10.1021/acsmaterialsau.5c00111
ACS Mater. Au 2026, 6, 140−153