ARTICLES
https://doi.org/10.1038/s42256-020-00249-z
nature machine intelligence

Prediction of water stability of metal–organic frameworks using machine learning

Rohit Batra[1], Carmen Chen[2], Tania G. Evans[2], Krista S. Walton[2] and Rampi Ramprasad[1]✉

Owing to their highly tunable structures, metal–organic frameworks (MOFs) are considered suitable candidates for a range of applications, including adsorption, separation, sensing and catalysis. However, MOFs must be stable in water vapour to be considered industrially viable. It is currently challenging to predict water stability in MOFs; experiments involve time-intensive MOF synthesis, while modelling techniques do not reliably capture the water stability behaviour. Here, we build a machine learning-based model to accurately and instantly classify MOFs as stable or unstable depending on the target application, or the amount of water exposed. The model is trained using an empirically measured dataset of water stabilities for over 200 MOFs, and uses a comprehensive set of chemical features capturing information about their constituent metal node, organic ligand and metal–ligand molar ratios. In addition to screening stable MOF candidates for future experiments, the trained models were used to extract a number of simple water stability trends in MOFs. This approach is general and can also be used to screen MOFs for other design criteria.

Metal–organic frameworks (MOFs) are a class of porous and crystalline materials that are increasingly being studied for gas separation[1,2], storage[3,4] and catalysis[5,6] applications. They consist of inorganic metal ions or clusters connected to organic ligands through coordination bonds, overall forming a highly porous three-dimensional (3D) crystalline structure[7]. They are known for their easily tunable components—modifications can be made to the metals, organic linkers, associated functional groups or the metal–ligand bond to customize their intrinsic properties for various applications[8–10]. From a theoretical standpoint, however, the infinite combinations possible between metal ions and organic ligands make it difficult to efficiently screen for MOFs with desired properties.

To be industrially applicable, a key property of a MOF is its water stability, because many industrial processes, such as gas separation and storage, involve some amount of water. Unfortunately, the majority of MOFs (for example, MOF-5[11] and MOF-508[12]) are unstable in water vapour, which poses a challenge for future commercialization efforts[13,14]. Based on past empirical and theoretical efforts, some general chemical trends elucidating the water stability of MOFs have been established[13,15–17]. Although MOFs with strong coordination bonds between the metal nodes and organic ligands are thermodynamically stable, the presence of significant steric hindrance or hydrophobic functional groups impart MOFs with high kinetic stability. Over the years, these general rules have been applied in the synthesis of several water-stable MOFs, including lanthanide-based $[La(pyzdc)_{1.5}(H_2O)_2].2H_2O$ (ref. [18]) and $([Dy(Cmdcp)(H_2O)_3](NO_3).2H_2O)_n$ (ref. [19]), Zr-based PCN-228/-229/-230 (ref. [20]), the metal azolate framework (MAF) series[21] and super hydrophobic fluorinated MOFs[22,23]. Although useful, these rules require a priori knowledge of the MOF atomic arrangement, which cannot be used to efficiently screen stable MOF candidates.

Thus, in this work, we have developed an efficient and instant machine learning (ML)-based strategy for screening water-stable MOFs, as portrayed by the schematic in Fig. 1. A dataset[13] of experimentally determined water stabilities for over 200 MOFs was used to construct a ML model capable of classifying a given MOF as stable or unstable. Starting from their activated formula unit, each MOF was fingerprinted, or uniquely represented by a vector of chemical features capturing information on the metal node, the organic linker and the molar ratios of the metal ions to the ligand and the associated H_2O, OH and O sites. With these MOF fingerprints as inputs, two types of classification models were constructed to represent situations with different levels of water exposure. The first, a two-class model, distinguishes between unstable and stable MOFs, while the second, a three-class model, classifies MOFs as unstable, kinetically stable or thermodynamically stable. Three different ML methods were tested, and those with the best performance, evaluated using learning curves and confusion matrix analysis on unseen cases, were selected for future predictions. Good performances of the trained ML models for MOFs that have been synthesized more recently, that is, after the publication of the dataset used to train the models, strongly suggest the applicability and merit of the surrogate models developed in this work. The penultimate step in the workflow presented in Fig. 1 was to screen new MOFs with unknown water stability behaviour. Thus, a ranked-ordered list of candidate MOFs predicted to be stable (with otherwise unknown water stability behaviour) is provided for future experiments. Further, simple chemical guidelines supporting water stability in MOFs were derived using the developed ML models and available experimental data. For example, the presence of metal ions of large atomic radius and lower ionization potential, or ligands with a low count of six-member rings and high number of cyclic divalent nodes, correlate with enhanced water stability. Besides being able to efficiently screen MOFs with a desired water stability, these ML models can be iteratively improved as more empirical measurements on MOF water stabilities become available.

Data preparation and ML methods

MOF water stability datasets. Figure 2 summarizes the MOF water stability dataset used in this work. The dataset, which includes 207 MOFs, was obtained from Burtch et al.[13]. Each MOF is categorized

[1]School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [2]School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: rampi.ramprasad@mse.gatech.edu

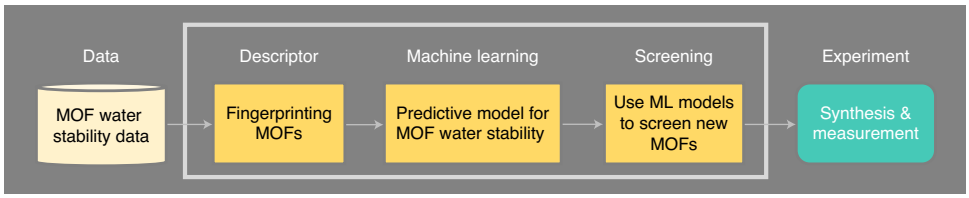


Fig. 1 | Workflow adopted to build ML models of water stability in MOFs. Of the steps shown, the middle three steps in the white box highlight the focus of this work.

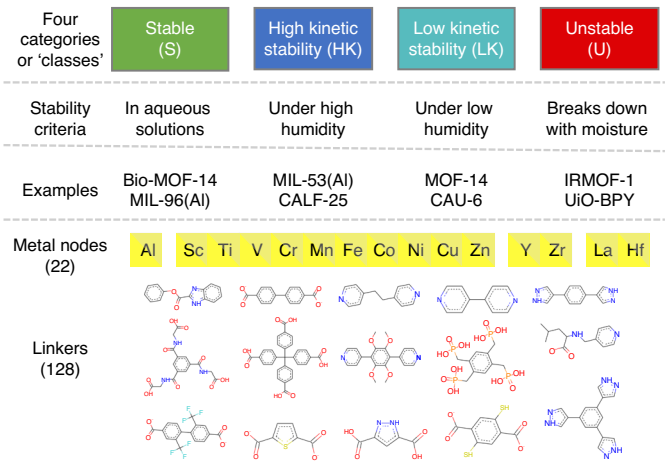


Fig. 2 | MOF water stability training data. Details are provided for four different categories of MOFs, along with the associated humidity condition for stability and a pair of exemplary cases. A few representative metal nodes and linkers constituting this dataset are also included.

into one of four classes of stability: stable (S), high kinetic stability (HK), low kinetic stability (LK) or unstable (U). The classification was based on the type of water exposure (aqueous, humid or dry) tested, the length of each exposure (weeks, days or hours), the characterization techniques (powder X-ray diffraction and Brunauer–Emmett–Teller surface area measurements) employed to ascertain any material degradation, and some other additional exposure conditions (acidic/basic environment or boiling temperatures). From an applications standpoint, the most important design criterion is the amount of water to which a MOF can be exposed without degrading. In terms of chemical diversity, the dataset consists of 22 different metal nodes and 128 different ligands (a few example cases are shown in Fig. 2). Additional statistics on metals favouring high water stability are included in Extended Data Fig. 1. Hereafter, we will refer to this dataset as the Burtch dataset.

It is important to note that the number of MOFs within each stability class is slightly imbalanced: there are 25, 118, 42 and 22 MOFs for S, HK, LK and U, respectively. However, in contrast to the Burtch dataset, we expect a large percentage of MOFs to fall under the LK or U categories[24,25]. To diminish the impact of class imbalance and the implicit bias towards 'positive examples' or MOFs with high water stability, the aforementioned four classes of MOFs were combined strategically. Two classes were obtained by combining S with HK (denoted class 1) and U with LK (denoted class −1), giving 143 and 64 cases, respectively. From a classification standpoint, the boundaries between classes U and LK, or S and HK, are less clear and defined compared to the boundary between U and S. Accordingly, the ML models should be considered poorly performing when a MOF of class S is mistakenly predicted as U, as opposed to a prediction of HK. For the three-class model, only U and LK were combined, giving the following three classes: S (denoted as 1), HK (denoted as 0) and LK + U (denoted as −1), as illustrated in Fig. 3. Such a grouping also makes sense from an application standpoint—a MOF of type LK is probably unsuitable for industrial processes in which small amounts of water cannot be avoided[12,26].

Beyond this Burtch dataset of 207 MOFs, which was specifically used for ML model development, two additional datasets were collected from the literature. The first of these consists of 10 MOFs synthesized and assessed for their water stability after publication of the Burtch dataset. These 10 cases allow for an unbiased evaluation of the ML models. The second dataset includes 88 MOFs for which no water stability measurements have been reported, but have been found to be useful for other applications, such as C_2H_4 or CO_2 capture. This dataset will be used to screen for MOF candidates likely to exhibit high water stability.

MOF feature set and dimensionality reduction. To build accurate and reliable ML models, it is important to include relevant features that collectively capture the water stability trends across different families of MOFs. The features should uniquely represent a MOF, and be readily available for new cases. Thus, in line with the general definition of a MOF, we used three sets of chemical descriptors: (1) the metal set, to capture information about the metal node(s), (2) the linker set, representing the organic ligand(s), and finally (3) the molar set, which encodes the molecular ratios of the linkers and the O, OH and H_2O species with respect to the metal nodes. The different subtypes of descriptors included within each set, along with their counts, are provided in Table 1.

Starting from a MOF activated formula unit, which is typically available via various empirical methods and is often reported in the literature, we extracted their constituent metal ions, organic ligands and molar ratios. Although commonly available chemical properties (Table 1) were used to describe the metal ions, the organic ligands were converted to their corresponding canonical SMILES representation[27], from which a hierarchy of features were derived. Based on our previous experience in fingerprinting polymers[28,29], we used hierarchical descriptors that capture different geometric and chemical information about ligands at multiple length scales. At the atomic scale, a count of a predefined set of motifs[30] consisting of atomic triplets (for example, C3–O1–N1, where C3, O1 and N1 define three coordinated C (two single and one double bond) and single coordinated O (double bond) and N (triple bond) atoms, respectively) are included. At a slightly larger length scale, quantitative structure–property relationship (QSPR) descriptors, often used in chemical and biological sciences, and implemented in the RDKit Python library, were used[31,32]. Finally, at the highest length scale, 'morphological descriptors' such as length of the largest side chain, shortest topological distance between rings and so on were considered. More details on the different hierarchical descriptors are provided in Supplementary Table 1 and our previous works[28,33]. In cases where multiple metal ions and ligands are present, the descriptors were obtained by taking a weighted molar average of all individual species. Finally, to uniquely represent a MOF, we added four features corresponding to the molar set described above and in Table 1.

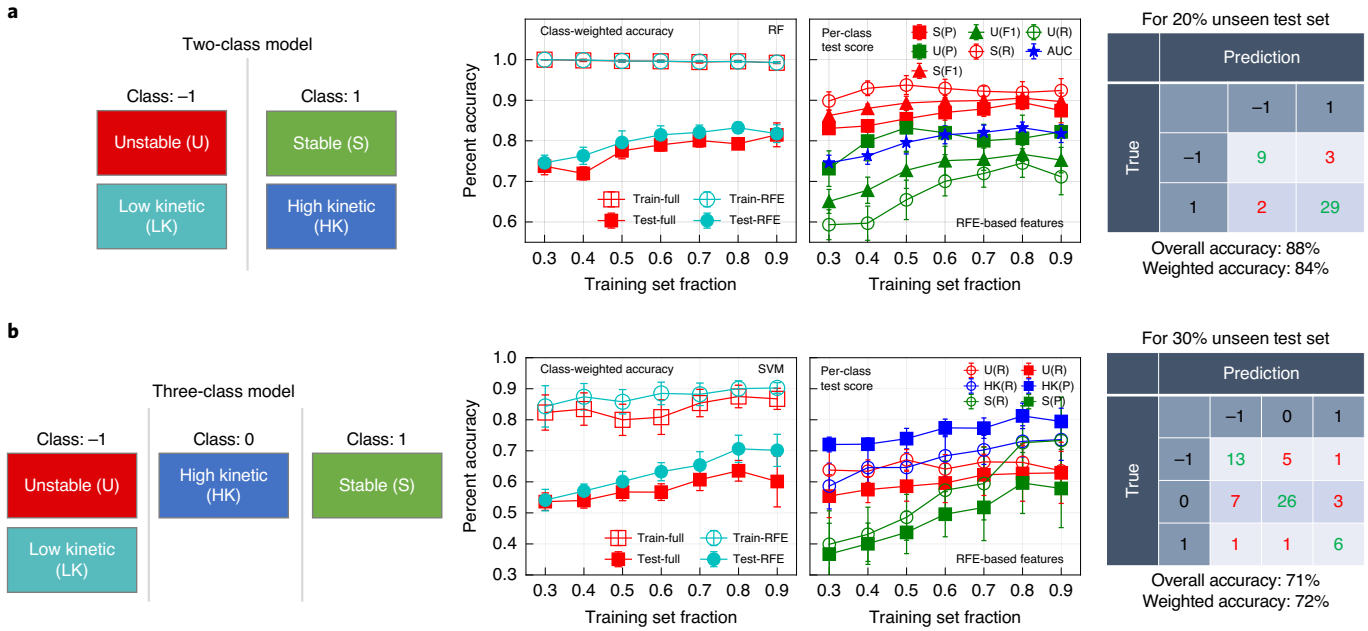


Fig. 3 | Performance of the classification models to predict water stability in MOFs. a,**b**, Class definitions, learning curves and an example confusion matrix for the two-class (**a**) and three-class (**b**) models. In the learning curves, error bars indicate 1σ standard deviation. AUC, area under the curve.

Table 1 | The MOF chemical descriptors used to numerically represent a MOF to learn ML-based water stability models

Category	Descriptors	Count
Metal node	Valency, atomic radius[34], affinity[35], ionization potential[35], electronegativity[36]	5
	Atomic triples	99
Organic linker	QSPR	32
	Morphological	8
Molar ratio	Linker, O, OH and H_2O w.r.t. metal	4

To retain only the relevant features, and to simplify the learning process, recursive feature elimination (RFE) using a support vector machine (SVM) algorithm (with fivefold cross-validation (CV) and linear kernel) was performed on the initial 149-dimensional fingerprint and the Burtch dataset. RFE recursively eliminates irrelevant features by ranking their importance and pruning the least important ones (one at a time in this work)[37]. The feature importance is obtained from the ML model itself (here, coefficients of the linear kernel SVM model), and the pruning procedure is repeated until the feature set with minimum CV score is achieved. For both the two-class and three-class models, RFE not only increased the model accuracy, but also reduced the model dimensionality from 149 to 37 and 29, respectively. The post-RFE reduced feature sets for both models contained many (~25) common descriptors, providing more confidence to the dimensionality reduction step performed here (the complete set of features is provided in Supplementary Table 1). Furthermore, features from each of the different subcategories were retained post-RFE, highlighting the need for capturing information at multiple length scales.

ML algorithms. Three ML classification methods were tested in this work: SVMs, random forest (RF) and gradient boosting (GB). Each method was used to train two MOF water stability classification models (two-class and three-class) using the Burtch dataset, with their respective hyperparameters determined using fivefold CV. For this work, SVM, RF and GB classifier libraries were used as implemented in the scikit-learn Python package[38].

SVM is a non-probabilistic binary linear classifier, in which the hyperplane or the classification boundary separating any two classes (for example, stable or unstable MOFs) is obtained by maximizing the margin between some special data points, called support vectors, and the hyperplane. Although, by definition, SVM performs linear classification, a kernel trick is often employed to obtain nonlinear classification boundaries. For this work, the SVM method with the radial basis function (RBF) was used. The SVM hyperparameters, RBF length scale and the regularization C parameters were estimated by minimizing the validation error during fivefold CV, which better generalizes the models and avoids overfitting.

Both RF and GB fall under the umbrella of ensemble methods, which are often the winning solutions in ML competitions. RF is an ensemble of decision trees that averages predictions from a large group of 'weak models' to result, overall, in a better prediction. The main hyperparameters for RF include the number of decision trees and the count of features accessible to an individual decision tree. Similarly, GB builds a set of additive models in a stagewise manner, wherein the next predictor is fit to the residual errors made by the previous predictor. The GB hyperparameter optimized in this work was the number of predictors.

To tackle the problem of class imbalance, the models were trained by minimizing the class-weighted accuracies. The performances of the ML models were evaluated using overall and class-weighted accuracy and per-class recall, precision and F1 (the harmonic mean of precision and recall) scores. To estimate prediction errors on unseen data, learning curves were generated by varying the sizes of the training and test sets. Test sets were obtained by excluding the training points from the Burtch dataset. Additionally, for each random test–train split, statistically meaningful results were obtained by averaging over 10 runs. Another dataset of recently reported MOFs (containing 10 points) was not included in the learning process and was used solely for model evaluation purposes.

Table 2 | Model validation on recently synthesized MOFs

ID	Common name	Activated formula	True stability	Two-class ML prediction	Three-class ML prediction
1	CAU-1(Al)	$[Al_4(OH)_2(OCH_3)_4(H_2N\text{-}bdc)_3]$	S	**S,HK**	HK
2		$[Cd_2(TBA)_2(bipy)(DMA)_2]$	LK	S,HK	S
3		$[La_2(pyzdc)_3(H_2O)_4]$	S	**S,HK**	**S**
4	MAF-X25ox	$[Mn_2(OH)Cl_2(bbta)]$	HK	**S,HK**	**HK**
5	MIL-68	Fe(OH)(bdc)	HK	**S,HK**	**HK**
6	MIL-160	$Al(O_5C_6H_2)(OH)$	S	**S,HK**	HK
7	Na-HPAA	$Na_2(OOCCH(OH)PO_3H)(H_2O)_4$	HK	**S,HK**	S
8	NU-1100	$Zr_6(OH)_4(OH)_4(L)_4$	HK	**S,HK**	**HK**
9	PCN-230	$Zr_6(OH)_4O_4(TCP\text{-}3)_3DMF_{30}(H_2O)_{10}$	HK	**S,HK**	**HK**
10	PCP-33	$(Cu_4Cl)(BTBA)_8((CH_3)_2NH_2)(H_2O)_{12}$	HK	**S,HK**	**HK**

Shown is a comparison of water stability predictions using two-class and three-class models against measurements reported in the literature for MOFs available after publication of the Burtch dataset. Text bolding in the 'Two-class prediction' and 'Three-class prediction' columns reflects if the model predictions were correct (bold) or not (non-bold). In increasing order of the MOF IDs, the following references were used to determine the true water stability[18,20,39–46].

Model performance and validation

Figure 3 presents the performance results for the two- and three-class MOF classification models. The learning curves provide the class-weighted accuracy (with error bars denoting 1σ deviation) on the training and test sets, using both the initial 149-dimensional (labelled '-Full') and reduced (labelled '-RFE') feature sets. We include results for the best performing ML methods, that is, RF for the two-class and SVM for the three-class (the comparative performances for the different algorithms are provided in Extended Data Figs. 2 and 3). From the learning curves, it is clear that the RFE dimensionality reduction scheme resulted in a better set of features by eliminating redundant cases. This caused an improvement in the model performance: the RFE test accuracy increased from 80 to 83% and 64 to 71% for the two- and three-class models, respectively. As expected, the test accuracy for the two- and three-class models (for both the RFE and full feature set) increased when the training set included more cases. The RFE-based models reached convergence in the test accuracy at 83% (two-class) and 71% (three-class). Owing to the class imbalance, the corresponding overall (unweighted) accuracy was found to converge at slightly higher values of 86 and 72%. For the case when the training set constitutes 90% of the data, there are not enough test examples from different classes, especially from class −1, to calculate adequate error statistics. This case is only included in Fig. 3 for completeness.

Other important error metrics to consider, especially when dealing with an imbalanced classification problem, are the precision, recall and F1 score. For a class, the former is defined as the ratio of correctly labelled points (or true positives) divided by the total number of data points predicted to belong to that class. Recall is defined as the number of true positives divided by the total number of data points that actually belong to a class. F1 score is the harmonic mean of precision and recall, and is often used as an important metric for imbalanced data. Because all the model parameters were optimized to maximize the class-weighted accuracy or recall, the classes with lower representation (that is, class −1) have a higher recall rate than precision, while the opposite is true for classes with higher representation. This is because the class-weighted accuracy metric penalizes the classifier more when it incorrectly classifies a data point from the underrepresented class. As a result, the classifier learns to predict a larger number of cases as belonging to the underrepresented class, thereby lowering the precision in these cases. For example, in the three-class model, the precision and recall values for classes with increasing order of representation were 60, 73 for class 1; 63, 66 for class −1 and 81, 73 for class 0. However, all of the recall, precision and F1 scores for the different classes were found to reasonable. For example, for the two- and three-class models, the F1 scores for the underrepresented classes were 76 and 63%, respectively, although they constituted only 30 and 12% of the overall data. This clearly demonstrates that the ML models developed here are not biased towards the more represented classes. In Fig. 3, example confusion matrices for both the two- and three-class models also confirmed the accuracy trends discussed above. Additionally, for the three-class model, a higher misclassification rate for the neighbouring classes can be observed—when predicted incorrectly, class 1 and −1 points are classified as class 0, but not each other. The high classification accuracy, recall and precision rate achieved by the two models, along with the misclassification preference in the neighbouring classes for the three-class model, suggest that good quality surrogate models have indeed been learned.

To further validate the generality and accuracy of our water stability models, we used the two- and three-class models trained on the entire Burtch dataset of 207 points to predict the water stability for 10 MOFs reported after the year 2014. The same classification criteria used for the Burtch dataset were used to determine the true water stability values (S, HK, LK or U) of these 10 MOFs. This comparative exercise presents a clean and unbiased evaluation of the ML scheme and the final two- and three-class models trained in this work. From the results presented in Table 2, it is clear that both models performed well. The two- and three-class models had 9 out of 10 and 6 out of 10 predictions correct, respectively—3 of 4 incorrect predictions in the case of the three-class models are between the more similar S and HK classes, which can still be considered acceptable. Furthermore, both models made consistent stability predictions; that is, for cases where the two-class model predicted class 1 (S or HK), the three-class model also predicted either class 1 (S) or class 0 (HK). Both models incorrectly predicted the MOF $Cd_2(TBA)_2(bipy)(DMA)_2$ to be stable, although our literature analysis suggests it has LK water stability.

Chemical insights from ML models

Past works have already suggested some chemical trends that promote water stability in MOFs, including inertness of the metal node, stronger and larger number of metal–ligand bonds through higher ligand basicity and higher-valent metal nodes, and higher steric shielding[16]. However, in the following we employ the developed ML methods to mine more such chemical trends or insights. For this, we first identified the most important features using the two-class RF models. In RF, the relative importance of a feature can be defined using the relative rank (or depth) of that feature when used as a decision node in a tree, because features used at the top

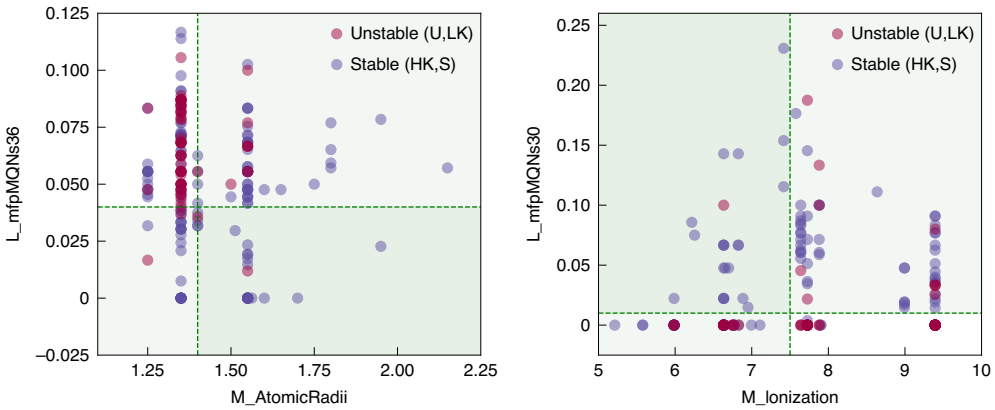


Fig. 4 | Mined chemical trends. Water stability trends in MOFs from the Burtch dataset using the important features obtained from the two-class RF models. MOFs with metal ions of large atomic radius (M_AtomicRadii > 1.4 Å) and lower ionization potential (M_Ionization < 7.5 eV), and ligands with low MQNS36 (L_mfpMQNs36 > 0.04) and high MQNS30 (L_mfpMQNs30 > 0.01) are shown to capture many water-stable MOFs (of the Burtch dataset).

Table 3 | Screened water-stable MOFs

ID	MOF	DOI	Two-class ML prediction	Three-class ML prediction
1	$[Ca_3(NTB)_2(DEF)_2(H_2O)_2]$	10.1134/S0022476619110192	1 (S, HK)	1 (S)
2	$[Ca_3(BTB)_2(NMP)_2(H_2O)_2]$	10.1134/S0022476619110192	1 (S, HK)	1 (S)
3	$[Cd_3(OABDC)_2(eurea)_4]$	10.1080/00958972.2016.1180371	1 (S, HK)	1 (S)
4	$[Mn_2(HL_6)(DMF)(H_2O)]$	10.1039/C6DT02846B	1 (S, HK)	1 (S)
5	$[Ce(BTC)(H_2O)]$	10.1002/ejic.201000541	1 (S, HK)	1 (S)

Shown are the top five identified 'new' water-stable MOFs using the two-class RF and three-class SVM models. The meaning of class definition 1 for the two-class and three-class models are provided in parentheses and can be seen from Fig. 3. More details on the MOFs can be found by using the DOIs.

of a tree contribute towards the final prediction for a larger fraction of the input samples. Based on this philosophy, the relative importance of different features towards MOF water stability prediction is presented in Extended Data Fig. 4. As expected, the atomic radius and ionization potential of the metal ion and the ligand versus metal ratio were found to be quite important. However, a variety of molecular quantum numbers (MQNs)[47] for the ligands were also found to be important descriptors; topological features, such as the number of cyclic divalent nodes (MQNs30) or six-member rings (MQNs36), and polarity-based descriptors, such as the number of hydrogen-bond acceptor sites (MQNs20), were among the top features.

Next, using these top identified features, we searched for simple chemical rules for water stability in MOFs. As illustrated in Fig. 4 by blue shaded regions, the MOFs from the Burtch dataset containing metal ions of large atomic radius and lower ionization potential, and ligands with low MQNS36 and high MQNS30 were found to display high water stability. In particular, 66 of 75 (88%), 71 of 82 (87%), 75 of 90 (83%) and 51 of 55 (93%) cases were found to be stable (S or HK) when the metal ion atomic radius was >1.4 Å, the metal ion ionization potential was <7.5 eV, the ligand had MQNS36 <0.04 and the ligand had MQNS30 >0.01, respectively. Furthermore, when considering two properties at a time, 21 of 22 (95%) MOFs with metal ion atomic radius >1.4 Å and the ligand with MQNS36 <0.04, or metal ion ionization potential <7.5 eV and the ligand with MQNS36 <0.04 were found to be stable. Similar insights derived from linear correlations between MOF features and water stability are provided in Extended Data Fig. 5. The identified trends provide new insights on water stability in MOFs and can serve as guidelines for future exploration of stable MOFs.

Screening new water-stable MOFs

Having established the accuracy levels achieved by the models, we next used them to screen MOF candidates with unknown water stabilities. This test demonstrates the ease with which instant water stability predictions can be made for 'new' MOFs, if given only their activated formula unit. Eighty-eight MOFs were selected from the literature based on their adsorption, separation or catalytic capabilities, and predictions were made for each using the two- and three-class models. In Table 3, we list the top five candidates, rank-ordered based on their probability to be of class 1 (that is, S or HK for the two-class model and S for the three-class model), while predictions for all 88 candidates are provided in Supplementary Table 2. This exemplifies how the developed ML models can be used to screen or prioritize MOF synthesis, and efficiently explore water-stable MOFs. We note that, for most cases (75 of 88), the predictions for the two models were consistent. Although the true water-stability nature of these 88 new candidate MOFs is not entirely unknown, following a literature search, we found some information regarding 12 of these candidates. Among those, the ML models correctly identified the two stable MOFs, that is, $[Ca(C_4O_4)(H_2O)]$[48] and [AgTPB][49]. Furthermore, 5 of 6 HK stability MOFs were correctly identified by the two-class model, while only 1 of 4 of the cases was correctly classified as U. Although these cases are too limited to draw statistical conclusions, they suggest that the developed ML models are indeed helpful in screening water-stable MOFs.

A caveat should be noted for the developed ML models. Analysis of the model predictions on the 88 new MOFs suggested a bias towards water-stable MOFs. Although the two-class model predicted only ~20% of the MOFs to be unstable, the three-class model predictions had a distribution of 28, 32 and 40% of new

MOFs to be in category U or LK, HK and S, respectively. Two potential sources could lead to such a bias in the model predictions. First, the ML models are biased towards water-stable MOFs as the Burtch training dataset is dominated by MOFs of classes HK and S. Second, the dataset of 88 new MOFs, although collected randomly from the literature, suffers from an inherent publication bias, that is, selective reporting of only those MOFs that have some degree of stability. Although the lack of a truly unbiased test dataset precludes resolution on the source of bias, we note that the classification distribution of the 88 new MOFs matches reasonably well with that of the Burtch dataset, which was taken from a comprehensive review paper reporting an unbiased state of this field in 2014, without any presumption that this dataset would be later used for ML studies. Additionally, some of the best ML practices (class-balanced accuracy and CV) have been adopted to avoid class bias in the models.

It is important to note that, because all the features used in this work (metal, linker and molar set) can be derived using only the MOF formula unit, no structural information is required a priori, making these proposed ML models versatile and easily applicable. However, this also highlights a limitation of the current models; that is, they cannot differentiate between different phases of a MOF or when ligand arrangement varies despite preserving the same metal–linker molar ratio. Although this issue could be resolved by expanding our feature set to include structural information (for example, pore limiting diameter and density), all of which is expected to improve the accuracy of the current water-stability models, adding structural features will limit the applicability of the ML model to only those MOFs for which accurate structural measurements are available. For this reason, we opted to not add more MOF structure-based features to our models, instead choosing to rely on the more readily available information given by the MOF formula unit.

Conclusions

In summary, we have developed simple and generalized ML models to predict the water stability of MOFs. Two classification models (two-class and three-class) were learned using a dataset of experimentally determined water stabilities for 207 MOFs. These models provide a quick and inexpensive estimate of MOF water stability. To train the models, a comprehensive set of chemical features was compiled to capture information about the MOF metal node, the organic linker and their molar ratios. These feature sets were further refined using dimensionality reduction schemes. The classification models were trained using RF and the SVM algorithm, while their performance was evaluated through class-weighted accuracies and per-class precision and recall rates. Not only were the models used to predict the water stabilities of 10 recently reported MOFs that have had experimental stability measurements done, they were also used to screen new MOF candidates predicated to be stable under aqueous conditions. Overall, this work can be used for the rational design and screening of new MOFs with a desired level of water stability, as well as for obtaining a better fundamental understanding of MOF degradation behaviour.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

The MOF water-stability data (illustrated in Fig. 2) used to train the models were obtained from ref. [13]. The water-stability data used for validation (recent 10 MOFs) and screening (88 new MOFs) were obtained from the literature as cited in the Article. These datasets, including MOF features, are deposited at https://doi.org/10.5281/zenodo.4014333. Source data are provided with this paper.

Code availability

The machine learning training and prediction codes underlying this work are freely available for general use under GNU General Public Licence v3.0 and are deposited at https://doi.org/10.5281/zenodo.4014333.

Received: 5 November 2019; Accepted: 7 October 2020;
Published online: 9 November 2020

References

1. Yoon, J. W. et al. Selective nitrogen capture by porous hybrid materials containing accessible transition metal ion sites. *Nat. Mater.* **16**, 526–531 (2017).
2. Adil, K. et al. Gas/vapour separation using ultra-microporous metal–organic frameworks: insights into the structure/separation relationship. *Chem. Soc. Rev.* **46**, 3402–3430 (2017).
3. Mason, J. A., Veenstra, M. & Long, J. R. Evaluating metal–organic frameworks for natural gas storage. *Chem. Sci.* **5**, 32–51 (2014).
4. Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal–organic frameworks. *Science* **341**, 1230444 (2013).
5. Dusselier, M. & Davis, M. E. Small-pore zeolites: synthesis and catalysis. *Chem. Rev.* **118**, 5265–5329 (2018).
6. Yang, D. & Gates, B. C. Catalysis by metal–organic frameworks: perspective and suggestions for future research. *ACS Catal.* **9**, 1779–1798 (2019).
7. Furukawa, H. et al. Ultrahigh porosity in metal–organic frameworks. *Science* **329**, 424–428 (2010).
8. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal–organic framework. *Nature* **402**, 276–279 (1999).
9. Cohen, S. M. Postsynthetic methods for the functionalization of metal–organic frameworks. *Chem. Rev.* **112**, 970–1000 (2011).
10. Zhang, Y.-B. et al. Introduction of functionality, selection of topology and enhancement of gas adsorption in multivariate metal–organic framework-177. *J. Am. Chem. Soc.* **137**, 2641–2650 (2015).
11. Kaye, S. S., Dailly, A., Yaghi, O. M. & Long, J. R. Impact of preparation and handling on the hydrogen storage properties of Zn_4O(1,4-benzenedicarboxylate)$_3$ (MOF-5). *J. Am. Chem. Soc.* **129**, 14176–14177 (2007).
12. Ma, D., Li, Y. & Li, Z. Tuning the moisture stability of metal–organic frameworks by incorporating hydrophobic functional groups at different positions of ligands. *Chem. Commun.* **47**, 7377–7379 (2011).
13. Burtch, N. C., Jasuja, H. & Walton, K. S. Water stability and adsorption in metal–organic frameworks. *Chem. Rev.* **114**, 10575–10612 (2014).
14. Schoenecker, P. M., Carson, C. G., Jasuja, H., Flemming, C. J. & Walton, K. S. Effect of water adsorption on retention of structure and surface area of metal–organic frameworks. *Ind. Eng. Chem. Res.* **51**, 6513–6519 (2012).
15. Bosch, M., Zhang, M. & Zhou, H.-C. Increasing the stability of metal-organic frameworks. *Adv. Chem.* **2014**, 182327 (2014).
16. Rieth, A. J., Wright, A. M. & Dinca, M. Kinetic stability of metal–organic frameworks for corrosive and coordinating gas capture. *Nat. Rev. Mater* **4**, 708–725 (2019).
17. ul Qadir, N., Said, S. A. & Bahaidarah, H. M. Structural stability of metal–organic frameworks in aqueous media–controlling factors and methods to improve hydrostability and hydrothermal cyclic stability. *Micropor. Mesopor. Mater.* **201**, 61–90 (2015).
18. Plessius, R. et al. Highly selective water adsorption in a lanthanum metal–organic framework. *Chem. Eur. J.* **20**, 7922–7925 (2014).
19. Qin, L. et al. A water-stable metal–organic framework of a zwitterionic carboxylate with dysprosium: a sensing platform for Ebolavirus RNA sequences. *Chem. Commun.* **52**, 132–135 (2016).
20. Liu, T.-F. et al. Topology-guided design and syntheses of highly stable mesoporous porphyrinic zirconium metal–organic frameworks with high surface area. *J. Am. Chem. Soc.* **137**, 413–419 (2014).
21. Zhang, J.-P., Zhu, A.-X., Lin, R.-B., Qi, X.-L. & Chen, X.-M. Pore surface tailored SOD-type metal–organic zeolites. *Adv. Mater.* **23**, 1268–1271 (2011).
22. Nijem, N. et al. Water cluster confinement and methane adsorption in the hydrophobic cavities of a fluorinated metal–organic framework. *J. Am. Chem. Soc.* **135**, 12615–12626 (2013).
23. Yang, C. et al. Fluorous metal–organic frameworks with superior adsorption and hydrophobic properties toward oil spill cleanup and hydrocarbon storage. *J. Am. Chem. Soc.* **133**, 18094–18097 (2011).
24. Shih, Y.-H. et al. A simple approach to enhance the water stability of a metal–organic framework. *Chem. Eur. J.* **23**, 42–46 (2017).
25. Taylor, J. M., Vaidhyanathan, R., Iremonger, S. S. & Shimizu, G. K. Enhancing water stability of metal–organic frameworks via phosphonate monoester linkers. *J. Am. Chem. Soc.* **134**, 14338–14340 (2012).

26. Canivet, J., Fateeva, A., Guo, Y., Coasne, B. & Farrusseng, D. Water adsorption in MOFs: fundamentals and applications. *Chem. Soc. Rev.* **43**, 5594–5617 (2014).
27. OpenSMILES; http://opensmiles.org
28. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer genome: a data-powered polymer informatics platform for property predictions. *J. Phys. Chem. C* **122**, 17575–17585 (2018).
29. Mannodi-Kanakkithodi, A. et al. Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond. *Mater. Today* **21**, 785–796 (2018).
30. Huan, T. D., Mannodi-Kanakkithodi, A. & Ramprasad, R. Accelerated materials property predictions and design using motif-based fingerprints. *Phys. Rev. B* **92**, 014106 (2015).
31. Nantasenamat, C., Isarankura-Na-Ayudhya, C. & Prachayasittikul, V. Advances in computational methods to predict the biological activity of compounds. *Expert Opin. Drug Discov.* **5**, 633–654 (2010).
32. *RDKit Open Source Toolkit for Cheminformatics*; http://www.rdkit.org/ (accessed 3 September 2019).
33. Jha, A., Chandrasekaran, A., Kim, C. & Ramprasad, R. Impact of dataset uncertainties on machine learning model predictions: the example of polymer glass transition temperatures. *Model. Simul. Mater. Sci. Eng.* (2018); https://doi.org/10.1088/1361-651X/aaf8ca
34. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).
35. Haynes, W. M. *CRC Handbook of Chemistry and Physics* (CRC Press, 2014).
36. Pauling, L. The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **54**, 3570–3582 (1932).
37. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**, 389–422 (2002).
38. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
39. Xie, L., Liu, D., Huang, H., Yang, Q. & Zhong, C. Efficient capture of nitrobenzene from waste water using metal–organic frameworks. *Chem. Eng. J.* **246**, 142–149 (2014).
40. Wang, D., Zhang, L., Li, G., Huo, Q. & Liu, Y. Luminescent MOF material based on cadmium(II) and mixed ligands: application for sensing volatile organic solvent molecules. *RSC Adv.* **5**, 18087–18091 (2015).
41. Liao, P.-Q. et al. Drastic enhancement of catalytic activity via post-oxidation of a porous Mn^{II} triazolate framework. *Chem. Eur. J.* **20**, 11303–11307 (2014).
42. Jing, F. et al. Mil-68(Fe) as an efficient visible-light-driven photocatalyst for the treatment of a simulated waste-water contain Cr(VI) and malachite green. *Appl. Catal. B Environ.* **206**, 9–15 (2017).
43. Cadiau, A. et al. Design of hydrophilic metal organic framework water adsorbents for heat reallocation. *Adv. Mater.* **27**, 4775–4780 (2015).
44. Bazaga-Garcia, M. et al. Tuning proton conductivity in alkali metal phosphonocarboxylates by cation size-induced and water-facilitated proton transfer pathways. *Chem. Mater.* **27**, 424–435 (2015).
45. Gutov, O. V. et al. Water-stable zirconium-based metal–organic framework material with high-surface area and gas-storage capacities. *Chem. Eur. J.* **20**, 12389–12393 (2014).
46. Duan, J., Jin, W. & Krishna, R. Natural gas purification using a porous coordination polymer with water and chemical stability. *Inorg. Chem.* **54**, 4279–4284 (2015).
47. Nguyen, K. T., Blum, L. C., Van Deursen, R. & Reymond, J.-L. Classification of organic molecules by molecular quantum numbers. *ChemMedChem* **4**, 1803–1805 (2009).
48. Lin, R.-B. et al. Molecular sieving of ethylene from ethane using a rigid metal–organic framework. *Nat. Mater.* **17**, 1128–1133 (2018).
49. Sun, Y. & Han, H. A novel 3D Ag^{I} cationic metal–organic framework based on 1,2,4,5-tetra(4-pyridyl) benzene with selective adsorption of CO_2 over CH_4, H_2O over C_2H_5OH, and trapping $Cr_2O_7^{2-}$. *J. Mol. Struct.* **1194**, 73–77 (2019).

Acknowledgements

This work was supported as part of the Center for Understanding and Control of Acid Gas-Induced Evolution of Materials for Energy (UNCAGE-ME), an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0012577. C.C. gratefully acknowledges a fellowship from the Achievement Rewards for College Scientists (ARCS) Foundation. R.B. acknowledges insightful discussions with D.S. Sholl.

Author contributions

R.B. and R.R. initiated this research project. R.B. developed and analysed the ML models. C.C. and T.G.E. contributed to data collection. All co-authors contributed to the model analysis, discussions and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

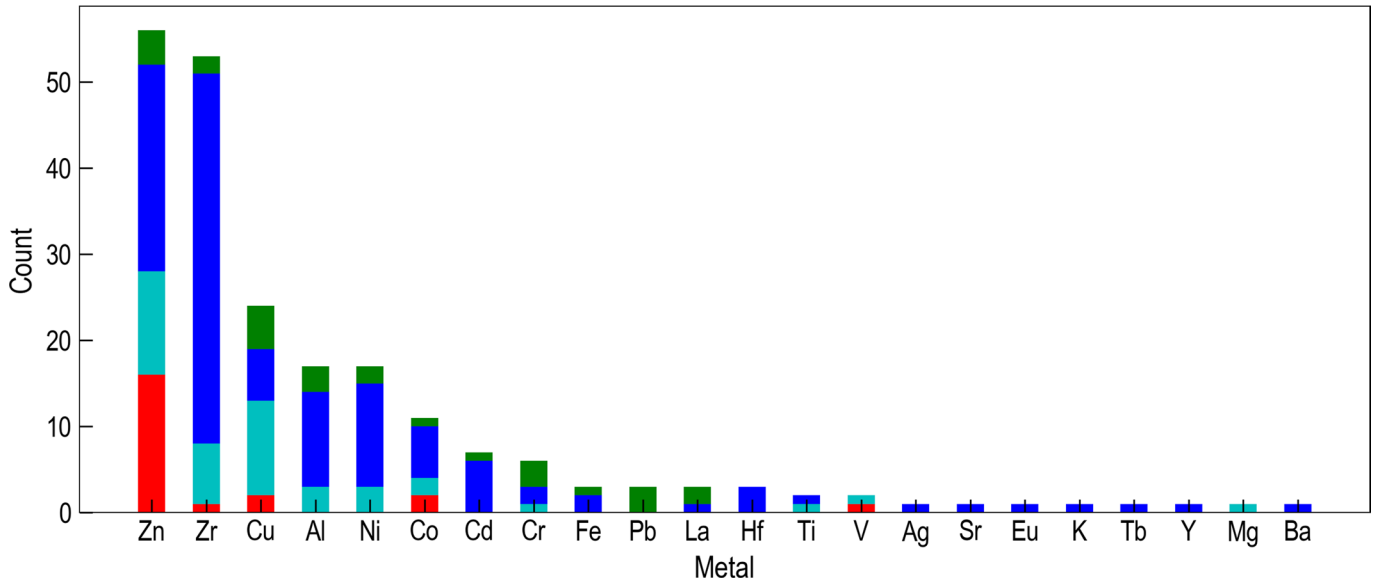
Extended data is available for this paper at https://doi.org/10.1038/s42256-020-00249-z.

Supplementary information is available for this paper at https://doi.org/10.1038/s42256-020-00249-z.

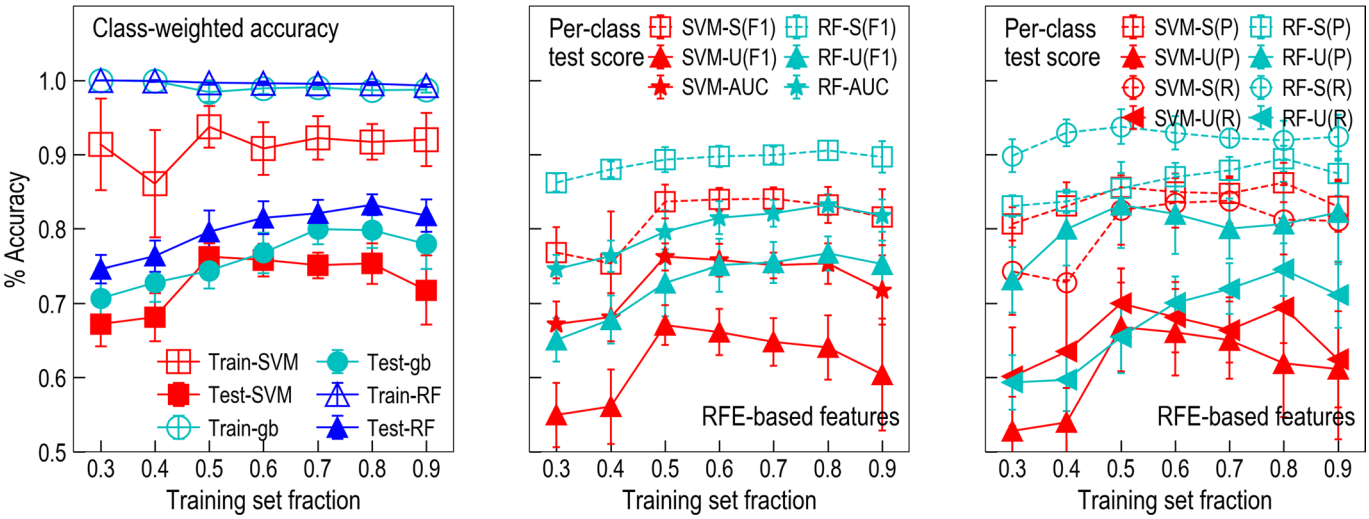
Correspondence and requests for materials should be addressed to R.R.

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

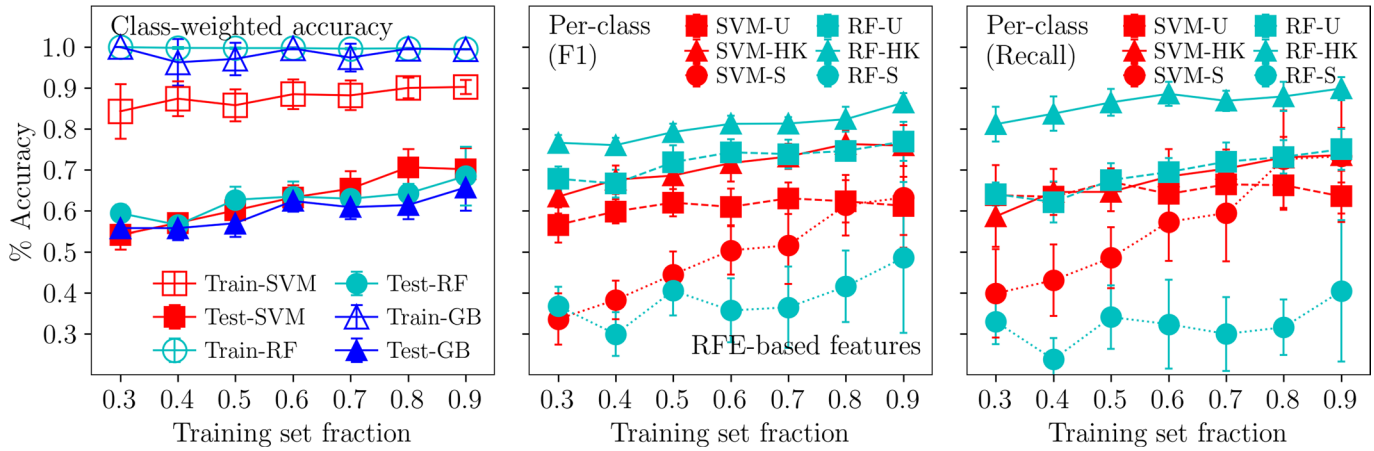
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



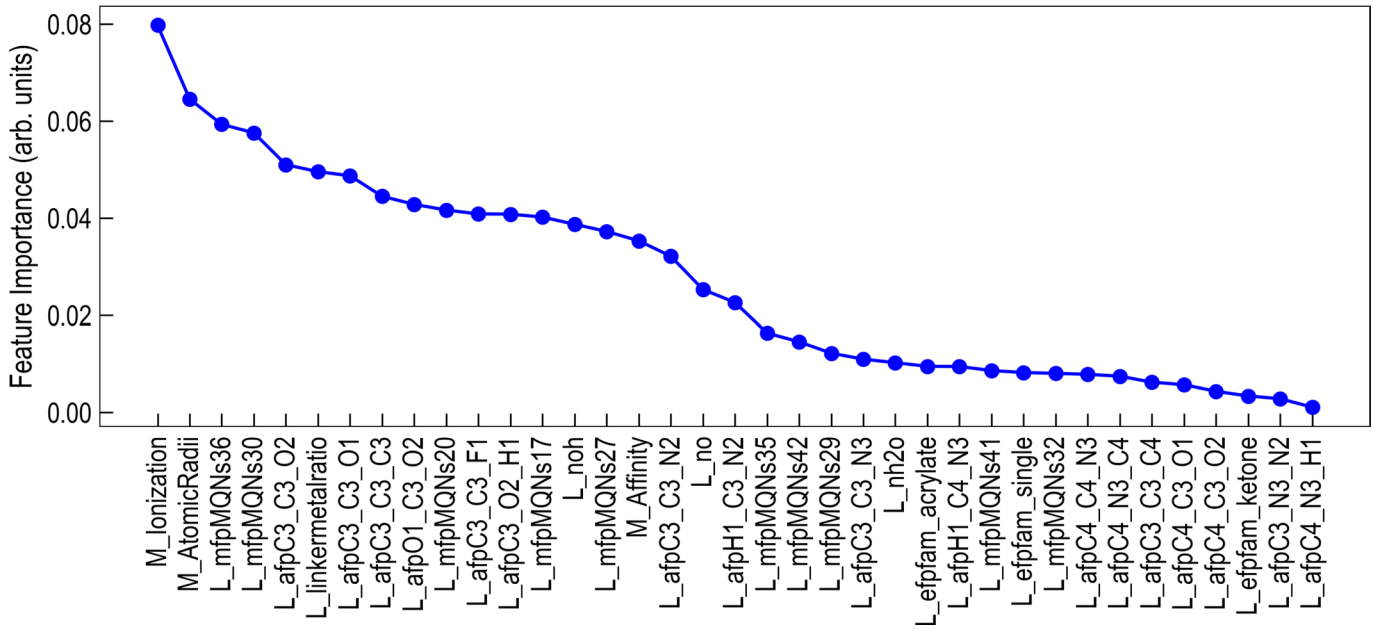
Extended Data Fig. 1 | Statistics on water stability in MOFs. Distribution of MOFs into 4 categories of water stability based on the constituting metal node.



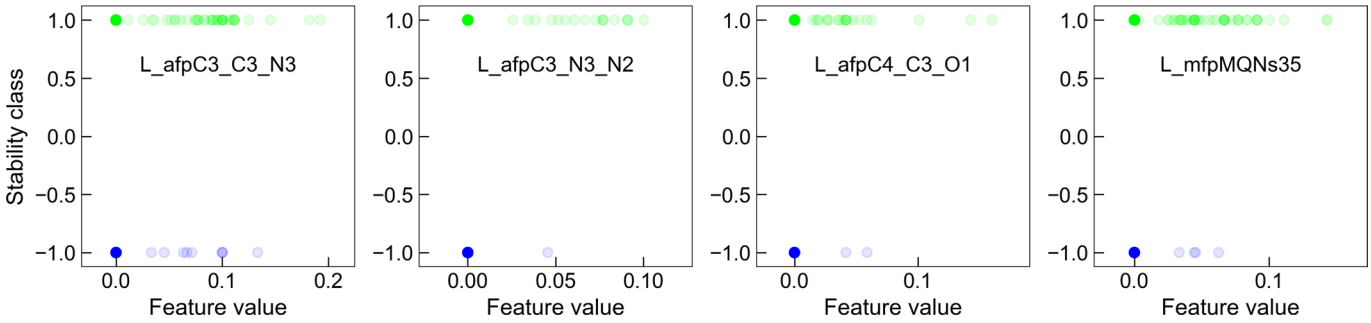
Extended Data Fig. 2 | Performance comparison of ML algorithms for 2-class model. Performance comparison of SVM, RF and GB methods for the 2-class model ('S', stable and 'U', unstable MOFs) using the RFE based reduced feature set. Left panel shows the overall class-weighted accuracies, while the right two panels show the per-class test scores, that is F1, area under the ROC curve (AUC), precision (P) and recall (R), for the RF and SVM models. The RF model can be seen to outperform in all accounts and was selected as the 2-class model in this work.



Extended Data Fig. 3 | Performance comparison of ML algorithms for 3-class model. Performance comparison of SVM, RF and GB methods for the 3-class model ('S', stable, 'HK', high kinetic stable, and 'U', unstable MOFs) using the RFE based reduced feature set. Left panel shows the overall class-weighted accuracies, while the right two panels respectively show the per-class F1 and recall scores, for the RF and SVM models. The RF model can be seen to have poor performance for the underrepresented stable (S) class, although it was trained to maximize the class-weighted accuracy. Similar results were found for GB algorithm as well. Thus, SVM with best performance for all classes was selected as the 3-class model in this work.



Extended Data Fig. 4 | Important MOF water stability descriptors. Relative feature importance as extracted from the random forest (RF) 2-class model. The feature importance in case of RF is based on the concept of mean decrease in impurity (MDI), as explained here (G. Louppe, Understanding Random Forests: From Theory to Practice, PhD Thesis, U. of Liege, 2014). The features with relatively high importance were selected to mine important chemical trends of water stability in MOFs. The first letter of the descriptor, that is, M or L, denotes the metal or the ligand associated features, respectively (see main article for details). Features with high importance were used to derive important stability trends as discussed in the main article.



Extended Data Fig. 5 | Correlation between MOF water stability and its descriptors. A subset of post-RFE features were analyzed to see if linear correlations between MOF water stability for the case with two classes (S+HK and U+LK) and the features values could be used to derive some chemical trends. This figure suggests that the presence of certain chemical motifs, especially those containing N or ketone groups, and 5-member rings, tend to enhance the water stability in MOFs. Each marker in the figure represents a MOF from the Burtch data set. See Supplementary Information for details on the different descriptors.

nature research

Corresponding author(s): Rampi Ramprasad (A19111130)

Last updated by author(s): Nov 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The water stability data for MOFs was collected manually from literature. All references have been included in the main manuscript.
Data analysis	The SVM classification model was learned using SVC library in the scikit-learn package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The MOF water stability data (illustrated in Figure 2) used to train the SVM models was obtained from Ref. 14. The water stability data for more recent MOFs used to validate the ML models were obtained from literature cited in the main article.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*

Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging