

Supporting Information for Interpretable Graph Transformer Network for Predicting Adsorption Isotherms of Metal-Organic Frameworks

Pin Chen,^{†,§} Rui Jiao,^{‡,¶,§} Jinyu Liu,[†] Yang Liu,^{*,‡,¶} and Yutong Lu^{*,†}

[†]*National Supercomputer Center in Guangzhou, School of Computer Science and Engineering, Sun Yat-sen University, 132 East Circle at University City, Guangzhou, 510006, P. R. China.*

[‡]*Department of Computer Science and Technology, Tsinghua University, Haidian District, Beijing, 100084, P. R. China*

[¶]*Institute for AI Industry Research, Tsinghua University, Haidian District, Beijing, 100084, P. R. China.*

[§]*These authors contributed equally to this work.*

E-mail: liuyang2011@tsinghua.edu.cn; luyutong@mail.sysu.edu.cn

Global features

The crystal geometric attributes, e.g., gravimetric surface area (GSA), largest cavity diameter (LCD), pore-limiting diameter (PLD), volumetric surface area (VSA), channel number (CN), crystal density, pore volume (PV), void fraction (VF) and open metal site per cell (OMS), are used in our model. These features were calculated using open-source

Table S1: The dataset processing workflow and relevant dataset size.

Processing workflow	Dataset size		
	N ₂	CO ₂	CH ₄
Adsorption gas			
MOFs data in CSD database		72,618	
After filtering by PLDs	10,701	13,203	9628
After removing GCMC failed MOFs	7304	6997	8539

Table S2: The detailed information for local and global environment features.

Local Node features	Size	Description
Atom symbol	82	Covered most elements in element periodic table (one-hot)
Connectivity	8	The attributes of a point include number of atoms connected to the point(one-hot)
Hydrogen bond acceptor	1	Whether it is a hydrogen bond acceptor(bool)
Chirality	1	Whether it has chirality(bool)
Ring	1	Whether it belongs to the ring system(bool)
Metal	1	Whether it is a metal(bool)
Spiro	1	Whether it is a spiro(bool)
Local Edge features	Size	Description
Bond length	1	Atom distance(float)
Global features	Size	Description
GSA	1	Gravimetric surface area (float)
LCD	1	Largest cavity diameter (float)
PLD	1	Pore-limiting diameter (float)
VSA	1	Volumetric surface area (float)
CN	1	Channel number per cell (integer)
Density	1	Crystal Density (float)
PV	1	Pore volume per cell (float)
VF	1	Void fraction (float)
OMS	1	Open metal site per cell (float)

software Zeo++,¹ and the relevant generated code is released in a Github repository at <https://github.com/Matgen-project/MOFNet>. The features of GSA, LCD, PLD, VAS, Density and VF do not vary with the size of the supercells, and the CN, PV and OMS only obtain values from one cell. More descriptions of features can be found at <http://www.zeoplusplus.org/about.html>.

CSD-MOFDB dataset evaluation

We show the distribution of calculated VAS and GSA in our calculated dataset against the dataset of hypothetical MOF (hMOF) from Berner et al² in Figure S3. Compared with hMOF dataset, the plots show that the structures from our calculated database have a denser distribution for the low value of VSA and GSA. There are about 90.75% of VSA less than 2000 m²/cm³ and 67.76% of GSA less than 1000 m²/g in our calculated database, while 88.26% of VSA more than 2000 m²/cm³ and 86.89% of GSA more than 1000 m²/g in hypothetical MOF dataset. The CSD codes of top-performing structures in our calculated database for N₂, CO₂ and CH₄ are HIWZAP, XAHQAA and BAZGAM at high pressure of 80 kPa, 20,000 kPa and 10,000 kPa, respectively. The trends for how N₂, CO₂ and CH₄ uptake capacities correlate with different textural properties at high pressure are shown in Figure S4.

We evaluated the accuracy of our calculated dataset by comparing with another dataset. As shown in Figure S5, we compare the N₂ uptake capacities between CoRE MOF 2019-ASR³ and our calculated dataset at the same temperature and pressure. We achieved MAE performance of 2.884 mol/kg and PCC performance of 0.887, which proves the rationality of our data in structure collection workflow and adsorption calculation method. Though we used the same force field and charge calculation method as CoRE MOF 2019-ASR, the GCMC method contains a random initialization algorithm that may cause deviations. Our dataset overlaps with CoRE MOF 2019-ASR, while there are still data differences. As the

data sources of CoRE MOF 2019-ASR are extensive, which includes CoRE MOF users, CSD database, Web of Science searching, and disordered structures reconstruction. In the calculation of the adsorption properties of N_2 , there are about 6000 structures that are not in our data set. As a consequence, our dataset can be used as an alternative to CoRE MOF.

All the calculated data are open in public in our website¹. One can click the name of elements from the periodic table, or input the CSD code to search materials. After selecting one material in the result list, the MOF structure can be viewed by our previous developed software, named 3DStructGen.⁴ Moreover, the geometry related computational properties and different gas adsorption isotherms are shown in the webpage.

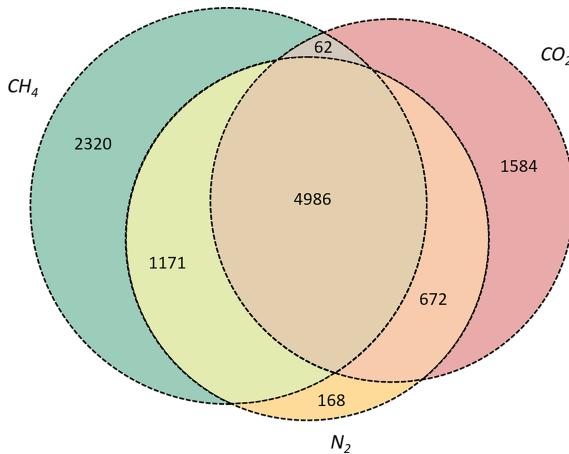


Figure S1: The structure distribution of dataset for N_2 , CO_2 and CH_4 , respectively.

Methods for all machine learning (ML) algorithms

We have provided the details of the data splitting scheme and the searched hyperparameters of MOFNet in Table S3 and Table S4. The traditional ML framework was built by scikit-learn package. Four ML algorithms, including Support Vector Regression (SVR), Decision Trees (DT), Gradient Boosted Regression Trees (GBRT) and Random Forests (RF), were designed for training and testing on the dataset, which only contains global features. The

¹<https://matgen.nscc-gz.cn/mof.html>

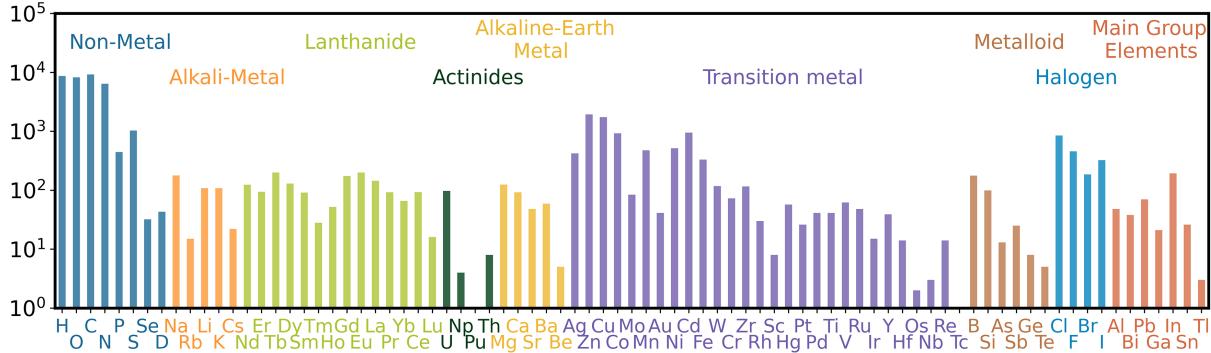


Figure S2: Number of appearances of various chemical species (81 element types in total) in the 9399 MOFs of our database examined in this study. Notice the logarithmic scale in the y-axis.

main hyperparameters set in SVR, DT, RF and GBRT are shown in Table S5. We have provided more hyperparameters for training the equivariant GNN-based baselines in Table S6. Moreover, we (1) adopt the same 10-fold data splits as the MOFNet model, (2) utilize the same atomic features as node-level inputs, and (3) combine each baseline model with an additional MLP to encode global features for fair comparison with our model.

Table S3: The searched hyperparameters of MOFNet model. The **bold** combination leads to the best validation performance.

Hyperparameter	Description	Range
d_model	The size of hidden states in graph transformer	128, 512, 1024
N	The number of layers in graph transformer network	2 , 3, 4
h	The number of heads in the multi-head attention mechanism	4,8, 16
n_rbf	The number of Bessel basis functions	20 , 25
warmup_steps	The initial steps for linear warmup	2000 , 4000
dropout	The dropout ratio	0.0, 0.1

Model performance on simulation dataset.

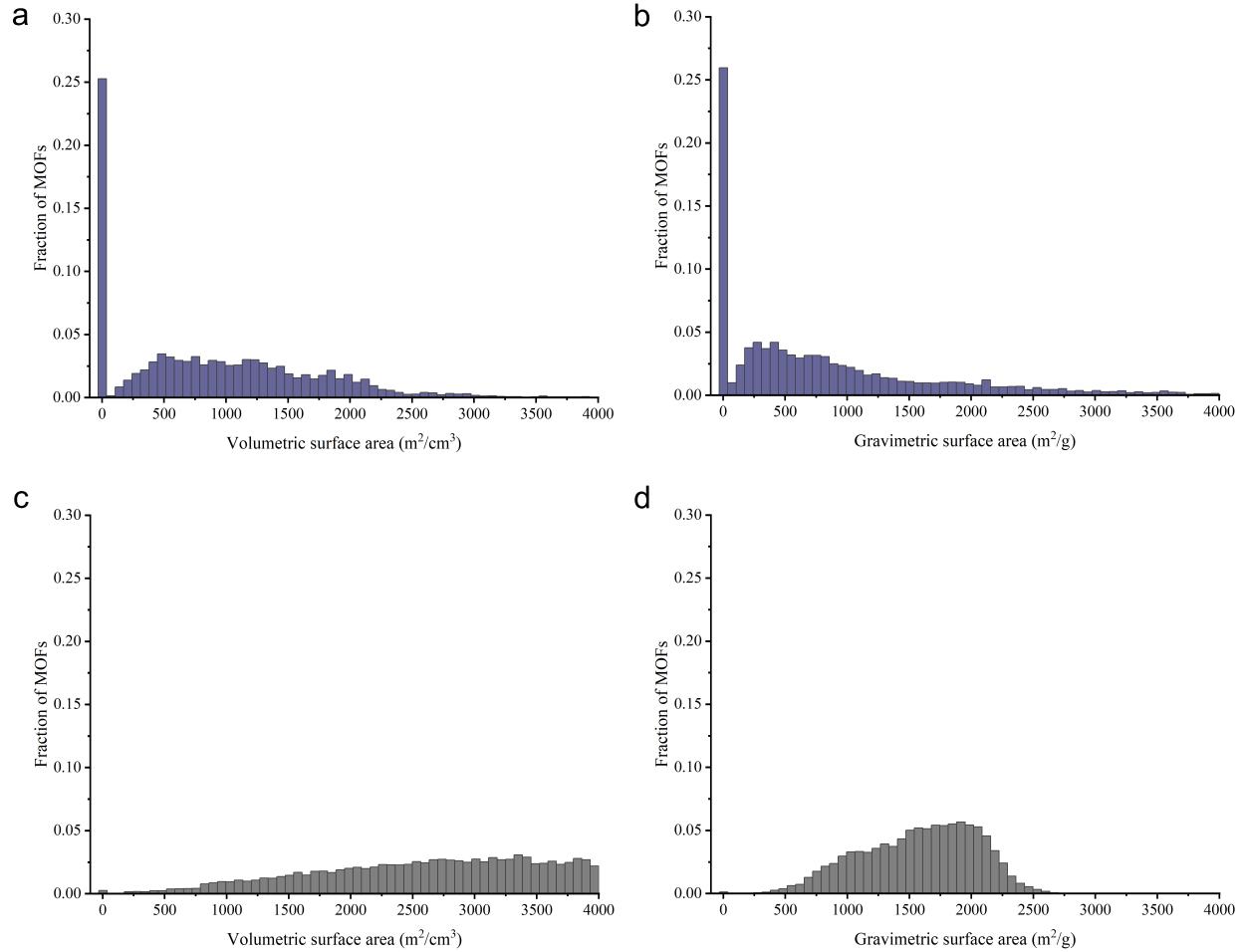


Figure S3: Probability distribution of calculated volumetric and gravimetric surface area (VSA) and gravimetric surface area (GSA): **a**, VSA from our calculated database; **b**, GSA from our calculated database; **c**, VSA from the hypothetical MOF database of Berner et al.; **d**, GSA from the hypothetical MOF database of Berner et al.

Table S4: All datasets and exact number of training/validation/test data sizes used in paper.

Data type	Data size	Training/validation/test data size
N_2 adsorption values in CSD-MOFDB	7304	5844/730/730
CO_2 adsorption values in CSD-MOFDB	6997	5599/699/699
CH_4 adsorption values in CSD-MOFDBs	8539	6833/853/853
$\text{N}_2/\text{CO}_2/\text{CH}_4$ adsorption isotherms in NIST-ISODB	54	-

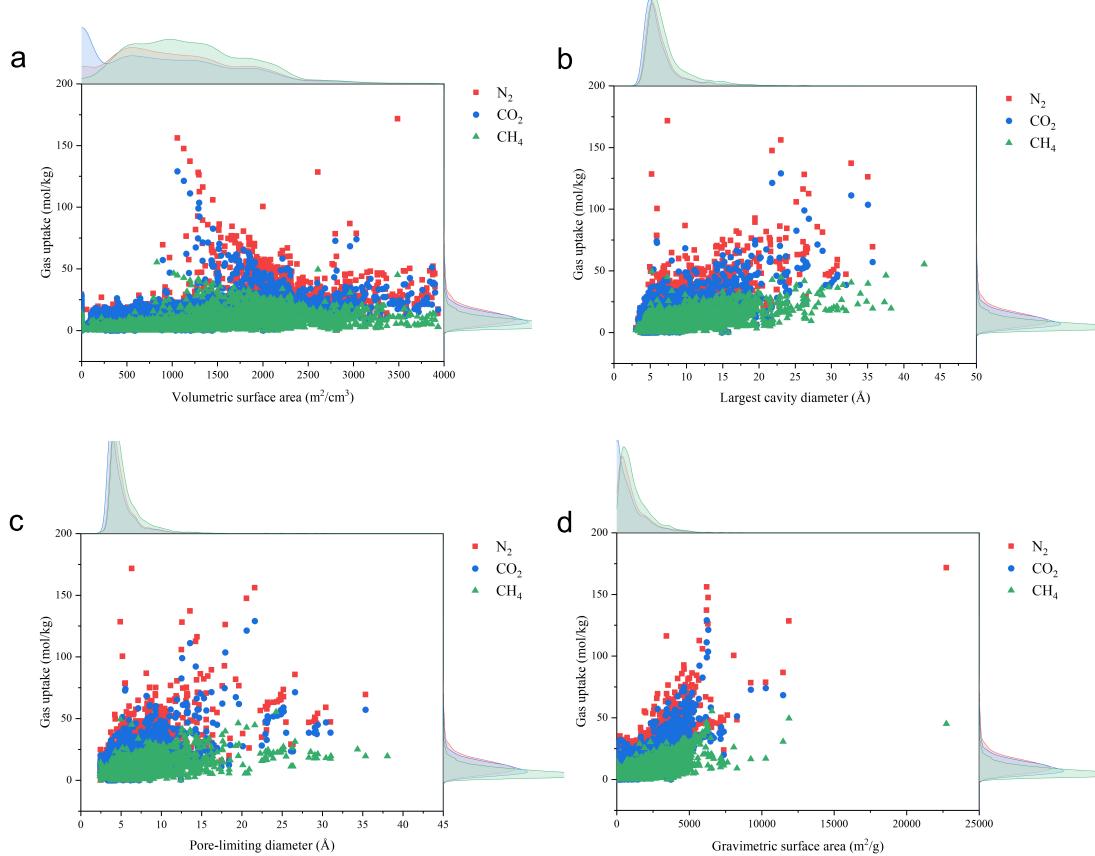


Figure S4: Uptake loading of our calculated database for CH_4 , N_2 and CO_2 at the highest pressure against **a**, volumetric surface area, **b**, largest cavity diameter, **c**, pore-limiting diameter and **d**, gravimetric surface area.

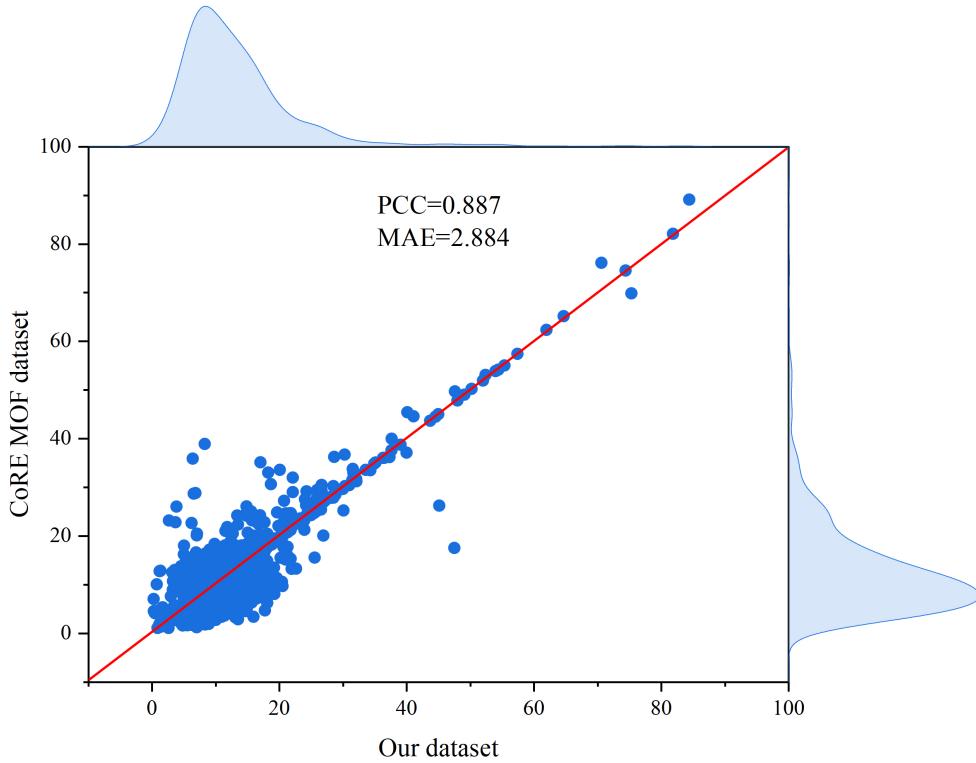


Figure S5: Comparison between the N_2 uptake loading of CoRE MOF 2019-ASR and our data set at the same temperature and pressure. Here, the temperature is 77 K, while the pressure is selected from the overlap of these two datasets.

Table S5: The key hyperparameters of four traditional ML algorithms.

Algorithm	Name	Description	value
SVR	kernel	The type of kernel	'rbf'
	degree	The degree of polynomial kernel function	3
	gamma	The coefficients of kernel	'scale'
DT	criterion	The function to measure the quality of a split	'gini'
	splitter	The strategy used to choose the split at each node	'best'
	min_samples_split	The minimum number of samples required to split an internal node	2
	min_samples_leaf	The minimum number of samples required to be at a leaf node	1
GBRT	n_estimators	The number of boosting stages to perform	100
	criterion	The function to measure the quality of a split	'friedman_mse'
	min_samples_split	The minimum number of samples required to split an internal node	2
	min_samples_leaf	The minimum number of samples required to be at a leaf node.	1
	max_depth	The maximum depth of the individual regression estimators	3
	The loss	loss function to be optimized	'ls'
	learning_rate	learning rate shrinks the contribution of each tree	0.1
RF	n_estimators	The number of submodels	100
	criterion	The measure the quality of a split	'mse'
	min_samples_split	The minimum number of samples required to split an internal node	2
	min_samples_leaf	The minimum number of samples required to be at a leaf node	1
	max_features	The number of features to consider when looking for the best split	'auto'

Table S6: The key hyperparameters of four GNN-based algorithms.

Algorithm	Name	Description	value
SchNet	n_atom_basis	The size of atomic embeddings and hidden states	128
	n_filters	The number of filters in the continuous-filter convolution	128
	n_interactions	The number of the interaction blocks	3
	n_gaussians	The number of Gaussian functions in the RBF kernel	25
DimeNet++	hidden_channels	The size of hidden states	128
	num_blocks	The number of the interaction blocks	3
	int_emb_size	The embedding size in the interaction blocks	64
	basis_emb_size	The basis embedding size in the interaction blocks	8
	out_emb_channels	The embedding size in the output blocks	256
	num_spherical	The number of spherical harmonics	7
	num_radial	The number of radial basis functions	6
	act	The activation function	'swish'
	hidden_nf	The size of hidden states	128
EGNN	n_layers	The number of layers	3
	act_fn	The activation function	'silu'
	activation	The activation function	'silu'
PaiNN	n_atom_basis	The size of atomic embeddings and hidden states	128
	n_interactions	The number of the interaction blocks	3
	n_rbf	The number of Bessel basis functions	20
	activation	The activation function	'silu'

 Table S7: All models predicting performance (sMAPE) for N₂ uptake at different pressures.

Model	Pressure (kPa)							
	0.2	5	10	20	40	60	80	100
SVR	0.289±0.013	0.251±0.008	0.247±0.008	0.244±0.009	0.242±0.009	0.241±0.009	0.241±0.009	0.240±0.009
DT	0.355±0.017	0.312±0.014	0.311±0.010	0.304±0.013	0.305±0.008	0.302±0.007	0.297±0.012	0.305±0.010
GBRT	0.296±0.010	0.258±0.008	0.254±0.008	0.251±0.009	0.249±0.009	0.248±0.009	0.247±0.009	0.246±0.009
RF	0.273±0.010	0.238±0.006	0.234±0.007	0.231±0.007	0.229±0.007	0.227±0.006	0.227±0.007	0.227±0.007
SchNet	0.293±0.011	0.257±0.011	0.251±0.011	0.247±0.012	0.245±0.012	0.244±0.011	0.243±0.013	0.244±0.010
PaiNN	0.281±0.015	0.252±0.015	0.252±0.013	0.250±0.014	0.244±0.014	0.241±0.013	0.244±0.012	0.240±0.010
EGNN	0.295±0.012	0.254±0.010	0.250±0.009	0.249±0.010	0.246±0.011	0.246±0.009	0.246±0.013	0.245±0.010
DimeNet++	0.276±0.010	0.237±0.008	0.237±0.014	0.228±0.008	0.231±0.008	0.228±0.011	0.230±0.010	0.227±0.010
MOFNet-L	0.377±0.012	0.377±0.009	0.376±0.011	0.379±0.007	0.375±0.010	0.379±0.009	0.377±0.009	0.379±0.009
MOFNet-G	0.289±0.011	0.250±0.010	0.246±0.009	0.243±0.010	0.242±0.011	0.241±0.010	0.242±0.010	0.241±0.010
MOFNet	0.251±0.013	0.221±0.011	0.220±0.009	0.216±0.011	0.215±0.009	0.215±0.010	0.215±0.012	0.215±0.011

Table S8: All models predicting performance (sMAPE) for CO₂ uptake at different pressures.

Model	Pressure (kPa)							
	10	50	100	500	1000	1500	20000	50000
SVR	0.691±0.016	0.508±0.014	0.442±0.014	0.359±0.011	0.328±0.012	0.318±0.011	0.303±0.010	0.296±0.010
DT	0.732±0.017	0.583±0.018	0.509±0.019	0.424±0.013	0.388±0.019	0.378±0.017	0.361±0.018	0.350±0.016
GBRT	0.733±0.014	0.531±0.014	0.459±0.014	0.372±0.012	0.341±0.013	0.329±0.013	0.312±0.012	0.304±0.011
RF	0.671±0.011	0.488±0.012	0.420±0.011	0.342±0.008	0.311±0.008	0.301±0.008	0.286±0.008	0.278±0.008
SchNet	0.695±0.029	0.496±0.020	0.429±0.017	0.359±0.009	0.328±0.008	0.319±0.009	0.302±0.006	0.296±0.009
PaiNN	0.632±0.030	0.476±0.023	0.413±0.015	0.348±0.009	0.316±0.014	0.307±0.014	0.288±0.012	0.282±0.009
EGNN	0.697±0.022	0.502±0.011	0.439±0.012	0.356±0.008	0.331±0.009	0.322±0.009	0.303±0.008	0.297±0.008
DimeNet++	0.595±0.022	0.447±0.023	0.389±0.018	0.331±0.014	0.303±0.008	0.291±0.012	0.277±0.010	0.272±0.009
MOFNet-L	0.770±0.024	0.536±0.012	0.457±0.016	0.405±0.013	0.392±0.006	0.389±0.012	0.388±0.012	0.382±0.007
MOFNet-G	0.720±0.016	0.518±0.015	0.441±0.012	0.362±0.010	0.327±0.010	0.316±0.011	0.300±0.010	0.293±0.010
MOFNet	0.569±0.013	0.424±0.015	0.370±0.012	0.311±0.011	0.282±0.012	0.274±0.012	0.263±0.012	0.258±0.012

Table S9: All models predicting performance (sMAPE) for CH₄ at different pressures.

Model	Pressure (kPa)							
	50	100	150	200	500	1000	5000	10000
SVR	0.462±0.013	0.416±0.011	0.388±0.010	0.367±0.010	0.304±0.010	0.262±0.010	0.200±0.009	0.186±0.008
DT	0.548±0.012	0.507±0.013	0.467±0.016	0.444±0.020	0.373±0.012	0.334±0.010	0.266±0.007	0.249±0.010
GBRT	0.496±0.014	0.441±0.012	0.410±0.011	0.387±0.008	0.322±0.010	0.280±0.009	0.212±0.009	0.198±0.008
RF	0.443±0.014	0.395±0.011	0.369±0.011	0.348±0.010	0.287±0.009	0.252±0.009	0.194±0.008	0.182±0.008
SchNet	0.444±0.017	0.393±0.017	0.351±0.017	0.334±0.019	0.276±0.016	0.245±0.018	0.198±0.015	0.190±0.015
PaiNN	0.402±0.023	0.358±0.017	0.327±0.018	0.309±0.017	0.257±0.015	0.231±0.016	0.194±0.016	0.189±0.015
EGNN	0.445±0.018	0.389±0.016	0.359±0.009	0.334±0.012	0.273±0.015	0.243±0.016	0.197±0.011	0.188±0.011
DimeNet++	0.398±0.015	0.344±0.018	0.321±0.018	0.296±0.019	0.250±0.014	0.224±0.014	0.187±0.014	0.187±0.017
MOFNet-L	0.515±0.022	0.446±0.013	0.419±0.018	0.405±0.010	0.366±0.018	0.360±0.007	0.385±0.012	0.406±0.009
MOFNet-G	0.471±0.013	0.420±0.009	0.388±0.008	0.368±0.009	0.303±0.006	0.263±0.006	0.203±0.005	0.190±0.006
MOFNet	0.359±0.010	0.320±0.010	0.295±0.007	0.281±0.008	0.233±0.006	0.211±0.006	0.176±0.005	0.171±0.007

Table S10: All models predicting performance (MAE) for N₂ uptake at different pressures.

Model	Pressure (kPa)								
	0.2	5	10	20	40	60	80	100	
SVR	2.711±0.124	2.514±0.089	2.489±0.083	2.469±0.092	2.461±0.089	2.459±0.091	2.454±0.096	2.454±0.096	
DT	3.360±0.133	3.156±0.124	3.144±0.108	3.072±0.117	3.096±0.083	3.094±0.096	3.039±0.123	3.127±0.106	
GBRT	2.739±0.071	2.548±0.042	2.513±0.040	2.490±0.053	2.479±0.050	2.475±0.051	2.467±0.058	2.465±0.051	
RF	2.539±0.064	2.387±0.067	2.343±0.049	2.323±0.060	2.306±0.060	2.297±0.055	2.304±0.064	2.294±0.059	
SchNet	2.677±0.076	2.535±0.116	2.477±0.078	2.441±0.089	2.451±0.093	2.446±0.093	2.424±0.112	2.443±0.086	
PaiNN	2.606±0.123	2.498±0.125	2.507±0.091	2.484±0.127	2.454±0.095	2.409±0.118	2.437±0.104	2.403±0.092	
EGNN	2.720±0.092	2.524±0.083	2.472±0.065	2.473±0.072	2.451±0.095	2.450±0.073	2.462±0.091	2.441±0.080	
DimeNet++	2.553±0.090	2.365±0.097	2.365±0.113	2.287±0.066	2.314±0.095	2.300±0.117	2.324±0.080	2.294±0.090	
MOFNet-L	3.816±0.163	4.392±0.148	4.445±0.166	4.564±0.189	4.561±0.197	4.621±0.208	4.603±0.173	4.680±0.174	
MOFNet-G	2.700±0.067	2.487±0.075	2.440±0.054	2.426±0.076	2.421±0.069	2.418±0.070	2.425±0.073	2.416±0.081	
MOFNet	2.355±0.110	2.233±0.111	2.231±0.096	2.206±0.107	2.196±0.100	2.197±0.107	2.224±0.111	2.212±0.116	

Table S11: All models predicting performance (MAE) for CO₂ uptake at different pressures.

Model	Pressure (kPa)							
	10	50	100	500	1000	1500	20000	50000
SVR	3.248±0.151	3.103±0.131	3.020±0.122	2.903±0.099	2.755±0.119	2.700±0.116	2.643±0.104	2.607±0.103
DT	3.775±0.206	3.728±0.192	3.637±0.221	3.527±0.146	3.371±0.171	3.286±0.151	3.239±0.230	3.167±0.195
GBRT	3.436±0.135	3.234±0.114	3.134±0.099	3.003±0.097	2.852±0.106	2.773±0.100	2.689±0.093	2.650±0.092
RF	3.141±0.139	2.960±0.125	2.861±0.103	2.767±0.100	2.626±0.093	2.560±0.092	2.486±0.089	2.443±0.090
SchNet	3.048±0.129	2.907±0.119	2.836±0.099	2.841±0.064	2.724±0.064	2.681±0.073	2.596±0.062	2.576±0.086
PaiNN	2.827±0.084	2.800±0.174	2.774±0.088	2.809±0.057	2.656±0.110	2.602±0.139	2.511±0.096	2.481±0.116
EGNN	3.081±0.124	2.942±0.112	2.908±0.082	2.857±0.067	2.738±0.096	2.699±0.066	2.620±0.069	2.590±0.067
DimeNet++	2.734±0.125	2.660±0.115	2.646±0.143	2.691±0.113	2.575±0.101	2.487±0.131	2.427±0.116	2.409±0.108
MOFNet-L	3.487±0.185	3.180±0.130	3.080±0.127	3.350±0.113	3.524±0.082	3.604±0.092	3.808±0.096	3.811±0.133
MOFNet-G	3.242±0.110	3.094±0.104	3.001±0.098	2.933±0.069	2.745±0.067	2.670±0.075	2.590±0.069	2.562±0.077
MOFNet	2.626±0.174	2.559±0.139	2.516±0.095	2.527±0.096	2.394±0.078	2.352±0.082	2.325±0.061	2.303±0.078

Table S12: All models predicting performance (MAE) for CH₄ at different pressures.

Model	Pressure (kPa)							
	50	100	150	200	500	1000	5000	10000
SVR	0.266±0.012	0.375±0.015	0.446±0.016	0.497±0.016	0.656±0.017	0.758±0.024	0.915±0.051	0.957±0.071
DT	0.346±0.012	0.493±0.017	0.575±0.017	0.630±0.024	0.835±0.023	0.993±0.022	1.202±0.053	1.271±0.059
GBRT	0.283±0.010	0.397±0.010	0.473±0.011	0.528±0.011	0.702±0.016	0.815±0.024	0.983±0.053	1.026±0.070
RF	0.258±0.010	0.358±0.010	0.427±0.010	0.474±0.011	0.620±0.012	0.730±0.017	0.898±0.045	0.951±0.067
SchNet	0.239±0.008	0.333±0.013	0.383±0.015	0.429±0.018	0.567±0.029	0.670±0.052	0.863±0.058	0.936±0.082
PaiNN	0.225±0.015	0.312±0.016	0.366±0.019	0.405±0.022	0.536±0.032	0.641±0.038	0.852±0.060	0.927±0.069
EGNN	0.245±0.010	0.336±0.011	0.394±0.015	0.432±0.017	0.570±0.032	0.677±0.050	0.870±0.050	0.932±0.059
DimeNet++	0.224±0.010	0.303±0.017	0.360±0.021	0.393±0.022	0.523±0.026	0.624±0.037	0.821±0.068	0.926±0.086
MOFNet-L	0.290±0.011	0.408±0.016	0.488±0.016	0.558±0.019	0.816±0.026	1.094±0.023	1.960±0.068	2.385±0.071
MOFNet-G	0.264±0.010	0.370±0.013	0.440±0.012	0.491±0.013	0.649±0.013	0.757±0.016	0.928±0.044	0.985±0.062
MOFNet	0.208±0.007	0.288±0.012	0.339±0.011	0.378±0.011	0.497±0.012	0.595±0.016	0.778±0.036	0.844±0.062

Table S13: All models predicting performance (PCC) for N₂ at different pressures.

Model	Pressure (kPa)							
	0.2	5	10	20	40	60	80	100
SVR	0.725±0.045	0.833±0.068	0.847±0.067	0.855±0.069	0.859±0.068	0.861±0.068	0.862±0.070	0.862±0.070
DT	0.645±0.083	0.804±0.024	0.821±0.018	0.825±0.033	0.835±0.027	0.834±0.030	0.840±0.039	0.840±0.030
GBRT	0.779±0.055	0.886±0.014	0.898±0.012	0.904±0.015	0.907±0.016	0.909±0.015	0.910±0.016	0.911±0.016
RF	0.784±0.057	0.887±0.016	0.901±0.015	0.907±0.017	0.911±0.017	0.912±0.018	0.912±0.018	0.913±0.017
SchNet	0.783±0.057	0.882±0.029	0.901±0.019	0.908±0.018	0.911±0.017	0.911±0.019	0.915±0.017	0.915±0.016
PaiNN	0.770±0.084	0.884±0.025	0.895±0.020	0.902±0.026	0.901±0.022	0.913±0.022	0.912±0.021	0.911±0.022
EGNN	0.774±0.050	0.887±0.026	0.900±0.017	0.904±0.022	0.909±0.020	0.913±0.017	0.912±0.020	0.915±0.017
DimeNet++	0.778±0.059	0.887±0.024	0.897±0.02	0.909±0.015	0.909±0.019	0.910±0.022	0.911±0.017	0.914±0.017
MOFNet-L	0.539±0.051	0.579±0.044	0.593±0.062	0.585±0.048	0.589±0.048	0.585±0.050	0.611±0.041	0.583±0.036
MOFNet-G	0.776±0.028	0.883±0.017	0.899±0.014	0.905±0.016	0.908±0.015	0.910±0.014	0.911±0.015	0.912±0.015
MOFNet	0.792±0.043	0.889±0.021	0.901±0.013	0.909±0.015	0.911±0.014	0.913±0.016	0.915±0.013	0.915±0.013

Table S14: All models predicting performance (PCC) for CO₂ at different pressures.

Model	Pressure (kPa)							
	10	50	100	500	1000	1500	20000	50000
SVR	0.490±0.030	0.464±0.038	0.449±0.034	0.583±0.035	0.709±0.036	0.755±0.033	0.799±0.033	0.816±0.031
DT	0.389±0.036	0.355±0.037	0.352±0.064	0.475±0.034	0.585±0.052	0.658±0.045	0.720±0.045	0.744±0.044
GBRT	0.523±0.035	0.502±0.040	0.495±0.066	0.613±0.041	0.727±0.029	0.776±0.032	0.827±0.028	0.843±0.026
RF	0.576±0.030	0.556±0.041	0.554±0.056	0.643±0.040	0.747±0.026	0.792±0.028	0.837±0.025	0.853±0.024
SchNet	0.581±0.064	0.578±0.087	0.576±0.071	0.651±0.044	0.741±0.024	0.783±0.029	0.836±0.023	0.851±0.024
PaiNN	0.597±0.032	0.575±0.029	0.562±0.050	0.630±0.047	0.724±0.047	0.778±0.019	0.825±0.023	0.841±0.024
EGNN	0.574±0.078	0.567±0.080	0.551±0.083	0.642±0.058	0.733±0.049	0.771±0.052	0.833±0.021	0.847±0.023
DimeNet++	0.607±0.040	0.600±0.032	0.583±0.043	0.648±0.033	0.735±0.023	0.784±0.021	0.830±0.020	0.846±0.026
MOFNet-L	0.480±0.046	0.509±0.043	0.514±0.048	0.516±0.038	0.555±0.046	0.568±0.056	0.587±0.049	0.598±0.047
MOFNet-G	0.538±0.023	0.517±0.044	0.517±0.053	0.617±0.037	0.735±0.027	0.783±0.028	0.832±0.027	0.846±0.026
MOFNet	0.644±0.045	0.632±0.039	0.627±0.037	0.681±0.027	0.771±0.021	0.805±0.024	0.840±0.025	0.854±0.021

Table S15: All models predicting performance (PCC) for CH₄ at different pressures.

Model	Pressure (kPa)							
	50	100	150	200	500	1000	5000	10000
SVR	0.606±0.023	0.655±0.024	0.681±0.023	0.698±0.022	0.758±0.018	0.811±0.016	0.893±0.035	0.894±0.073
DT	0.431±0.027	0.467±0.035	0.507±0.021	0.542±0.027	0.621±0.026	0.680±0.024	0.825±0.041	0.825±0.096
GBRT	0.596±0.024	0.643±0.023	0.665±0.022	0.684±0.020	0.745±0.015	0.798±0.013	0.892±0.033	0.900±0.072
RF	0.642±0.016	0.686±0.014	0.708±0.014	0.725±0.014	0.785±0.013	0.824±0.013	0.900±0.035	0.900±0.073
SchNet	0.703±0.023	0.743±0.018	0.770±0.012	0.781±0.014	0.822±0.015	0.849±0.030	0.910±0.032	0.900±0.081
PaiNN	0.707±0.037	0.750±0.022	0.777±0.017	0.791±0.019	0.829±0.017	0.859±0.017	0.907±0.031	0.904±0.073
EGNN	0.686±0.026	0.735±0.017	0.761±0.016	0.778±0.016	0.823±0.014	0.851±0.020	0.908±0.037	0.908±0.073
DimeNet++	0.708±0.024	0.76±0.018	0.775±0.019	0.797±0.013	0.838±0.015	0.868±0.016	0.908±0.037	0.896±0.080
MOFNet-L	0.569±0.027	0.615±0.031	0.632±0.035	0.638±0.038	0.658±0.027	0.658±0.017	0.648±0.029	0.617±0.054
MOFNet-G	0.629±0.019	0.678±0.016	0.700±0.015	0.716±0.014	0.775±0.013	0.819±0.014	0.896±0.034	0.901±0.072
MOFNet	0.736±0.014	0.772±0.017	0.794±0.014	0.805±0.015	0.845±0.011	0.869±0.015	0.915±0.035	0.912±0.078

Table S16: All models predicting performance with different pressure mechanisms on NIST-ISODB data set.

Model	N ₂		CO ₂		CH ₄	
	MAE (mol/kg)	PCC	MAE (mol/kg)	PCC	MAE (mol/kg)	PCC
MOFNet_MP ^a	4.723±3.366	0.797±0.140	2.999±1.998	0.960±0.085	1.206±1.033	0.971±0.041
MOFNet_AD ^b	4.112±2.667	0.824±0.285	2.643±1.842	0.966±0.058	1.106±1.153	0.983±0.021

^a Inputting pressure as the feature.

^b Pressure adaptive mechanism.

References

- (1) 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 and Mesoporous Materials* **2012**, *149*, 134–141.
- (2) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO₂ Adsorption Properties of Metal–Organic Frameworks. *The Journal of Physical Chemistry C* **2020**, *124*, 27996–28005.
- (3) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Q, S. R. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019. *Journal of Chemical & Engineering Data* **2019**, *64*, 5985–5998.
- (4) Chen, P.; Wang, Y.; Yan, H.; Gao, S.; Xu, Z.; Li, Y.; Mo, Q.; Huang, J.; Tao, J.; Pan, G.; Li, j.; Du, Y. 3DStructGen: an interactive web-based 3D structure generation for non-periodic molecule and crystal. *Journal of cheminformatics* **2020**, *12*, 1–11.