Supplementary Information

Accelerated screening of water-stable MOF structures using digital reticular chemistry method

Zeyang Song[a], Shaopeng Lu[a], Shanshan Cheng[a], Haijian Li[c], Hongyi Gao[a]*, Zhimeng Liu[a], Yuqiao Su[a], Wenqing Li[a], Yujing Ji[a], Xu Jin[b]*, Ge Wang [a]*

a Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Key Laboratory of Function Materials for Molecule & Structure Construction, School of Materials Science and Engineering, University of Science and Technology Beijing, Beijing 100083 (China)

b Research Institute of Petroleum Exploration&Development, PetroChina, Beijing, 100083, PR China

c National Key laboratory of Energetic Materials, Xi'an Modern Chemistry Research Institute, Xi'an 710065, China

E-mail: hygao@ustb.edu.cn (H. Gao), jinxu@petrochina.com.cn (X. Jin), gewang@ustb.edu.cn (G. Wang)

Table S1. Criteria for Water Stability Classifications.[1]

Category	Criteria
Thermodynamically stable	stable after long-term exposure to aqueous solutions: week or greater in pure water, day(s) in acidic/basic or boiling conditions strong potential for a wide range of applications
High kinetic stability	stable after exposure to high humidity conditions: decomposes after short exposure times in liquid water strong potential for industrial applications with high humidity conditions
Low kinetic stability	stable under low humidity conditions potential for applications with predried gas conditions
unstable	quickly breaks down after any moisture exposure potential for applications under moisture-free conditions

Table S2. Examples of keywords used in string matching methods for locating water stability verification paragraphs.

Keyword	relevant description in the original literature
water	as well as in boiling water for 24 h. The PXRD patterns of treated samples are well retained
aqueous	maintain its crystallinity and structural integrity after exposure to air for 3months, or immersion in water and aqueous solutions of pH1 and 12 for 3days
stable	sample was soaked in water for 5 days; the PXRD patterns demonstrated that its framework remained stable
humidity	The results reveal that the structural integrity of TIFSIX-2-Cu-i was unaffected by humidity , as was the BET surface area
H_2O	Their frameworks remain stable under H2O for 2 months and in particular
exposure	retains high crystallinity and is resistant to water upon exposure to aqueous solutions
collapse	collapse to nonporous amorphous materials after water vapor isotherm
unstable	loses crystallinity in water within minutes, loses crystallinity in air over several hours or days, desolvated sample unstable in air
vapor	Full loss of BET surface area and loss of peaks in PXRD after vapor adsorption isotherm in air at 298 K.

moisture	the CO_2uptakes were minimally affected by moisture exposure/ found to be unstable when exposed to moisture

Table S3. Data distribution of the combined dataset obtained according to the 3-class strategy.

	Stable(S)	**High-stability (HS)**	**Low-stability (LS)**
Category tags	2	1	0
Numbers	114	107	90

Figure S1. Data distribution of the combined dataset obtained according to the 3- and 2-class strategy.

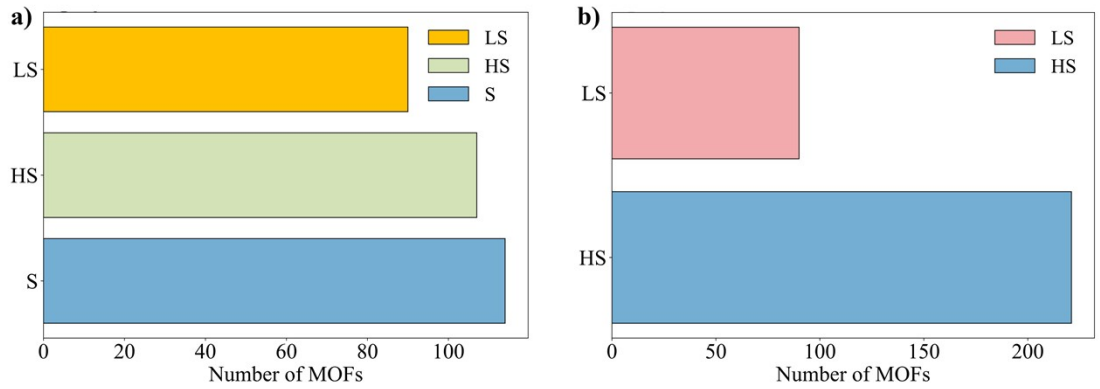


Table S4. Data distribution of the combined dataset obtained according to the 2-class strategy.

	High-stability (HS)	**Low-stability (LS)**
Category tags	1	0
Numbers	221	90

Text S1. Revised autocorrelation functions (RACs)[2] is an improved compact vector descriptor based on Autocorrelation functions (ACs), which is stable for different system sizes and compositions and does not change with the change of bond lengths and bond orders.RACs are based on different attribute values for each different connection depth of a particular molecular graph vertex, and the related formulas are shown in (1.1)~(1.4):

$$ {}^{\mathrm{f}}_{ax/\mathrm{eq}}P_d = \frac{1}{|\,ax/eq\ \text{ligands}\,|}\sum_{i}^{n_{ax/eq}}\sum_{j}^{n_{ax/eq}} P_iP_j\delta(d_{ij},d) \qquad \backslash^{*}\ MERGEFORMAT\ (1.1) $$

$$ {}^{\mathrm{mc}}_{all}P_d = \sum_{i}^{mc}\sum_{j}^{all} P_iP_j\delta(d_{ij},d) \qquad \backslash^{*}\ MERGEFORMAT\ (1.2) $$

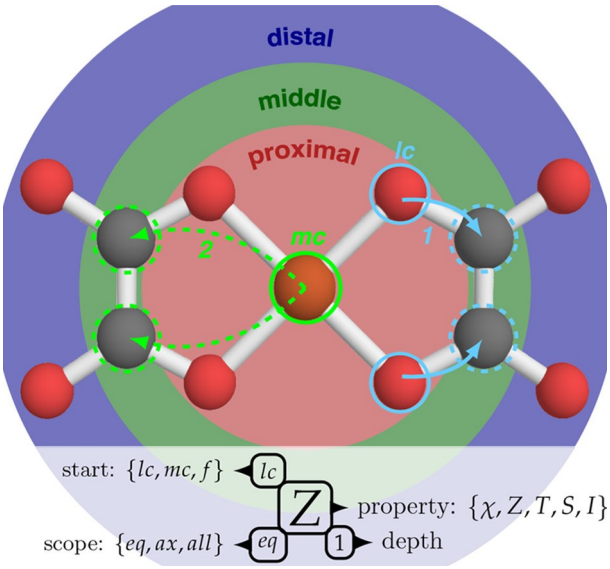
$$ {}^{\mathrm{lc}}_{ax/\mathrm{eq}}P_d = \frac{1}{|\,ax/eq\ \text{ligands}\,|}\frac{1}{|\,lc\,|}\sum_{i}^{lc}\sum_{j}^{n_{ax/eq}} P_iP_j\delta(d_{ij},d) \backslash^{*} $$

$$ MERGEFORMAT\ (1.3) $$

$$ {}^{\mathrm{lc}/mc}_{ax/eq/all}P_d{}' = \sum_{i}^{lc/mc}\sum_{j}^{scope} (P_i - P_j)\delta(d_{ij},d) \qquad \backslash^{*}\ MERGEFORMAT\ (1.4) $$

An example of a RAC descriptor: D-mc-Z-3-all is calculated based on the atomic nuclear charge number (-Z) property, using metal atoms as the starting atoms (-mc), with a maximum bonding depth of 3 (-3), covering all atoms within the entire unit cell (-all), and formulated as a "difference" (D-) type descriptor.

Schematic of RACs[2]



Five heuristic atomic properties are employed in this research：(i) nuclear charge, Z (ii) Pauling electronegativity, χ (iii) topology(coordination number), T (iv) identity, I (v) covalent atomic radius, S. We apply these properties to the product and difference forms, respectively, and take the maximum depth of 3 to compute all RAC descriptors. In the case of linker connecting atom centered and functional group centered RACs, an additional atom-wise property polarizability (α) was considered.[3] A total of 176 RACs were generated, with the specific types and numbers shown in Table S5.

Table S5. The different kinds of RACs descriptors used in this study and their number.

type of starting points	type and number of descriptors
metal centered	20 products, 20 differences
functional group centered	24 products, 24 differences
full	20 products
full linker	20 products
linker connecting atom centered	24 products, 24 differences

Text S2. Smoothed Overlap of Atomic Positions (SOAP) is a descriptor that encodes geometrical regions of atoms by using local expansions of Gaussian-coated atomic densities and orthogonal functions based on spherical harmonic functions and radial basis functions. SOAP allows for quantification of similarity between atoms without being constrained by discretisation or specific coordination coefficients. Although the features of SOAP descriptors in each dimension do not have actual physical significance, their combined features can provide rich local structural information for machine learning models, thus improving the predictive performance of the models. The relevant formulas for the calculation of SOAP are shown in (1.5)~(1.8):

$$p_{nn'l}^{Z_1Z_2} = \pi\sqrt{\frac{8}{2l+1}}\sum_m c_{nlm}^{Z_1} * c_{nl'm}^{Z_2} \qquad \backslash* \ MERGEFORMAT\ (1.5)$$

$$c_{nlm}^{Z} = \iiint_{R^3} dV g_n(r) Y_{lm}(\theta,\phi)\rho^Z(r) \qquad \backslash* \ MERGEFORMAT\ (1.6)$$

$$\rho^Z(r) = \sum_i^{|Z_i|} e^{-1/2\sigma^2|r-R_i|^2} \qquad \backslash* \ MERGEFORMAT\ (1.7)$$

$$K^{SOAP}(p,p') = \left(\frac{p\cdot p'}{\sqrt{p\cdot pp'\cdot p'}}\right)^{\xi} \qquad \backslash* \ MERGEFORMAT\ (1.8)$$

The main parameters used in this study for generating SOAP descriptors by using DScribe 2.1.1 are shown in Table S6.

Table S6. The main parameters employed when generating SOAP descriptors by using DScribe 2.1.1.[4]

r_cut(Å)	σ	n_max	l_max	method	crossover
5.0	0.2	6	4	inner	False

Text S3. The Henry constant, which is used to describe the adsorption behaviour of a gas at low pressure on the surface of a material such as a MOF, is an important concept in the field of gas adsorption. It is derived from Henry's Law, which states that the solubility of a gas in a solid or liquid at low pressure is proportional to the partial pressure of the gas:

$$C = K_H \cdot P \qquad \text{* MERGEFORMAT (1.9)}$$

The Henry constant characterises the adsorption behaviour of a material at lower gas partial pressures and is a quantitative indicator of the initial adsorption capacity. The magnitude of this value often implies the strength of water adsorption by the material. The main parameters used in this study for the calculation of Henry constant for the adsorption of water molecules by MOF using the Sorption module of Material Studio version 2019 are shown in Table S7.

Table S7. Main parameters employed for Henry constant calculations using the sorption module of Material Studio 2019.

Task	Henry constant	**Forcefield**	Universal
Method	Metropolis	**Charges**	Use current
Production steps	100000	**Electrostatic**	Ewald&Group
Temperature	298(constant)	**Van der Waals**	Atom based

Table S8. The relevant parameters of the t-SNE algorithm and the UMAP algorithm used in this study. T-Distributed Stochastic Neighbor Embedding (t-SNE) is a probabilistic-based dimensionality reduction algorithm that is typically effective in capturing the local structure of the data and is therefore well-suited for the detection of details in local neighborhoods. The algorithm maps high-dimensional data into two- or three-dimensional space, with the objective of maintaining the similarity between the original data points in a low-dimensional space. It is typically effective in distinguishing complex data with nonlinear structure. Uniform Manifold Approximation and Projection (UMAP) is a topology-based dimensionality reduction algorithm that better preserves the global structure of the data while maintaining the relationships between local neighbouring points.

t-SNE	'angle': 0.5, 'early_exaggeration': 12.0, 'init': 'pca', 'learning_rate': 'auto', 'method': 'barnes_hut', 'metric': 'euclidean', 'metric_params': None, 'min_grad_norm': 1e-07, 'n_components': 2, 'n_iter': 1000, 'n_iter_without_progress': 300, 'n_jobs': None, 'perplexity': 50, 'random_state': 123, 'verbose': 0
UMAP	'a': None, 'angular_rp_forest': False, 'b': None, 'dens_frac': 0.3, 'dens_lambda': 2.0, 'dens_var_shift': 0.1, 'densmap': False, 'disconnection_distance': None, 'force_approximation_algorithm': False, 'init': 'spectral', 'learning_rate': 1.0,

'local_connectivity': 1.0, 'low_memory': True, 'metric': 'euclidean', 'metric_kwds': None, 'min_dist': 0.1, 'n_components': 2, 'n_epochs': None, 'n_jobs': -1, 'n_neighbors': 30, 'negative_sample_rate': 5, 'output_dens': False, 'output_metric': 'euclidean', 'output_metric_kwds': None, 'precomputed_knn': (None, None, None), 'random_state': None, 'repulsion_strength': 1.0, 'set_op_mix_ratio': 1.0, 'spread': 1.0, 'target_metric': 'categorical', 'target_metric_kwds': None, 'target_n_neighbors': -1, 'target_weight': 0.5, 'tqdm_kwds': {'desc': 'Epochs completed', 'bar_format': '{desc}: {percentage:3.0f}%| {bar} {n_fmt}/{total_fmt} [{elapsed}]', 'disable': True}, 'transform_mode': 'embedding', 'transform_queue_size': 4.0, 'transform_seed': 42, 'unique': False, 'verbose': False

Figure S2. T-SNE plot and UMAP plot for dimensionality reduction of the original dataset under 2-class strategy. **b)** t-SNE plot and UMAP plot for dimensionality reduction of the original dataset under 3-class strategy.

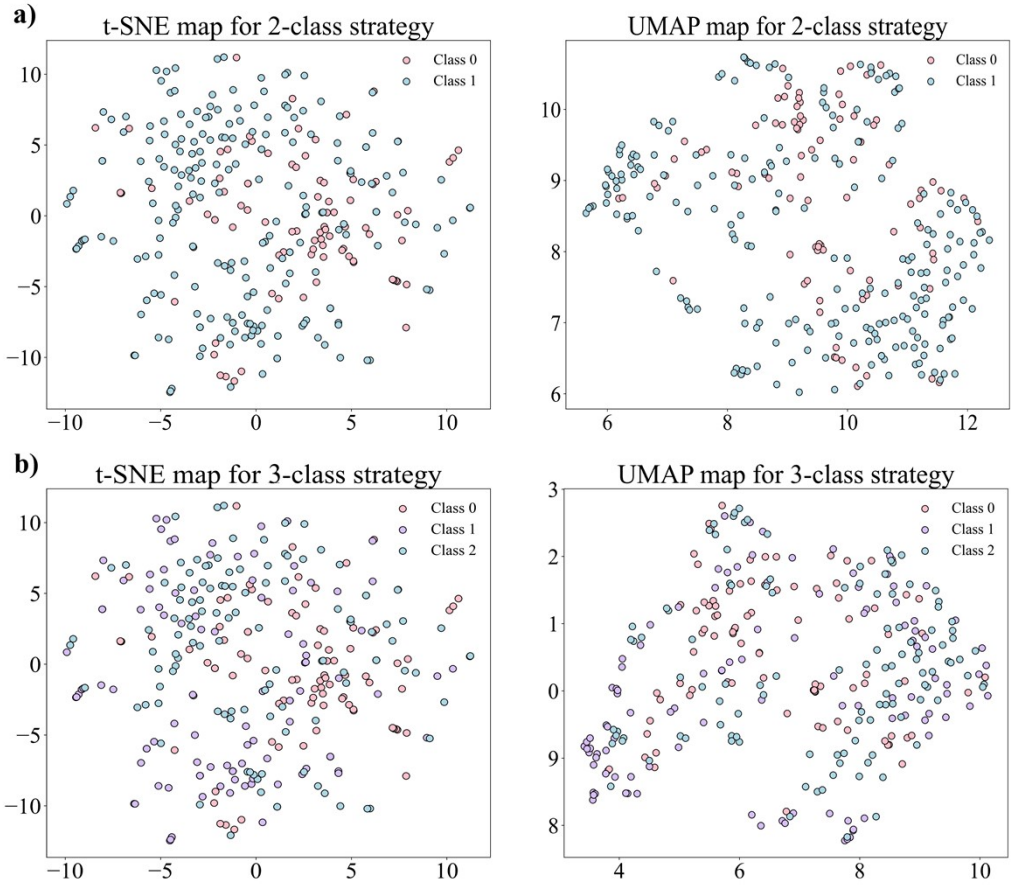


Table S9. 2-class models trained using the PyCaret workflow on the dataset after dimensionality reduction using the RFECV algorithm and their performance (performance is sorted from high to low, only the top five are shown).

	Model	Accuracy	AUC	Recall	Prec.	F1
et	Extra Trees Classifier	0.8571	0.8687	0.9479	0.8661	0.9038
xgboost	Extreme Gradient Boosting	0.8522	0.8524	0.9471	0.8595	0.9006
catboost	CatBoost Classifier	0.8431	0.8781	0.9413	0.8539	0.8944
rf	Random Forest Classifier	0.8340	0.8555	0.9283	0.8522	0.8876
lightgbm	Light Gradient Boosting Machine	0.8108	0.8455	0.8958	0.8469	0.8687

Table S10. 3-class models trained using the PyCaret workflow on the dataset after dimensionality reduction using the RFECV algorithm and their performance (performance is sorted from high to low, only the top five are shown).

	Model	Accuracy	AUC	Recall	Prec.	F1
rf	Random Forest Classifier	0.6686	0.8380	0.6686	0.6961	0.6628
gbc	Gradient Boosting Classifier	0.6502	0.8223	0.6502	0.6657	0.6471
lightgbm	Light Gradient Boosting Machine	0.6498	0.8266	0.6498	0.6772	0.6440

et	Extra Trees Classifier	0.6455	0.8312	0.6455	0.6874	0.6412
catboost	CatBoost Classifier	0.6455	0.8394	0.6455	0.6815	0.6381

Table S11. 2-class models trained using the PyCaret workflow on the dataset after dimensionality reduction using the PCA algorithm and their performance (performance is sorted from high to low, only the top five are shown).

	Model	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**
catboost	CatBoost Classifier	0.8247	0.8312	0.9871	0.8091	0.8888
xgboost	Extreme Gradient Boosting	0.8197	0.8447	0.9275	0.8409	0.8798
rf	Random Forest Classifier	0.8065	0.7896	0.9871	0.7936	0.8788
et	Extra Trees Classifier	0.8065	0.8154	0.9812	0.7970	0.8784
lightgbm	Light Gradient Boosting Machine	0.8050	0.8440	0.9275	0.8216	0.8694

Table S12. 3-class models trained using the PyCaret workflow on the dataset after dimensionality reduction using the PCA algorithm and their performance (performance is sorted from high to low, only the top five are shown).

	Model	**Accuracy**	**AUC**	**Recall**	**Prec.**	**F1**
catboost	CatBoost Classifier	0.5942	0.7424	0.5942	0.6151	0.5772

xgboost	Extreme Gradient Boosting	0.5935	0.7259	0.5935	0.5964	0.5818
et	Extra Trees Classifier	0.5617	0.7400	0.5617	0.5796	0.5541
lightgbm	Light Gradient Boosting Machine	0.5385	0.7086	0.5385	0.5599	0.5280
rf	Random Forest Classifier	0.5342	0.7170	0.5342	0.5702	0.5239

Text S4. Specific formula for the metrics used to assess model performance and their meanings.

Accuracy: The proportion of correctly classified samples among the total samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \backslash * MERGEFORMAT\ (1.10)$$

AUC (Area Under the Curve): Represents the area under the ROC curve, showing the model's ability to distinguish between classes. AUC does not have a simple formula as it is computed by integrating the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels.

Recall (Sensitivity or True Positive Rate): The ability of the model to identify all positive samples.

$$Recall = \frac{TP}{TP + FN} \quad \backslash * MERGEFORMAT\ (1.11)$$

Precision: The proportion of true positive predictions among all positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \quad \backslash * MERGEFORMAT\ (1.12)$$

F1 Score: The harmonic mean of Precision and Recall, providing a balance between the two, especially when there is an uneven class distribution.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad \backslash * MERGEFORMAT\ (1.13)$$

True Positive (TP): The number of instances where the model correctly predicts the positive class. This means that a sample that is actually positive is correctly classified as positive.
True Negative (TN): The number of instances where the model correctly predicts the negative class. This indicates that a sample that is actually negative is correctly classified as negative.
False Positive (FP): The number of instances where the model incorrectly predicts the positive class. In this case, a sample that is actually negative is mistakenly classified as positive.
False Negative (FN): The number of instances where the model incorrectly predicts the negative class. Here, a sample that is actually positive is wrongly classified as negative.

Figure S3. ROC curves of a) 2-class ETC model and b) 3-class RF model based on the reduced dataset using the RFECV algorithm.

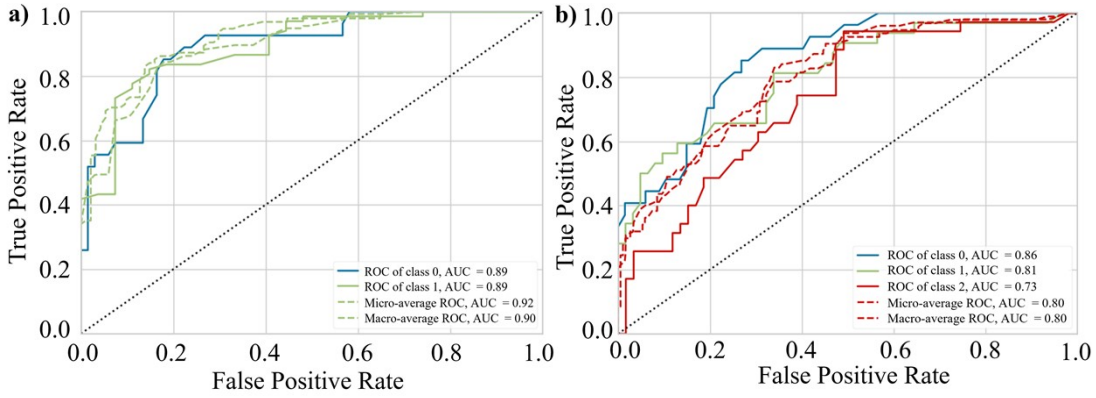


Table S13. The parameter grid used for hyperparametric tuning.

Parameter	Value
n_estimators	100,300,500,700,800,900, 1200,1300
max_features	'log2', 'sqrt', 0.5,0.6, 0.7,0.8
max_depth	None, 5,10, 15,20
min_samples_split	2, 5, 10, 15, 20
min_samples_leaf	1, 2, 4, 6, 8, 10
bootstrap	True, False

Figure S4. a) ROC curves of the optimal hyperparameter-based 2-class ETC model on the test set, the ROC curves of all categories show that the model exhibits good performance at different thresholds. **b)** Confusion matrix of the optimal hyperparameter-based 2-class ETC model on

the test set, where 88% of the data points are correctly classified.

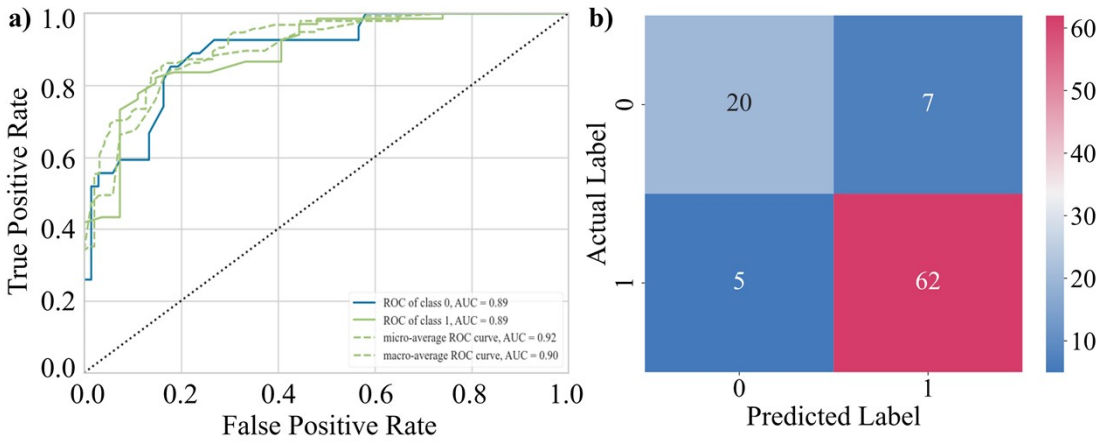


Table S14. Optimal hyperparameter combinations for the 2-class ETC model after hyperparameter tuning.

hyperparameter	Value	hyperparameter	Value
bootstrap	FALSE	**min_samples_leaf**	1
ccp_alpha	0	**min_samples_split**	2
class_weight	None	**min_weight_fraction_leaf**	0
criterion	gini	**n_estimators**	100
max_depth	None	**n_jobs**	-1
max_features	sqrt	**oob_score**	False
max_leaf_nodes	None	**random_state**	123
max_samples	None	**verbose**	0
min_impurity_decrease	0	**warm_start**	False

Table S15. Optimal hyperparameter combinations for the 3-class ETC model after hyperparameter tuning.

hyperparameter	Value	hyperparameter	Value
bootstrap	FALSE	**min_samples_leaf**	1
ccp_alpha	0	**min_samples_split**	2
class_weight	None	**min_weight_fraction_leaf**	0
criterion	gini	**n_estimators**	1200
max_depth	10	**n_jobs**	None
max_features	sqrt	**oob_score**	False
max_leaf_nodes	None	**random_state**	0
max_samples	None	**verbose**	0
min_impurity_decrease	0	**warm_start**	False

Figure S5. Partial Dependency Plot (PDP) of the 3 most important features selected by the feature_importances_ parameter of the 3-class ETC model.

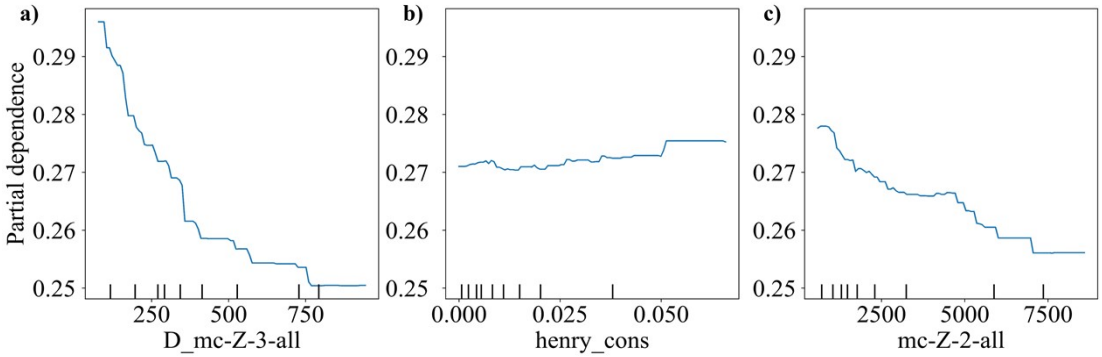


Figure S6. Kernel density estimation plots of MOF water stability for a) **D_mc-Z-3-all** and b) **mc-Z-2-all**.

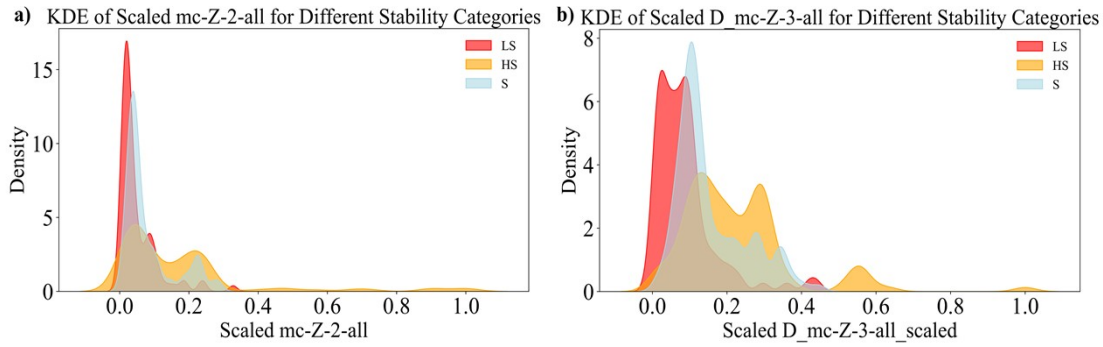


Table S16. Comparison of predicted labels with true labels. The predicted labels are the result of the prediction using the 3-class ETC model on the validation set.

filename	prediction	true
BIQHIU	0	1
CAVRAU	0	0
EWESEF	1	2
FUYCAF	1	1
GUCTAB	1	2
HUFKOK	2	2
IXIKOR	1	2
KUMGIK	0	0
KUMGOQ	0	0
LAFFEG	2	2
ONOCOK	0	0
RUYVIS	1	2
SOYWOT	0	2
WUHCUZ	1	1
YEZKIZ	2	2
ADABUE	2	2
AVIPAX	0	0

AVIPEB	0	0
BAFCAP	0	1
BICPOT	0	0
CIMTEZ	2	2
EDOVEA	1	1
ELEREU	2	2
ELEROE	0	1
EZIMIK	0	0
EZOGAC	0	0
FUJCOC01	0	0
GUVCIJ	0	0
GUVCOP	0	0
HOHMIB	1	1
HOWROC	0	0
IQECIS	2	2
JOFZOV	1	1
JUMZOI	2	2
JUMZUO01	2	2
KUFTAI	2	2
LATCAN	2	2
LULWUL	0	0
NABMUA01	0	0
OLUHAG	0	1
OSAVEK	0	0
OSAVIO	0	0
OSAVOU	0	0
OSAVUA	0	0
OSAWEL	0	0
OTENEH	1	1
PAFXUS	2	2
PORLAL	2	2

PUJJEL	2	2
PURNAT	1	1
PURNEX	1	1
PURNIB	1	1
PURNOH	1	1
QIFLOI	2	2
QIFLUO	2	2
RORJOZ	2	2
RUJGIN	0	0
RUVMIF	1	1
RUYQIN	2	2
RUYVEO	2	2
SENWAL	0	0
SENWEP	0	0
SENWIT	0	0
SENWOZ	0	0
SUTJEY	2	2
SUTJOI	2	2
TIDSEE	0	0
TIFBEP	0	0
TIFCAM	0	0
TILBOH	1	1
UGOCAW	1	2
VAGMAT	0	1
WAJHUM	0	0
WEHPIJ	0	0
WUPTUY	1	2
WUPVAG	1	1
WUTBAQ	1	2
WUTBEU	2	2
YOZROW	0	0

YUQMOO	2	2

Figure S7. Learning curves for training set fraction of 2- and 3-class models. (Random Forest Classifier based)

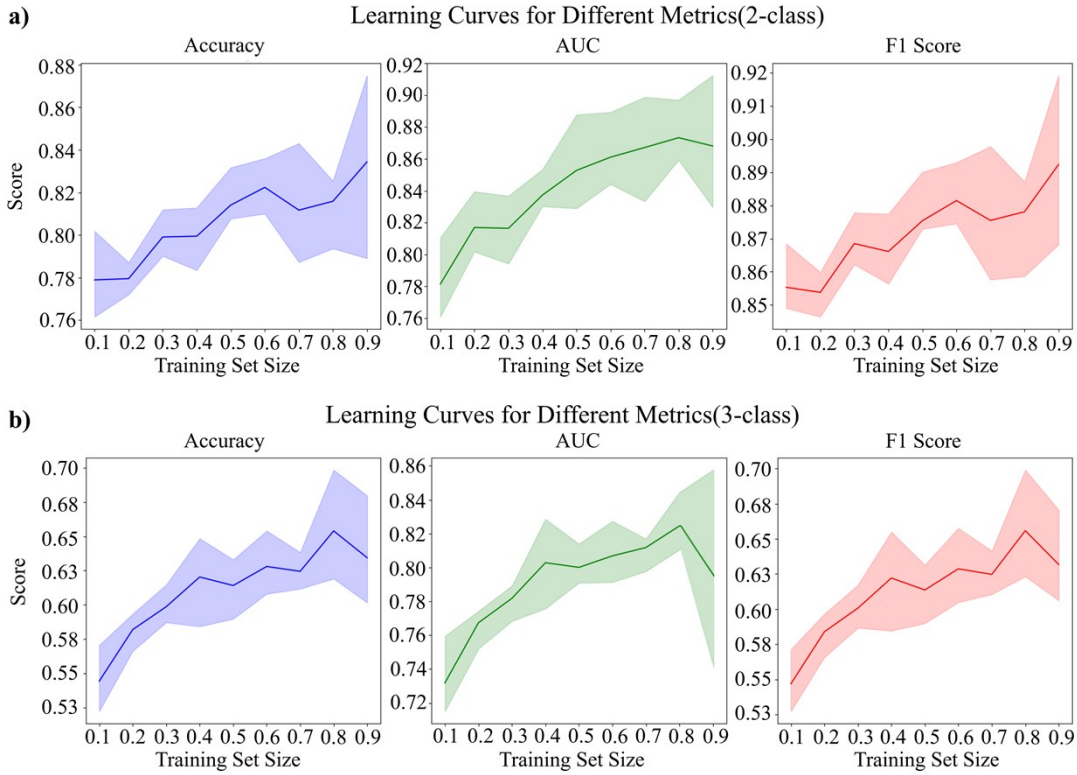


Figure S8. Boxplots for training set fraction of 2- and 3-class models. (Random Forest Classifier based)

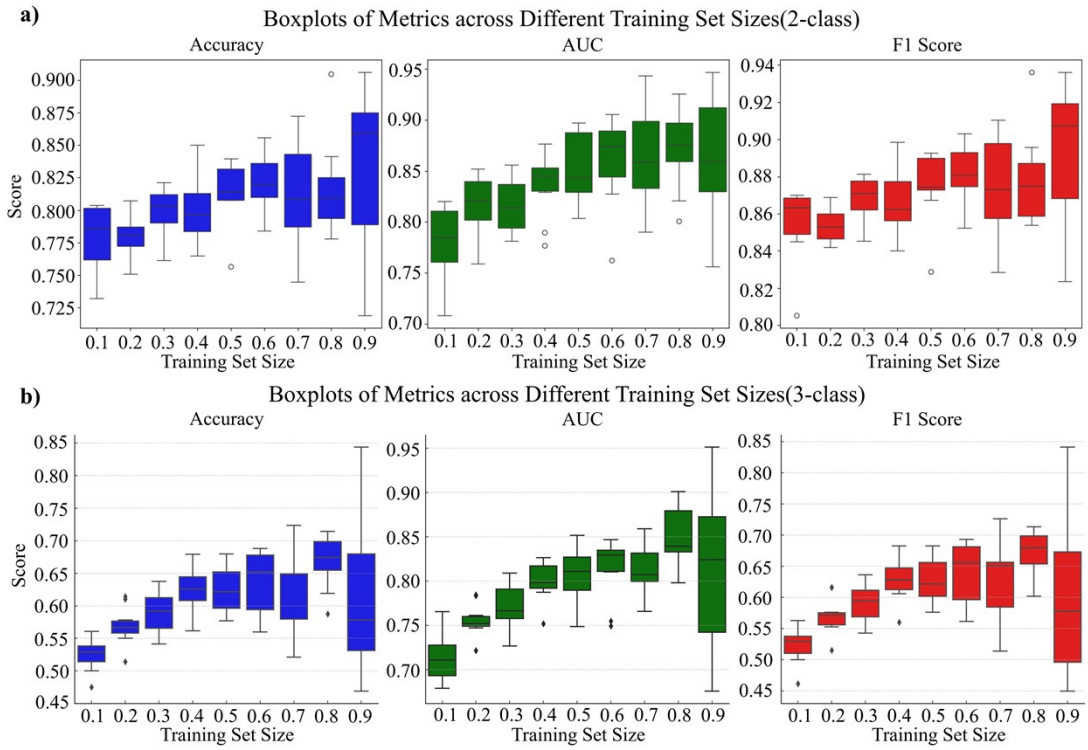


Figure S9. The treemap of the metal composition of the MOFs predicted to be water stable, where different colors represent different Lewis acid classes and the size of the rectangle represents the proportion of the corresponding category in all stable MOFs.

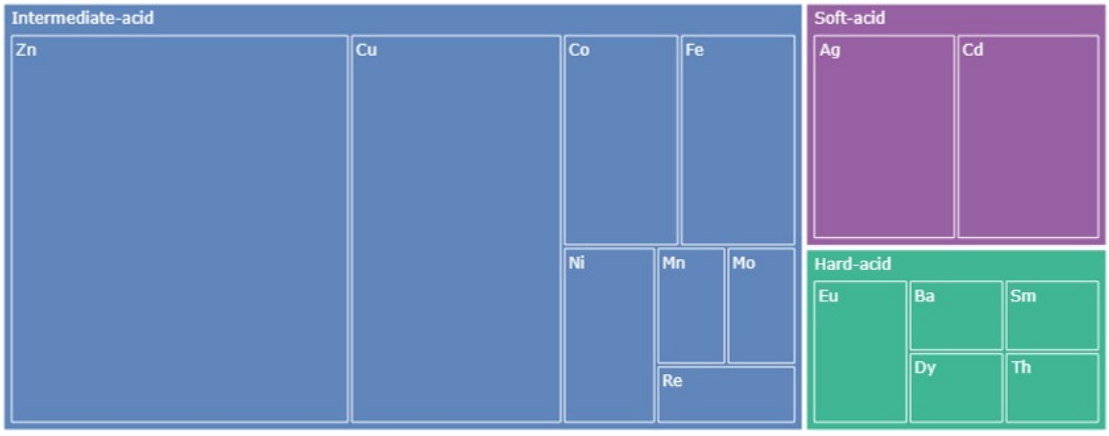


Figure S10. Photographs of the samples at different stages: **a)** as synthesized **b)** during soaking and **c)** after drying.

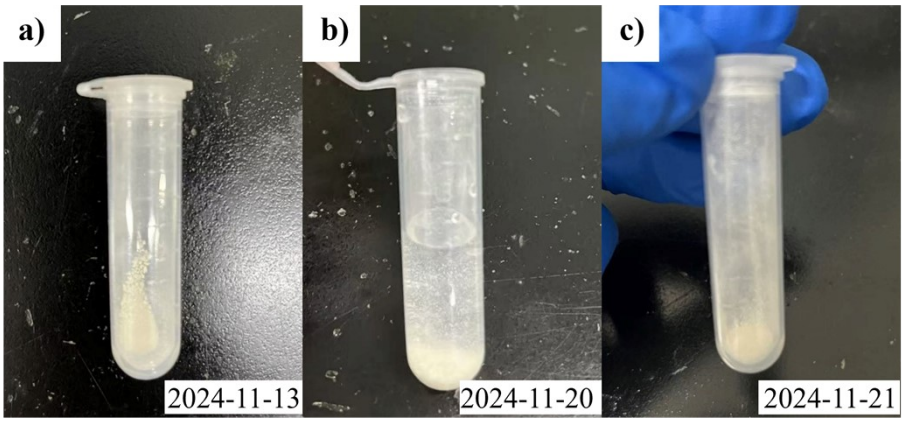
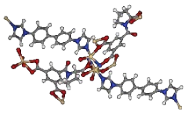
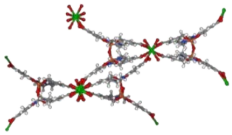
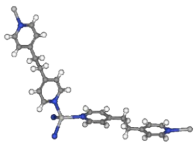
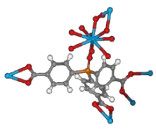
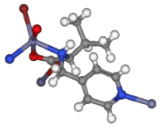
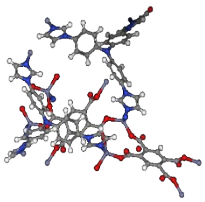
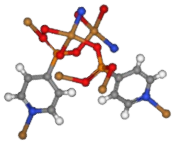
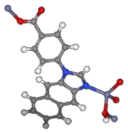
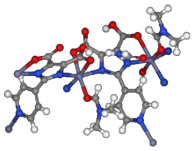
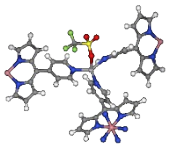
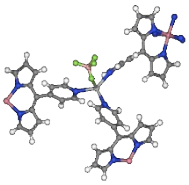
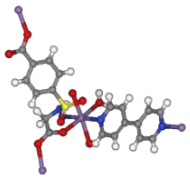
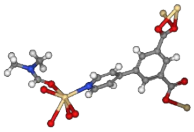
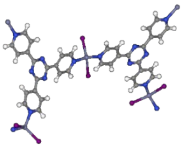
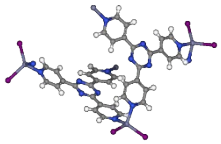
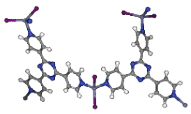
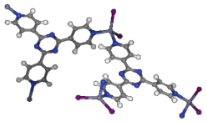
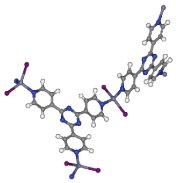
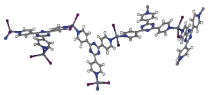
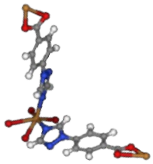
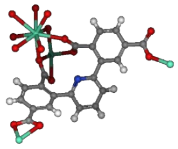
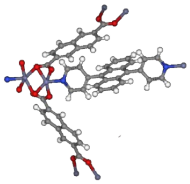
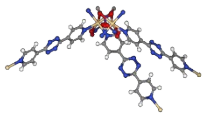
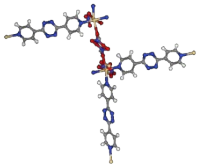
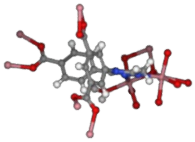
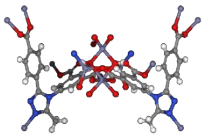
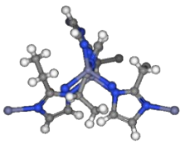
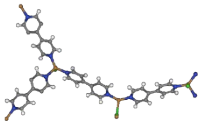


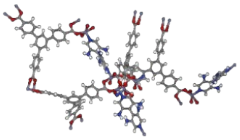
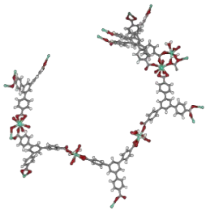
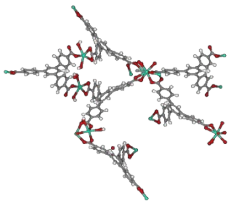
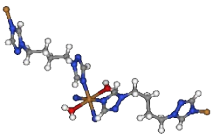
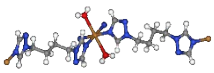
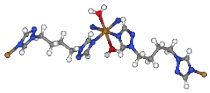
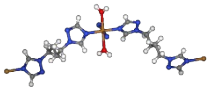
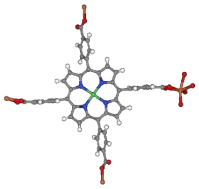
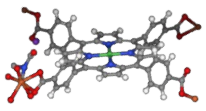
Table S17. Structural composition of MOFs predicted to be in the stable class and their possible directions of application or structural properties.

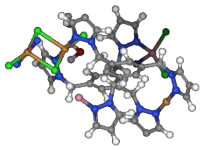
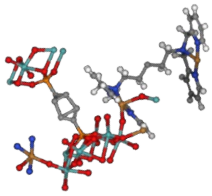
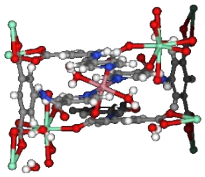
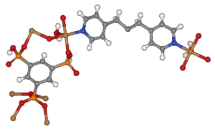
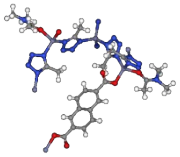
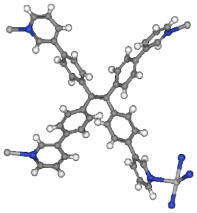
MOF	Unit cell	Metal	Application/Properties	Source
HUWHOY		Cd	homochiral metal–organic frameworks (HMOFs)	10.1021/acs.cgd.5b01359
HUVNAP		Ba/Fe	visible-light irradiation	10.1021/acs.cgd.5b00925
HUWDEJ		Ag/Re	harvesting solar energy	10.1021/ic901749r
HUWLUI		Th	centrosymmetric	10.1039/C5DT03363B
IBICON		Zn	proton conductors	10.1021/ja2078637

IBUYAH		Zn	photoluminescent	10.1039/C1CE05889D
ICIZAV		Cu	Catalysis ,molecular recognition ,ion exchanging ,nonlinear optics	10.1021/ic010609g
IFENOY		Zn	anisotropic thermal expansion	10.1021/ja401671p
IQEKOG		Zn	Luminescence	10.1016/j.inorganichem.2016.06.032
KESBOA		Ag	strong optical absorptions potential for the synthesis of chiral MOFs	10.1021/ja0645483
KESBUG		Ag	strong optical absorptions potential for the synthesis of chiral MOFs	10.1021/ja0645483
KESGAS		Mn	magnetic properties	10.1080/15533174.2012.654878
KETHEY		Cd	photoluminescence	10.1080/00958972.2012.742890
LUDTUZ		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-

				ANIE3392%3E3.0.CO;2-V
LUDVAH		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-ANIE3392%3E3.0.CO;2-V
LUDVEL		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-ANIE3392%3E3.0.CO;2-V
LUDVIP		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-ANIE3392%3E3.0.CO;2-V
LUDVOV		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-ANIE3392%3E3.0.CO;2-V
LUDVUB		Zn	has elastic properties while remaining crystalline	10.1002/1521-3773(20020916)41:18%3C3392::AID-ANIE3392%3E3.0.CO;2-V
LUFBUL		Cu	catalysis	10.4236/ojic.2012.23009

MAPLEX		Eu	multifunctional chemosensor/luminescent probe for diethyl ether vapor	10.1039/C7TC00508C
MAQMEZ		Zn	luminescence	10.1016/j.molstruc.2017.04.057
MARKOF		Cd	catalysis	10.1021/ja991698n
MARLAS		Cd	catalysis	10.1021/ja991698n
MARNUR		Co	magnetic properties	10.1016/j.jssc.2017.09.023
MARNOL		Zn	gas storage and separation	10.1016/j.jssc.2017.09.023
MECWIB		Zn	high thermal stability	10.1002/anie.200503778
MECWOH		Zn	high thermal stability	10.1002/anie.200503778
NETYIV		Cu	catalysis	10.1002/ejic.200600558

NEYVEU		Zn	selective sensing	10.1039/C2DT30689A
OMUMEQ		Eu	ion sensor	10.1021/acs.inorgchem.6b00190
OMUMIU		Dy	Magnetic properties	10.1021/acs.inorgchem.6b00190
ONAWAC		Cu	catalysis	10.1016/j.ica.2011.01.052
ONAWEG		Cu	catalysis	10.1016/j.ica.2011.01.052
ONAWEG01		Cu	catalysis	10.1016/j.ica.2011.01.052
ONAWEG02		Cu	high thermal stability	10.1080/00958972.2012.726713
PAQTIM		Ni/Fe	catalytic	10.1021/cm2025747
PAQTOS		Ni/Fe	catalytic	10.1021/cm2025747

PEBNEP		Cu/Co	catalytic	10.1039/DT9930000687
PECBOQ		Cu/Mo	catalytic	10.1016/j.ica.2012.01.011
PECMOB		Sm/Co	catalytic	10.1524/ncrs.2012.0096
PEDMOB		Cu	catalytic	10.1021/ic0509377
VOXQOQ		Zn	luminescence	10.1016/j.molstruc.2014.12.077
VOYLUS		Ag	luminescence	10.1016/j.poly.2014.11.024

The relevant datasets and codes used in this study have been uploaded to https://github.com/coco550/Water-stability/tree/main.

References

1. Burtch, N.C., Jasuja, H., and Walton, K.S. (2014). Water Stability and Adsorption in Metal–Organic Frameworks. Chem. Rev. *114*, 10575–10612. https://doi.org/10.1021/cr5002589.

2. Janet, J.P., and Kulik, H.J. (2017). Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. ACS Publications. https://doi.org/10.1021/acs.jpca.7b08750.

3. Moosavi, S.M., Nandy, A., Jablonka, K.M., Ongari, D., Janet, J.P., Boyd, P.G., Lee, Y., Smit, B., and Kulik, H.J. (2020). Understanding the diversity of the metal-organic framework ecosystem. Nat Commun *11*, 4068. https://doi.org/10.1038/s41467-020-17755-8.

4. Laakso, J., Himanen, L., Homm, H., Morooka, E.V., Jäger, M.O.J., Todorović, M., and Rinke, P. (2023). Updates to the DScribe library: New descriptors and derivatives. The Journal of Chemical Physics *158*, 234802. https://doi.org/10.1063/5.0151031.