Next-Generation MOFs for Atmospheric Water Harvesting: The Role of Machine Learning Techniques

M. Arjmandi[a], E. Aytaç[b], M. Khayet[a,c*], N. Hilal[d]

[a]Department of Structure of Matter, Thermal Physics and Electronics, Faculty of Physics, University Complutense of Madrid, Avda. Complutense s/n, 28040 Madrid, Spain

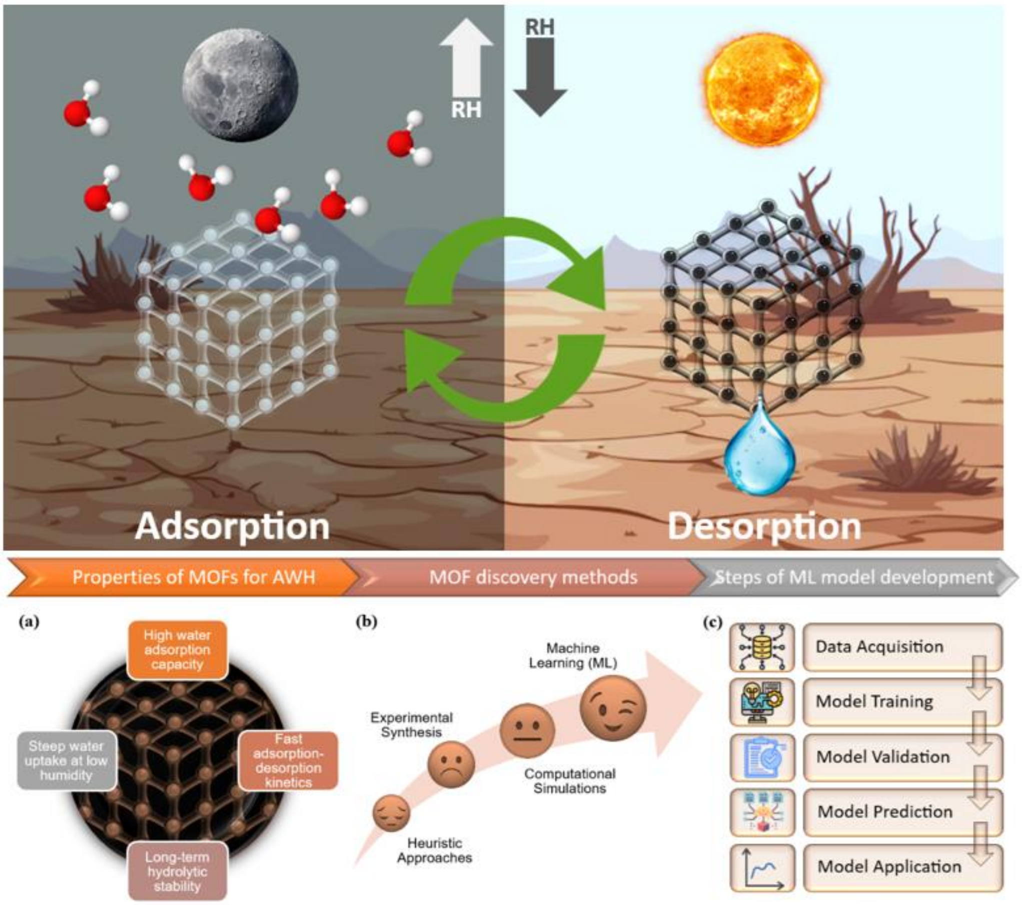
[b]Department of Environmental Engineering, Zonguldak Bülent Ecevit University, 67100 Zonguldak, Türkiye

[c]Madrid Institute for Advanced Studies in Water (IMDEA Water Institute), Avda. Punto Com N° 2, 28805 Alcalá de Henares, Madrid, Spain

[d]Water Research Center at New York University Abu Dhabi (NYUAD), P.O. Box 129188, Abu Dhabi 129188, United Arab Emirates

*Corresponding authors: khayetm@fis.ucm.es

Graphical Abstract



Abstract

Atmospheric Water Harvesting (AWH) using Metal-Organic Frameworks (MOFs) has emerged as a highly promising approach to mitigate water scarcity, especially in arid and semi-arid regions. The development of high-performance MOFs for AWH hinges on materials that exhibit optimal water uptake capacity, rapid adsorption-desorption kinetics, and robust hydrolytic stability. However, the structural complexity of MOFs and the inefficiencies of traditional experimental screening have made data-driven approaches, particularly machine learning (ML), increasingly indispensable for accelerating materials discovery. Among the ML techniques applied to MOF-based AWH, models such as Random Forest (RF), Random Forest Regression (RFR), Neighbor Component Analysis (NCA), Genetic Algorithms (GA), and Machine-Learned Atomistic Cluster Expansion (MACE) have demonstrated outstanding predictive performance. These models are especially valued for their ability to capture non-linear dependencies, improve interpretability, and optimize design strategies across diverse application domains. This review presents a comprehensive analysis of ML-assisted MOF discovery for AWH, focusing on the roles of explicit and latent descriptors, evaluation metrics, dataset curation challenges, and comparative model performance. By emphasizing the superior predictive capabilities of RF, MACE, NCA, RFR, and GA, this work highlights the transformative potential of ML in driving the rational design of next-generation MOFs for efficient and scalable AWH.

Keywords: *Metal Organic Framework, Machine Learning, Atmospheric Water Harvesting*

1. Introduction

Water scarcity is one of the most critical global challenges of the 21st century, impacting billions of people worldwide [1]. According to the United Nations, over two billion individuals currently reside in regions experiencing high water stress, a number expected to rise due to the compounding effects of climate change, over-extraction of groundwater, pollution, and population growth [2]. Freshwater resources, essential for human survival, agriculture, and industrial processes, are under increasing pressure and are rapidly depleting in many regions. In response, there is an urgent need to develop alternative water-harvesting technologies that are both efficient and sustainable.

Various technologies have been developed for water access, including Desalination [3, 4], Rainwater Harvesting (RWH) [5], and Atmospheric Water Harvesting (AWH) [6]. While desalination relies on existing water bodies and require extensive infrastructure, limiting their applicability to coastal regions, RWH depends on climate, weather events, catchment area and infrastructure and limited to health risks and social perceptions [7]. AWH has emerged as a promising alternative for freshwater access [6]. Unlike traditional sources such as rivers, lakes, and groundwater, the atmosphere represents a continuous yet diffuse reservoir of water vapor that can be harnessed through specialized technologies. The atmosphere holds an estimated 13000 trillion liters of water in the form of droplets and vapor at any given time, presenting a significant opportunity for mitigating water scarcity [8]. Current water collection strategies from air include fog harvesting [9], dewing/refrigeration [10], and sorption-based AWH [11]. Fog harvesting, inspired by nature, is efficient but requires high humidity and is geographically limited; while dewing and refrigeration methods are either passive and energy-efficient or active but energy-intensive. In contrast, sorption-based AWH offers a promising solution for arid regions, providing potable water with simple, mobile, and eco-friendly technology. This approach has gained significant attention as a potential method for augmenting freshwater supplies, particularly in areas where conventional water sources are scarce, unreliable, or contaminated. As illustrated in Fig. 1, sorption-based AWH captures water vapor at night when humidity is high and temperature is low, storing it in adsorbents. During the day, heat from solar energy, electricity, or magnetic fields releases the water vapor. The released vapor then condenses into liquid form and is collected for use [11].

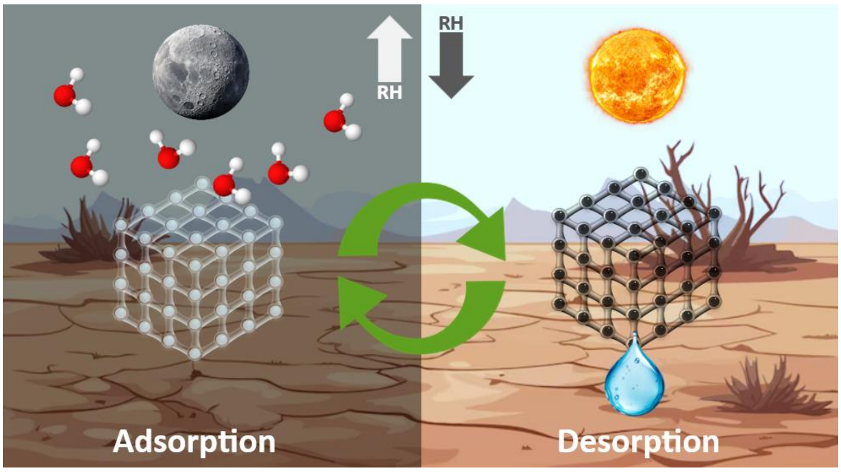


Fig. 1. Schematic of sorption-based AWH: Adsorbents capture water vapor and release it upon solar heating, leading to condensation and collection

The sorbent material used in AWH systems plays a fundamental role in determining their efficiency, water uptake capacity, and long-term operational performance [12]. To ensure the viability of AWH technologies, the materials must exhibit high water adsorption capacity, rapid adsorption and desorption kinetics, and exceptional stability under cyclic operation [12]. Various adsorbents have been developed for adsorption-related applications. As shown in Fig. 2, traditional adsorbents like zeolites and silica gel offer benefits, but they face limitations like high regeneration temperatures and low adsorption capacities, while hygroscopic salts and polymeric gels show promise but face challenges such as leakage and slow kinetics [13, 14, 15]. Polymeric gels, with tunable structures, enable moisture capture and energy conversion but struggle under low humidity conditions [16]. Among these, Metal-Organic Frameworks (MOFs) stand out due to their high porosity, diverse structures, and superior water adsorption properties, making them the focus of this review [8]. MOFs are a class of porous materials composed of metal ions or clusters coordinated with organic ligands, forming highly ordered, three-dimensional structures. They possess several advantages for AWH applications, including tunable porosity, large surface area, structural diversity, and exceptional water adsorption properties. These characteristics make MOFs highly attractive candidates for efficiently capturing water from the atmosphere. Among these, Metal-Organic Frameworks (MOFs) and Covalent Organic Frameworks (COFs) have both shown significant potential due to their large surface areas, ordered pore structures, and tunable functionalities [8, 12, 17] [18] [19] [20]. COFs, being entirely organic and metal-free, offer certain advantages such as chemical robustness, low density, and ease of structural modification for introducing hydrophilic sites [20]. However, they generally exhibit lower water uptake capacities and slower adsorption kinetics compared to MOFs. One key limitation arises from their lamellar structures, where strong van der Waals forces and π–π stacking interactions promote agglomeration. This agglomeration reduces the number of accessible active binding sites, thereby diminishing their water adsorption performance [21]. For instance, COF-432 has been reported to show significantly lower uptake compared to various MOFs [22] [23]. MOFs, with their metal-organic structures, often provide stronger interactions between water molecules and metal centers, leading to higher adsorption capacities and improved performance under low humidity and cyclic operation. In addition, many MOFs have been experimentally validated for their hydrolytic stability and long-term durability. Based on these comparative advantages, MOFs currently stand out as more promising candidates for AWH applications [24].

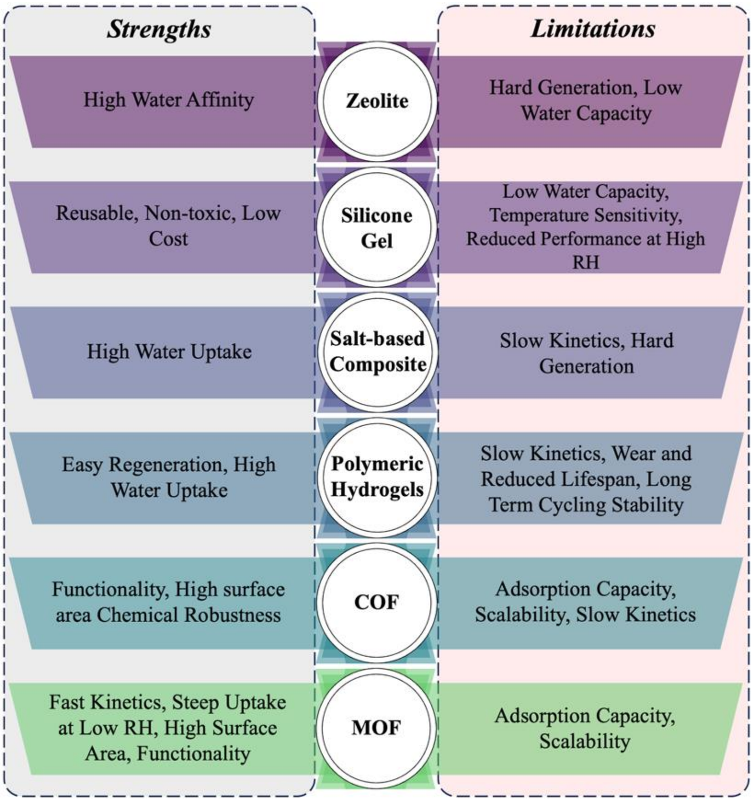


Fig. 2. Comparative evaluation of different classes of water adsorbents used in AWH systems, including zeolites, silica gel, hygroscopic salts, polymeric gels, COFs and MOFs. The figure highlights key strengths and limitations of each class in terms of water uptake capacity, regeneration requirements, kinetic performance, and operational stability.

Despite their promise, the discovery and optimization of MOFs for AWH remains challenging. The vast chemical space of MOFs, comprising potentially millions of distinct structures, renders exhaustive exploration impractical using traditional trial-and-error methods. As shown in Fig. 3a, designing MOFs for specific applications, such as AWH, requires identifying MOFs that achieve an optimal balance of properties, including high water adsorption capacity, fast adsorption-desorption kinetics, steep water uptake at low humidity, and long-term hydrolytic stability [12]. Thus, the main distinction between identifying new MOFs for the AWH process and discovering MOFs for other applications lies in the need for MOFs with multifaceted properties. Therefore, the discovery process for AWH-specific MOFs differs significantly, and using conventional methods such as heuristic approaches, experimental synthesis, and computational simulations (including Density Functional Theory (DFT) [25, 26] calculations, Grand Canonical Monte Carlo (GCMC) [25] simulations, response surface methodology (RSM) [27], analysis of variance (ANOVA) [28], and Molecular Dynamics (MD) [29] simulations would be highly time-consuming and resource-intensive (Fig. 3b).

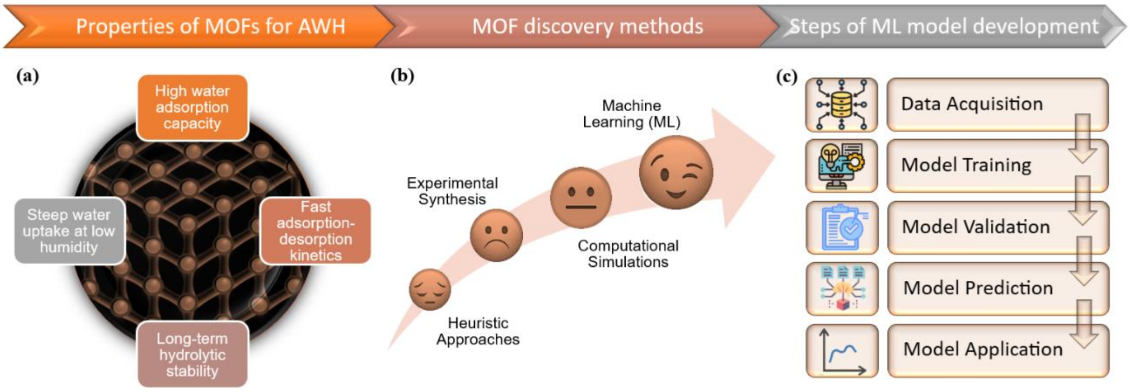


Fig. 3. (a) Key properties expected from MOFs for efficient AWH, including high water adsorption capacity, fast adsorption-desorption kinetics, steep water uptake at low humidity, and long-term hydrolytic stability. (b) Conventional and computational methods commonly used for MOF discovery and optimization, highlighting their complexities. (c) The five key steps in the machine learning workflow for MOF research: data acquisition, model training, model validation, model prediction, and model application for property estimation and accelerated materials discovery.

In recent years, Machine Learning (ML) has emerged as a powerful tool for accelerating the discovery and optimization of materials, including MOFs, for AWH applications [30]. ML, a subset of artificial intelligence (AI), has shown remarkable capabilities in predicting outcomes by identifying patterns within large datasets, typically within the framework of an algorithm [31, 32, 33]. As illustrated in Fig. 3c, the classical ML process consists of five key steps: (1) data acquisition, (2) model training, (3) model validation,

(4) model prediction, and (5) model application [34]. In the first step, data acquisition, raw datasets are collected, ensuring they are of high quality, sufficient size, diverse, and relevant to the problem. These data include the identification and investigation of the most suitable features (variables, attributes) that best reflect the adsorption characteristics of MOFs, so that the system can make accurate predictions [30]. The quality of features significantly affects the model's performance. Properly chosen features enhance predictive accuracy, while poorly selected ones may lead to misleading results [35]. Removing less relevant or highly correlated features improves the efficiency and accuracy of predictions, besides prevents model being overfitted [36]. The accuracy and comprehensiveness of these datasets directly impact on the reliability of the trained ML models [30]. For this reason, sometimes data preprocessing may be sufficient. The preprocessing of data includes data cleaning, data transformation, dimensionality reduction, data sampling, feature engineering, data augmentation, etc. The second step, model training, involves first splitting the dataset into training, validation, and testing (unseen) data subsets to ensure robust evaluation [30, 36]. It then focuses on selecting appropriate ML algorithms and optimizing them through hyperparameter tuning, cross-validation, and the use of ensemble methods such as bagging and boosting to further improve model performance. There are various ML algorithms such as linear, tree-based, probabilistic, ensemble, and neural network and different algorithms perform differently depending on the dataset [37]. The third step, model validation, ensures that the trained model generalizes well by evaluating its performance on the validation set. This step helps in identifying underfitting or overfitting issues and guides hyperparameter tuning to enhance robustness. In the fourth step, model prediction, the finalized model is applied to unseen test data or new MOF candidates to estimate their properties. This stage evaluates how well the model performs on data it has never encountered, simulating real-world use cases. In the final step, model application, the generalized ML model is used for targeted tasks to make predictions (i.e. predicting the adsorption performance of MOFs and accelerating material discovery) [30].

In this review, we first provide a detailed analysis of the key properties that an optimal MOF for AWH must exhibit, emphasizing the necessity of multifunctionality. An ideal MOF should not only demonstrate high water uptake and rapid adsorption/desorption kinetics but also maintain structural stability under cyclic operation while ensuring energy-efficient regeneration [12]. Given the vast and complex chemical design space of

MOFs, traditional trial-and-error approaches are inherently limited in discovering materials that simultaneously optimize these properties. Thus, we explore ML as the most powerful and efficient tool for accelerating MOF discovery [30]. We outline various ML methodologies ranging from supervised learning (SL) and unsupervised learning (UL) to reinforcement learning (RL) that have been leveraged to predict adsorption capacities, phase stability, and overall performance of MOFs in AWH applications. Furthermore, we discussed how ML can overcome fundamental challenges such as high-dimensional material space navigation, multi-objective optimization, and inverse design, enabling a paradigm shift in materials discovery. A critical aspect of this approach is data preparation, where we highlight the importance of feature engineering and dataset identification to enhance predictive accuracy and generalizability. In addition, we systematically review recent advancements in ML-driven MOF research for AWH, evaluating their methodologies, predictive power, and impact on accelerating the development of next-generation water-harvesting materials. Finally, we discuss the challenges, limitations, and future directions for the application of ML for the discovery of suitable MOFs in AWH technology.

2. MOFs for Atmospheric Water Harvesting

The goal of designing MOFs for AWH is to identify materials that achieve an optimal balance between high water uptake, fast adsorption-desorption kinetics, steep water uptake at low humidity, and long-term hydrolytic stability. Unlike conventional gas adsorption applications, where a single dominant property may suffice, AWH demands multifunctional MOFs that excel across multiple performance metrics simultaneously. This section focuses on the critical MOFs attributes required for efficient AWH and the key strategies employed to engineering their properties that meet these stringent criteria.

2.1. High water adsorption capacity

MOFs with ultra-high porosity and large pore volumes (PVs) are designed for gas adsorption, with high porosity providing more accessible adsorption sites for water molecules [12, 38]. For example, Cr-soc-MOF-1 demonstrates a water uptake of 1.95 gg^{-1} with a BET surface area of 4549 m^2g^{-1} and PV of 2.1 cm^3g^{-1} [39]. MIL-101(Cr) shows water uptake between 1 and 1.5 gg^{-1} with high stability, and a BET surface area of 3070 m^2g^{-1} [40]. Co_2Cl_2(BTDD) captures up to 0.97 gg^{-1} of water, with a BET surface area of 1912 m^2g^{-1} [41]. NU-1500-Cr exhibits a water uptake of 1.09 gg^{-1}, with a BET surface area of 3580 m^2g^{-1} and PV of 1.24 cm^3g^{-1} [42]. The newly synthesized Al-t,t-Ma MOF, an extended analogue of Al-Fum, exhibits a 115% higher PV, a 57.5% increase in water transfer, and stable hydrothermal performance [43]. This enhancement leads to a specific daily water production (SDWP) of 28 m^3t^{-1} at a low regeneration temperature of 55°C, making it a promising candidate for AWH. However, there is a trade-off between high porosity and stability, as capillary forces can lead to structural deformation during regeneration. Hydrophobic ligands or post-synthetic modifications can enhance stability. For instance, a water-stable csq-MOF (NU-1000-TFA) with trifluoroacetate capping achieved reversible water adsorption of 1.32 gg^{-1} [44]. Zr-MOFs, modified with hydrophobic paracyclophane (PCP) and trifluoroacetate, show improved water sorption properties. NU-602 captured 0.72 gg^{-1} of water at 70% relative humidity (RH), and NU-913-TFA captured 0.85 gg^{-1} [45].

Defects in MOFs can increase hydrophilicity and PV, improving water adsorption capacity. For example, hollow MIL-101, created via post-etching, showed increased water uptake from 1.64 to 1.99 cm^3g^{-1} and improved adsorption capacity [46]. Incorporating salts or polymers into MOF matrices further enhances water capacity. LiCl@MIL-101, with 51 wt% salt, achieved a sorption capacity of 1.9 gg^{-1}, while a PNIPAM@MIL-101 composite captured 440 wt% of water at 96% RH, compared to 120 wt% for pure MIL-101 [47]. Encapsulating ionic liquids (ILs) into MIL-101(Cr) MOFs significantly enhances water uptake (from 1.01 gg^{-1} to 1.45 gg^{-1}), improves adsorption/desorption kinetics by 48% and 55%, respectively, and increases water transfer (Δq) by 45%, leading to a higher SDWP from 38 to 56 m^3t^{-1} at 70°C regeneration temperature [48]. Also, flexible MOFs with gate-opening effects offer great potential for AWH with high capacity and selective adsorption behaviors [26].

2.2. Fast adsorption-desorption kinetics

The sorption kinetics of adsorbents depend on factors such as the size, thickness, packing density, and surface area (SA) of the material, as well as internal heat and mass transfer properties and environmental conditions (temperature, humidity, wind velocity, etc.) [49]. Water molecule transport within packed MOF adsorbents includes intraparticle and interparticle transport [50]. Intraparticle transport is influenced by the crystal diameter of the MOFs, with smaller crystals achieving faster kinetics due to a shorter diffusion path

[51]. Interparticle transport, influenced by the size and tortuosity of voids, impacts overall mass transport. Hyunho Kim et al. predicted how thickness (1 to 5 mm) and packing porosity (0.5-0.9) affect adsorption-desorption kinetics. At low porosities (~0.5 or less), Knudsen diffusion limits the adsorption kinetics, while at higher porosities, interparticle diffusion resistance becomes a factor [52]. Simulation calculations can predict the optimal packing density and thickness for fast water vapor diffusion and adsorption kinetics.

Since MOFs in the AWH process must release adsorbed water through a thermal process, their structural design cannot be optimized without accounting for heat transfer considerations. Heat and mass transfer are closely linked. Increasing the packing density benefits heat transfer but hinders interparticle vapor transport due to reduced porosity [51]. Adding thermal conductive materials to a thin adsorbent layer can reduce interfacial heat and mass transfer resistance. For example, metal foams incorporated with MIL-101 have been synthesized to improve heat transport under electric or magnetic field heating [53].

High-frequency water vapor sorption cycling can also improve AWH rates. MOF-801, when fluidized, achieved steady-state operation cycles with a high AWH rate of 0.33 $Lh^{-1}kgMOF^{-1}$ at 18% RH (40 cycles/day at 22°C) [54]. In solar-driven AWH, photothermal compositions enhance kinetics and multi-cycle performance.

The interaction between the adsorbent and water molecules affects the regeneration temperature. Van der Waals interactions have a bond strength of about 7.7 $kJmol^{-1}$, while hydrogen bonds have a strength of 9.7 to 29.0 $kJmol^{-1}$, facilitating faster water release. Covalent O-H bonds have higher bond strength (459.3 $kJmol^{-1}$), requiring more energy for breaking [49]. For mesoporous materials, overcoming capillary condensation requires extra energy, resulting in higher regeneration temperatures. Therefore, micropore filling based on hydrogen bonding is a promising approach for efficient water adsorption and desorption.

2.3. Steep water uptake at low humidity

MOFs designed for arid regions must exhibit high water uptake at low humidity (RH < 40%) as many of these areas face water scarcity. Some representative MOFs (e.g., MOF-801, MOF-303, MOF-74, MIL-101, Cr-soc-MOF-1, Co_2Cl_2(BTDD)) show limited water uptake (<1 gg^{-1}) at low humidity [39, 55, 56, 57, 58, 59, 60]. To improve water uptake, several strategies have been explored: (1) Hydrophilic microporous networks: Short organic linkers (fumaric acid, terephthalic acid, etc.) in Zr- or Al-based MOFs with pore sizes around 5-9 Å can create hydrophilic environments conducive to water adsorption via hydrogen bonding [61, 62, 55]. Research shows that a pore size of ~10 Å is optimal for fast water vapor transport in arid regions [63]. (2) Immobilizing adsorption sites: Li et al. reported nitrogen sites in Zr-adip MOFs enhance hydrophilicity, providing secondary adsorption sites and increase water uptake [64]. Song et al. showed that incorporating oxygenated and nitrogenated groups could improve water transport, but excessively large adsorption sites can increase diffusion resistance [63]. (3) Functional group modification: Introducing hydrophilic groups (e.g., -COOH, -OH, $-NH_2$, $-SO_2$) during MOF synthesis shifts the water adsorption inflection point (α) to lower relative pressures, enhancing uptake at low humidity. Hydrophobic groups, in contrast, shift the α to higher pressures. This approach was used in MIL-101-NH_2, UiO-66-NH_2, and CAU-10-NH_2, which showed improved water affinity compared to their parent MOFs [65, 62, 66, 67]. However, introducing functional groups may reduce PV and limit maximum adsorption capacity (q_{max}). (4) Heterocyclic atom incorporation: Substituting carbon atoms with heterocyclic atoms (e.g., O, N, S) in MOFs can increase hydrophilicity and PV, improving water uptake. For instance, replacing benzene rings with pyran- or furan-based linkers enhanced water adsorption in CAU-10-H [68]. MOF-303's pyrazole groups, according to DFT calculation results, serve as the strongest water adsorption sites, further demonstrating the role of heterocyclic atoms in improving AWH performance [69]. (5) Creating defects: Defects in MOFs, such as missing linkers or metal sites, provide unsaturated adsorption sites, improving water affinity. For example, defects in UiO-66 and MOF-801 shift the water adsorption α to lower pressures, enhancing water uptake [26, 55, 70]. Hollow MIL-101, prepared via acid etching, demonstrated superior water adsorption due to defect-induced hydrophilicity [46]. (6) Composite strategies: Composites, such as MOF-801@MIL-101 supraparticles, combine the properties of different MOFs to improve water uptake and kinetics at low humidity. The composite achieved a 225% and 390% higher water uptake at 8% RH compared to pure MOF-801 and MIL-101 [71]. Similarly, LiCl@MIL-101 composites showed enhanced water uptake and fast sorption kinetics [72]. A recent study has designed a water-harvester composite (MTC) made from MOF-303 and thiolated chitosan (TC) that exhibits excellent water uptake and desorption properties, with a rapid water uptake capacity of 0.135 gg^{-1} in 60 minutes at 12.5% RH. It shows ultrafast water desorption kinetics of 0.003 $gg^{-1}min^{-1}$ at 8.5% RH and high efficiency at low heat (~40°C), with water desorption and collection rates of 0.0195 and 0.0168 $gg^{-1}min^{-1}$, respectively [73].

These strategies highlight the significant role of structural modifications and composite approaches in enhancing MOFs' water adsorption capabilities for efficient AWH in arid regions.

2.4. Long-term hydrolytic stability

The hydrolytic stability of MOFs is crucial for their effectiveness in AWH [74, 75, 76]. The coordination bonds between metal clusters and organic linkers in MOFs are reversible, making them susceptible to degradation through hydrolysis or linker replacement when exposed to water. Additionally, H^+ or OH^- ions in acidic or alkaline solutions can replace the original linkers, disrupting the crystal structure. During vacuum drying, pressure or capillary forces can lead to structural collapse or phase changes. In contrast, methods like low-boiling solvent exchange or CO_2 supercritical drying can preserve MOF structural integrity, maintaining high porosity.

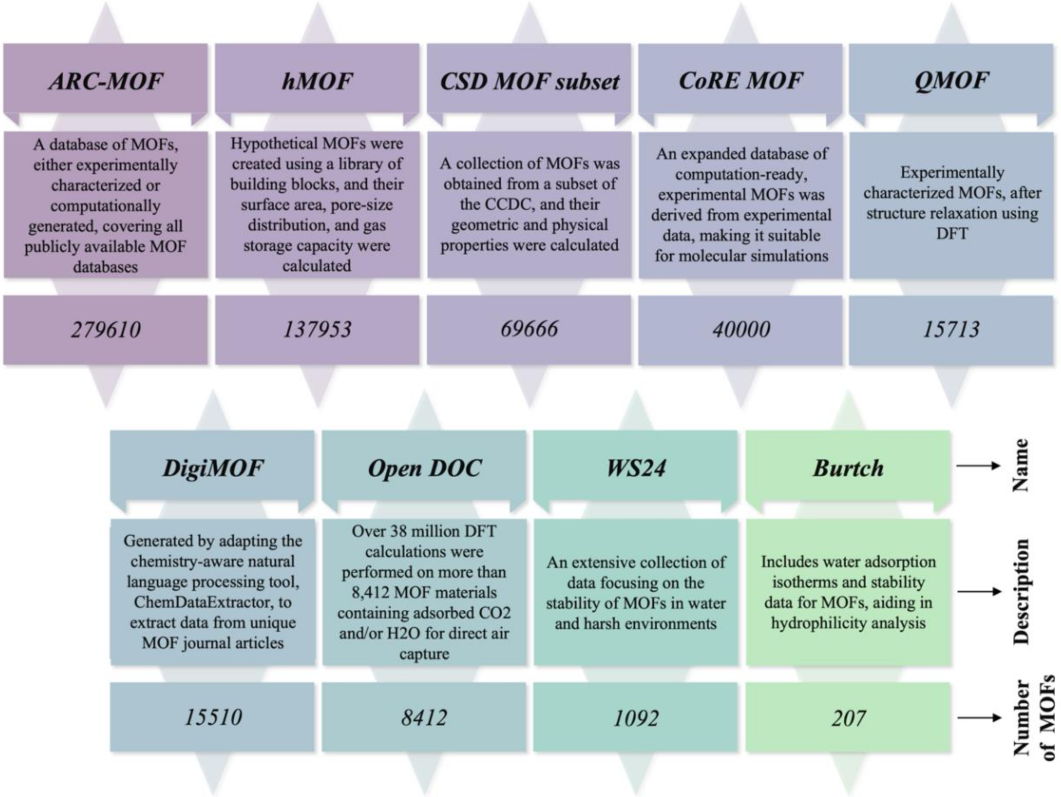
MOF stability is influenced by factors such as the choice of metal/organic ligands, metal-ligand coordination geometry, pore surface engineering, and steric effects [77]. According to the hard-soft acid-base concept, MOFs with high-valent metals (e.g., Zr^{4+}, Al^{3+}, Fe^{3+}, Cr^{3+}) coordinated with carboxylate groups and low-valent metals (e.g., Mg^{2+}, Co^{2+}) with azoles are generally more stable due to stronger coordination bonds [78]. MOFs with highly connected secondary building units (SBUs) and rigid linkers are also more resistant to degradation. For example, $Zr_6O_4(OH)_4(\text{-COO})_n$ or $Al(OH)(\text{-COO})_2$ structures help prevent water molecule attacks, improving stability. Short, rigid ligands such as fumarate can increase the activation energy required for degradation by allowing larger bending angles, thus enhancing stability [55, 79]. While steric effects can be strengthened by introducing bulky linkers or interpenetrating networks that prevent water attacks, this can also reduce PV and, in turn, water uptake capacity [26]. Therefore, this approach is not always suitable for improving stability in MOFs designed for AWH. Additionally, inert cations (e.g., Cr^{3+}) in octahedral coordination geometries, and hydrophobic functional groups (e.g., -F, -CH_3), can be incorporated into MOFs to improve their long-term stability for cyclic use in AWH devices [80]. For instance, Yang et al. synthesized the highly stable MOF material NU-1000-TFA, using trifluoroacetic acid (TFA) as a hydrophobic capping agent. Unlike NU-1000, which showed a 65% decrease in water uptake during the second cycle, NU-1000-TFA maintained a consistent uptake capacity of 1.32 gg^{-1} over cycles, demonstrating its stability [44]. Additionally, Laha et al. developed a ternary MOF nanocomposite consisting of aminoclay (AC), graphene oxide (GO), and CuBTC. This composite demonstrated a maximum water uptake of 0.67 gg^{-1} at 90% RH with high crystallinity, showing that AC and GO support improved the hydrothermal stability and water adsorption capacity of CuBTC [11]. Zhiyong et al. developed Zr-MOF (NJTech-4) with 12-connected Zr6 nodes exhibits a high-water vapor uptake of 995 cm^3g^{-1} and outstanding cycle stability, making it a strong candidate for AWH applications. As a result of this study, strategy of bidentate chelation of auxiliary ligands enhances node stability, improving both cycle and thermal stability against capillary forces during water desorption [81]. Junghye et al. reported that the porphyrin-based merged-net MOF, UPF-5, exhibits significantly enhanced hydrolytic stability and improved water sorption performance, maintaining a high PV. This design strategy, utilizing metallolinkers, enables the fine-tuning of Zr_6 nodes, optimizing water sorption properties and advancing reticular chemistry in MOF development [82].

These examples illustrate that MOF stability, particularly in terms of water and environmental stability, plays a significant role in enhancing their long-term performance for AWH applications.

3. MOF Databases and Descriptors

3.1. Unlocking Data: MOF Databases

Data quality and quantity play an essential role in the discovery and optimization of new materials. These data can be sourced from established databases, as shown in Fig. 4, which provides a comparative overview of major public MOF databases, categorizing them by their size and primary focus, or extracted from academic papers. The databases themselves encompass a wide array of information, derived from both experimental results and simulation data. With the constant progress in quantum chemical simulation methods and the increasing demand for ML applications, there has been a strong push to compile large datasets.



Name	Description	Number of MOFs
ARC-MOF	A database of MOFs, either experimentally characterized or computationally generated, covering all publicly available MOF databases	279610
hMOF	Hypothetical MOFs were created using a library of building blocks, and their surface area, pore-size distribution, and gas storage capacity were calculated	137953
CSD MOF subset	A collection of MOFs was obtained from a subset of the CCDC, and their geometric and physical properties were calculated	69666
CoRE MOF	An expanded database of computation-ready, experimental MOFs was derived from experimental data, making it suitable for molecular simulations	40000
QMOF	Experimentally characterized MOFs, after structure relaxation using DFT	15713
DigiMOF	Generated by adapting the chemistry-aware natural language processing tool, ChemDataExtractor, to extract data from unique MOF journal articles	15510
Open DOC	Over 38 million DFT calculations were performed on more than 8,412 MOF materials containing adsorbed CO2 and/or H2O for direct air capture	8412
WS24	An extensive collection of data focusing on the stability of MOFs in water and harsh environments	1092
Burtch	Includes water adsorption isotherms and stability data for MOFs, aiding in hydrophilicity analysis	207

Fig. 4. Overview of Public MOF Databases, including database identifiers and the number of MOFs available together with a brief description.

One of the pioneering efforts in this area was the creation of hMOF, which generated 137,953 hypothetical MOFs using a library of 102 building blocks [83]. Properties such as SA, pore-size distribution, and gas storage capacity were calculated for each structure using GCMC simulations. Another significant contribution came from Chung et al., who developed the CoRE MOF 2014 database, initially containing 4,764 MOF structures based on experimental data [84]. In 2019, this database, designed for immediate use in molecular simulations, has been expanded to include over 14,000 structures [85]. Recently, the CoRE MOF 2025 database was released as a comprehensive update to the previous versions [86]. It was presented with more than 40,000 experimental MOFs, addressing many of the previously reported limitations and incorporating new MOF structures published since CoRE MOF 2019. This new version improves data curation workflows and integrates machine-learning-predicted properties such as thermodynamic stability and heat capacity. It also includes updated MOFid annotations describing metal nodes, linkers, and topologies, and features DDEC6 partial atomic charges assigned via ML models. Furthermore, hydrophobicity classification was performed through Gibbs ensemble Monte Carlo simulations, making this dataset highly relevant for screening

MOFs under humid or water-rich conditions. This enhanced resource supports advanced ML applications by bridging high-quality structure data with predictive physical properties and process-level performance. Terrones et al. introduced the WS24 data set, a new database consisting of additional MOFs developed based on the CoRE MOF 2019 dataset [87]. Another notable contribution was made by Glasby et al., who created the DigiMOF database using the chemistry-aware natural language processing tool, ChemDataExtractor (CDE) [88]. By integrating web scraping tools with the Cambridge Structural Database (CSD) MOF set, they collected 43,281 unique MOF journal articles, extracted 15,501 unique MOF materials, and performed text mining (TM) on over 52,680 associated properties, including synthesis methods, solvents, organic linkers, metal precursors, and topologies. Moghadam et al. published a collection of 69,666 MOFs derived from a subset of the Cambridge Crystallographic Data Centre (CCDC) [89]. After solvent removal through the CSD Python API, properties such as surface area, PV, pore limiting diameter (PLD), largest cavity diameter (LCD), void fraction, and density were computed. Rosen et al. introduced the Quantum MOF (QMOF) database, which provides computed quantum-chemical properties such as electronic, optical, magnetic, and catalytic characteristics for 14,482 experimentally synthesized MOFs [90]. Similarly, Sriram et al. launched Open DAC 2023 (ODAC23), a dataset containing over 38 million DFT calculations on 8,412 MOF materials that adsorb CO_2 and/or H_2O for direct air capture applications [91]. Burtch et al. introduced the Burtch dataset, comprising 207 MOFs categorized by 22 metal nodes and 128 ligands [77]. Additionally, Burner et al. introduced the ab initio REPEAT charge MOF (ARC-MOF) database, which holds 279,610 MOFs, either experimentally characterized or computationally generated [92]. This comprehensive database spans across various public MOF resources and includes both experimental and hypothetical structures, ensuring a balanced representation of chemical space. Today, numerous public databases provide access to hundreds of thousands of MOFs. Understanding the historical development of these databases is crucial for advancing ML models. Selecting the right database is key to identifying MOFs with high water adsorption efficiency and stability. Nevertheless, challenges remain in addressing biases within existing databases and creating new ones that can overcome these limitations.

In the context of AWH, adsorption-desorption isotherms are indispensable for evaluating both the capacity and kinetics of water uptake, which are critical for selecting suitable

MOFs under varying humidity conditions. While structural and hypothetical MOF databases offer a solid foundation for screening materials, their predictive power is limited without incorporating experimentally validated adsorption properties. In this regard, the NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (NIST ISODB) serves as a valuable complementary resource [93]. This openly accessible database provides a large collection of both experimental and simulated adsorption isotherms for diverse adsorbent–adsorbate pairs, including water-MOF systems. Water isotherms extracted from NIST ISODB (alongside literature data) enable the derivation of key performance indicators such as Henry's coefficient (K_H), q_{max}, α, and working capacities in both low- and high-humidity regimes [94]. These metrics are essential for quantifying water–framework interactions, pore-filling behavior, and regeneration potential. Thus, incorporating property-based databases like NIST ISODB into ML workflows not only enhances the physical relevance of predictive models but also bridges the gap between theoretical screening and real-world performance assessment.

3.2. The Building Blocks: Descriptors

In the design and discovery of MOFs for AWH, descriptors play a crucial role in predicting the performance of these materials in water adsorption. In recent years, several advanced tools have been developed to automate the generation and validation of features, significantly improving the efficiency of computational studies on MOFs. One such tool is mofdscribe, a Python package designed to facilitate the extraction of both chemical and structural descriptors [95]. With over 40 built-in featurization methods, it allows direct computation of MOF descriptors, which can then be used to train ML models for property prediction. Another useful toolkit, molSimplify, provides a framework for extracting chemically relevant descriptors, including geometric properties such as metal-ligand bond lengths, ligand coordination angles, and coordination numbers [96]. Additionally, this tool incorporates electronic structure features derived from DFT calculations, such as highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), Fermi energy, and d-band center, all of which play a crucial role in understanding adsorption behavior. For a broader scope of feature generation, matminer (feature extraction library for materials science) offers an extensive library of featurization techniques, allowing users to extract descriptors from existing material databases [97]. This toolkit integrates seamlessly with ML frameworks like scikit-learn and Keras, enabling efficient model training and validation. It includes various featurizers, such as the ElementProperty featurizer, which calculates statistical properties of elements in a MOF, including average ionic radius and standard deviation of elemental melting points. Another key module, the BandFeaturizer, derives electronic descriptors from band structures, such as band gap values and k-point coordinates at conduction and valence band edges [98]. These tools have significantly enhanced the process of feature selection and predictive modeling in MOF research, enabling more accurate identification of materials with desirable adsorption properties. Generally, ML descriptors are divided into two main categories: explicit and latent. Explicit descriptors or traditional descriptors offer detailed insights into the material's structure and interactions with water. Beyond explicit descriptors, latent descriptors derived from computational models and ML techniques provide an alternative approach for predicting MOF performance, even in the absence of complete structural data. Fig. 5 shows types of ML descriptors schematically. By incorporating both types of descriptors, researchers can significantly enhance the accuracy and efficiency of MOF selection for AWH applications. In the following section, both types of descriptors and their various forms will be briefly elaborated, and a summary of the descriptors used in ML models for MOF-based AWH will be presented.

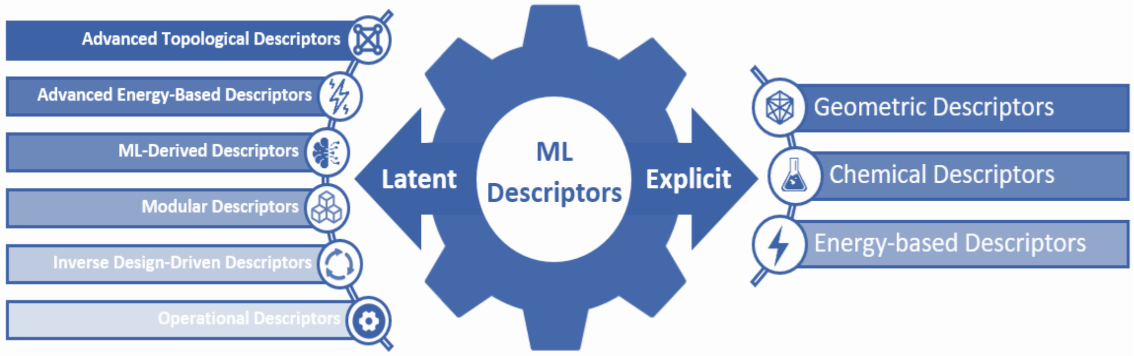


Fig. 5. Schematic illustration of the two main categories of ML descriptors used in MOF research: explicit descriptors and latent descriptors. Explicit descriptors, such as geometric, chemical and electronic-based features are directly calculated from MOF structures or simulations. Latent descriptors are learned representations extracted from data through ML algorithms often without requiring full structural information.

3.2.1. Explicit Descriptors

Explicit descriptors encompass geometric, chemical, and energy-based descriptors, providing a comprehensive representation of the material's behavior in water adsorption processes.

Geometrical descriptors primarily characterize the structural properties of MOFs and are computed using software tools such as Zeo++ [99], PoreSpy [100], VORO++ [101], and PoreBlazer [102]. Key descriptors include surface area, pore size, and porosity, which are critical for water adsorption, as they influence how water molecules access and interact with the framework. Unlike gas adsorption, where adsorption capacity typically increases with larger SA and porosity [103], water adsorption is more sensitive to pore size distribution and connectivity due to capillary condensation effects [12]. These geometrical descriptors, while useful, are inherently one-dimensional, meaning they are represented by a single numerical value, which may limit their accuracy in predicting water adsorption properties. A similar limitation applies to water adsorption, where the interaction between water molecules and the MOF structure depends not only on these geometric factors but also on their spatial arrangement. Examples of geometrical descriptors include LCD, probe-occupiable and accessible volume (POAV), PLD, largest free path diameter (LFPD), SA, volumetric surface area (VSA), volume (V_a), gravimetric surface area (GSA), void fraction (VF), PV, pore size, density (ρ), accessible surface area (ASA), pore dimensionality (PD), pore diameter (D_p), pore size standard deviation (PSSD), atom density (Metal atom, non-metal atom), and bond density [104, 105, 106]. Specifically, POAV represents the total volume accessible to the center of a water molecule, accounting for the material's extended volume [107]. ASA refers to the SA that can be reached by a water molecule, influencing initial water adsorption [107]. PSSD quantifies variations in pore sizes by calculating the weighted standard deviation of the pore size distribution, which affects the likelihood of capillary condensation [98]. LCD indicates the size of the largest spherical cavity that can fit inside the structure, influencing the maximum adsorbate size that can be accommodated [107]. PLD defines the smallest diameter along the pore channels, representing the size threshold for molecular diffusion [107]. LFPD refers to the longest uninterrupted path through the pore network, affecting transport properties and accessibility [108]. VSA measures the SA per unit volume, which influences volumetric adsorption capacity [109]. GSA is the SA normalized by the material's mass, indicating the potential for adsorption per gram of MOF [36]. VF represents the fraction of the material's total volume that is empty space, directly correlating with available pore space for adsorption [110]. PV quantifies the total pore volume, affecting the material's adsorption capacity [110]. Density (ρ) reflects the mass per unit volume, indirectly impacting adsorption performance by influencing material packing and pore accessibility [107]. Also, atom density and bond density describe the spatial distribution of atoms and bonds within the framework, which are crucial for understanding the structural integrity and adsorption behavior of MOFs [111]. Since these properties define the physical arrangement of atoms and their connectivity, they fall under geometrical descriptors rather than chemical ones.

Chemical descriptors provide numerical or categorical representations of MOFs' chemical properties, offering insights into their interactions with water molecules. The results of the studies showed that using geometric descriptors alone leads to relatively low prediction accuracy. However, when chemical descriptors are added, either alone or in combination with geometric descriptors, the prediction accuracy improves significantly. These descriptors typically analyze the composition of metals and functional groups within MOFs, which significantly influence water adsorption [112]. Unlike gas adsorption, where open-metal sites (OMS) primarily enhance adsorption via electrostatic interactions [107], water adsorption is more sensitive to factors such as hydrophilicity, hydrogen bonding, and framework stability in humid environments [12]. To further capture the local chemical interactions and coordination patterns that significantly influence water stability, descriptors are computed across different atomic scopes, such as metal-centered, linker-connecting atoms, and functional group-centered environments [113]. These atomic-level descriptors provide a more detailed representation of the electronic and coordination environment within MOFs, enabling a deeper understanding of how different structural components contribute to adsorption behavior. Examples of chemical descriptors include the (i) the metal set, which encompasses key parameters such as valency, atomic radius, ionization potential, electronegativity, atomic mass, valence electrons, radical electrons, hybridization, oxidation state, electron affinity, and metal radius [114, 115]. (ii) the linker set, which described using quantitative structure-property relationships (QSPR) and morphological data to characterize organic ligands [116]. (iii) the molar set, which focuses on the molar ratios between ligands and metal nodes, including species such as oxo, hydroxyl, and water molecules [117]. Furthermore, RDKit descriptors, which account for molecular weight, aromatic ring count, and other molecular features, provide additional characterization [118]. Molecular ACCess System (MACCS) fingerprints offer a molecular signature, capturing distinctive structural features, while revised autocorrelation functions (RACs) integrate data on metal nodes, organic linkers, and functional groups, thereby enhancing the overall understanding of the chemical system [119]. Among chemical descriptors, force field parameters such as

UFF4MOF capture atomic interactions, bond strengths, and coordination environments, providing a detailed representation of the MOF's chemical characteristics [120].

Energy-based descriptors provide an advanced layer of insight by quantifying adsorption energetics. During water adsorption in MOFs, host-guest interactions play a fundamental role in determining adsorption performance. While these features can offer higher predictive accuracy than purely structural descriptors, their computation often requires extensive sampling of the potential energy surface (PES), making them more computationally demanding [121]. Among the widely used energy-based descriptors, K_H and heat of adsorption (Q^{st}) are particularly significant, as they directly represent the interaction strength between water molecules and the MOF framework [122, 123]. Energy-based descriptors also include force field-derived parameters like UFF4MOF, which quantify intermolecular interactions and adsorption energetics, offering valuable insights into MOF stability and water affinity [120]. These features can be obtained experimentally or through computational methods with relatively low cost. Their predictive power is especially strong under low-pressure conditions, where host-guest interactions dominate adsorption behavior. To capture the PES more comprehensively, researchers have developed energy histogram-based features, which convert three-dimensional energy grids into lower-dimensional representations. One example is the one-dimensional energy histogram (1D-EH), which summarizes the energy interaction values across a unit cell into a histogram [124]. This method simplifies complex energy distributions while preserving key adsorption characteristics, making it computationally efficient and easily integrable into ML models. Expanding on this concept, two-dimensional energy histograms (2D-EH) incorporate energy gradients, allowing for better retention of spatial interaction details [125]. This refinement has been particularly useful for predicting adsorption properties of various adsorbates, including water, by improving the representation of energy variations within the MOF structure. Compared to full 3D energy grids, these histogram-based features reduce computational cost while maintaining predictive power. The continued development of advanced energy-based descriptors is enhancing the ability of ML models to accurately predict water adsorption behavior in MOFs, paving the way for more efficient material selection and design.

3.2.2. Latent Descriptors

While explicit descriptors provide detailed information about MOF structure and chemistry, latent descriptors offer an alternative approach to efficiently predict performance, even in cases where full structural data is unavailable, as they abstract complex material properties into computationally manageable representations, enhancing ML model generalization and predictive power. These advanced descriptors enhance the understanding of adsorption mechanisms by capturing complex structural and energetic interactions. As shown in Fig. 5, these features can be broadly classified into six categories: advanced topological descriptors, advanced energy-based descriptors, ML-derived descriptors, modular descriptors, inverse design-driven descriptors, and operational descriptors. Each of these approaches provides unique insights into adsorption behavior, expanding the scope of feature engineering beyond traditional methodologies.

Advanced topological descriptors provide a more detailed characterization of MOF structures by considering atomic connectivity, spatial arrangement, and geometric patterns. One key approach in this area is persistent homology, which transforms a MOF's structure into a point cloud and analyzes its connectivity through barcodes and persistent diagrams [126]. This technique has been used to identify structural voids, which are critical for adsorption behavior. Another method, geometric landscapes, represents MOFs in a high-dimensional space based on their geometric similarity [127]. By mapping these relationships, ML models can leverage topological patterns to predict water adsorption performance more effectively. An important tool used to analyze the topology of MOFs is MOFid [128]. This database specifically focuses on the structural connectivity and arrangement of metal nodes and linkers within the MOF framework. MOFid helps researchers classify and compare different MOF topologies, offering valuable insights into the material's network structure. This tool is particularly useful for designing and optimizing MOFs with specific topological features, improving their performance in adsorption applications, including water uptake.

Advanced energy-based descriptors go beyond explicit descriptors by offering a more detailed understanding of host-guest interactions, which is particularly valuable for predicting water adsorption behavior in MOFs. One emerging approach involves probabilistic distributions of electrostatic potential energy and surface curvature [129]. In this method, the accessible surface of a MOF is first determined based on Van der Waals energy. Key properties such as local mean curvature, electrostatic potential energy, and its gradient are then projected onto this surface. These values are subsequently transformed into normalized probability distributions, providing a quantitative representation of energy variations across the MOF structure. Such descriptors can enhance the prediction of water uptake under varying humidity conditions. Another effective feature is the radial distribution function (RDF) of energy adsorption sites, which helps assess the spatial distribution of adsorption energies on a MOF's accessible surface [130]. Adsorption energy values are categorized into different energy ranges, and RDFs are computed to determine how the density of adsorption sites varies as a function of distance from a reference point. By incorporating these descriptors into ML models, the spatial distribution of adsorption energy can be better understood, improving the accuracy of water adsorption predictions. A different strategy involves Effective Point Charge (EPoCh) descriptors, which estimate electrostatic interactions between adsorbent molecules and MOFs [131]. These descriptors are generated by simulating a hypothetical atom with varying charges and analyzing how these charge variations influence water uptake under different pressures. Unlike computationally expensive calculations such as K_H, EPoCh descriptors provide a cost-efficient way to capture electrostatic effects, making them highly useful for screening MOFs for water harvesting applications. Additionally, DFT-derived energy descriptors, such as adsorption energies and interatomic forces, provide a detailed description of the energy landscape, which is crucial for accurately modeling adsorbate-MOF interactions [132]. These descriptors are essential for training ML models to predict adsorption behavior and simulate framework flexibility.

ML-derived descriptors have emerged as a powerful approach for predicting water adsorption in MOFs by allowing models to self-learn features directly from structural data. Unlike manually designed descriptors, these self-learned features preserve spatial and energetic information, enabling models to identify complex adsorption patterns. However, the interpretability of these features remains a challenge, as it is often unclear which aspects of the structure contribute most to prediction accuracy. One major category of these descriptors is atomic position-based features, where MOF structures are represented as spatial grids or projected images for deep learning models. For instance, two-dimensional projections of MOFs have been used as inputs for convolutional neural networks (CNNs) to learn structural patterns that correlate with adsorption properties

[133]. Alternatively, voxel-based 3D grids, where occupied and unoccupied spaces are mapped, have been explored for predicting adsorption behavior [134]. While these methods have shown success in other adsorption contexts, their application to water adsorption remains challenging due to the complex nature of water-MOF interactions. To improve accuracy, energetically augmented atomic position-based features incorporate both structural and energetic information. These methods enhance pure position-based approaches by labeling atoms with energy-related properties such as electrostatic potential or Van der Waals interactions. For example, models trained with Lennard-Jones parameters or atomic point charges have demonstrated improved adsorption predictions, as they capture the energy landscape governing water uptake [135]. Another advanced approach involves crystal graph convolutional neural networks (CGCNNs), which model MOFs as graph structures where nodes represent atoms and edges encode bonding interactions [136]. By leveraging these learned representations, CGCNNs can predict water adsorption properties with high accuracy. Additionally, hybrid models that combine CGCNNs with traditional geometric descriptors, such as pore size and accessible volume, further enhance predictive performance, leveraging both data-driven learning and domain knowledge [137]. These ML-derived descriptors provide a scalable way to screen MOFs for water adsorption, offering a pathway toward discovering efficient materials for AWH applications.

Modular descriptors provide an efficient way to characterize MOFs for water adsorption by leveraging shared structural elements across different materials. Given the diversity of MOF structures, ranging from metal nodes to organic linkers and functional groups, these descriptors help capture key adsorption-relevant features without requiring extensive individual calculations for each material. One approach involves encoding the surface charge distribution of organic linkers as a modular feature, which can be applied across multiple MOFs. For instance, sigma profile descriptors, originally used for hydrocarbon adsorption [138], can be adapted to predict water adsorption by incorporating linker polarity and hydrophilicity. By integrating these descriptors with traditional structural descriptors such as pore size, ML models can achieve more accurate adsorption predictions. The advantage of modular descriptors lies in their transferability, enabling the construction of feature databases that eliminate the need for recalculating properties for each MOF. This approach is particularly beneficial for screening large MOF libraries for AWH applications, as it streamlines the identification of promising candidates while maintaining predictive accuracy.

Inverse design-driven descriptors have emerged as a powerful tool for developing predictive ML models for material properties. An encoder converts input data into meaningful representations, often as vectors or tensors, which allows for the inverse design of materials by targeting specific properties. The decoder then reconstructs the material from these representations. For example, Yao et al. introduced a method called supramolecular variational autoencoder (SmVAE), which decomposed MOFs into components such as metals, organic linkers, and topologies [139]. These were encoded into graph-based canonical codes (RFcodes), which were used as features to train a deep generative model. This model used the encoder to map RFcodes into a latent space, enabling property predictions. Afterward, the decoder reconstructed RFcodes, validating the representations and allowing for the inverse design of MOFs.

Operational descriptors encompass external conditions and experimental parameters that influence the adsorption behavior of MOFs but are not intrinsic to their structure or composition. These descriptors capture environmental and experimental factors such as temperature (T), pressure (P), RH, and adsorption kinetics, which are critical in determining the real-world performance of MOFs in applications like AWH [140]. Unlike explicit structural or chemical descriptors, operational descriptors account for system-dependent variations that can significantly impact adsorption isotherms and thermodynamic properties. Incorporating operational descriptors into ML models enhances predictive accuracy by adapting predictions to practical conditions rather than relying solely on idealized material properties [87]. As a result, these descriptors play a crucial role in bridging computational modeling with experimental validation, ensuring that MOF performance predictions align with real-world applications.

A comprehensive list of explicit and latent descriptors utilized in MOF-based AWH is presented in Table 1. A detailed discussion of each descriptor is provided in Section 6.

4. ML Models for MOF Discovery

ML has revolutionized the discovery and optimization of MOFs, enabling more efficient material design across various applications, including adsorption-based processes. While previous studies have explored the role of ML in screening MOFs for gas storage, separation, and general adsorption performance, AWH presents unique challenges that demand specific material properties, which were discussed in Section 2. In the context of AWH, ML can be particularly powerful in predicting water adsorption isotherms, identifying critical structural features influencing water capture efficiency, and streamlining the discovery of MOFs with tailored properties. Unlike traditional trial-and-error approaches, ML-driven methodologies allow for rapid screening of large MOF databases shown in Fig. 4 while incorporating experimental and computational insights to enhance prediction accuracy. Beyond structural screening, ML is instrumental in optimizing synthesis conditions and post-synthesis characterization. By analyzing historical synthesis data, ML models can predict the most effective parameters for obtaining MOFs with superior AWH performance, reducing experimental costs, and accelerating material development. Additionally, automated analysis of large-scale experimental datasets enables the identification of key stability and performance indicators, further refining the selection process. Given the diversity of ML approaches, no single algorithm universally outperforms others across all MOF-related tasks. Therefore, a comparative evaluation of ML models tailored to AWH applications is essential. This section systematically examines different ML methodologies applied to MOF discovery, emphasizing their role in addressing the specific challenges of AWH.

4.1. Classification of ML Models

ML approaches are difficult to classify because functionality, goals and methodologies often overlap, making it difficult to categorize them into clear and independent categories. Often algorithms serve multiple purposes or adapt to different learning paradigms based on the application domain, complicating categorization even further. Nevertheless, for a better understanding and conceptual clarity of the use of ML in MOFs discovery, it is necessary to establish a classification framework. Based on the type of learning, the methods in ML mainly categorized into SL, which predicts material properties using labeled data; UL, which uncovers patterns and groups data without labels; and RL, which optimizes material designs by exploring and interacting with design spaces autonomously [141]. These categories can be further branched, first by problem objective and then by algorithm type. Fig. 6 schematically illustrates the classification of ML approaches. The red oval shapes in this figure represent the three main types of learning (i.e., Supervised, Unsupervised, Reinforcement), which are considered the building blocks. These types are classified according to how data is interacted with and how the learning process is shaped. The brown arrowheads branching from them indicate specific approaches that are techniques or algorithms developed for specific types of problems. The oval shapes with stars at the bottom of the diagram represent hybrid models formed by combining the main types, which are needed where real-world problems often involve complex structures that cannot be adequately addressed by main learning types. The color coding explains this combination: purple indicates supervised + unsupervised, green indicates supervised + reinforcement, and gray indicates unsupervised + reinforcement learning combinations. While this categorization offers a systematic framework for comprehending ML methodologies, it is crucial to recognize that the domain of ML is a swiftly advancing area of study, with researchers frequently presenting novel methodologies and hybrid paradigms. As progress persists, new techniques frequently obscure the distinctions between established categories, underscoring the fluid and proliferating essence of ML inquiry.

As shown in Fig. 6, beyond traditional ML paradigms, various alternative approaches have been explored in MOF discovery to enhance prediction accuracy, optimize synthesis strategies, and accelerate material screening. One of them is hybrid learning approaches integrate elements of SL, UL, and RL. Among hybrids the most employed techniques are Hybrid Learning Models (HLMs) such as Semi-Supervised Learning (SSL) [142], Transfer Learning (TL) [143] and Machine Learning Potentials (MLPs) [144]. HLMs combine different ML techniques to leverage their respective strengths, improving generalization and robustness in MOF property prediction. For instance, SSL integrates labeled and unlabeled data to enhance model performance, while self-supervised learning extracts meaningful representations from large MOF datasets. TL is an ML paradigm that transfers knowledge learned in one task to another task. Korolev and Mitrofanov introduced an ML model combining density-derived electrostatic and chemical methods (DDEC) with charge equilibration methods (Q^{eq}), improving adsorption property predictions for nanoporous materials [145]. Wang et al. developed an integrated ML framework, incorporating both geometric and chemical features to link MOF structures with gas adsorption properties [146]. Ma et al. employed TL to enhance the predictive accuracy of ML models for gas adsorption in MOFs [147]. They trained DNN on H_2 adsorption data and successfully transferred knowledge to predict H_2 adsorption at different conditions and CH_4 adsorption, improving accuracy from 0.935 to 0.980.

Additionally, MLPs represent an emerging hybrid approach that combines SL with physics-informed modeling to accurately approximate PES for MOF systems. By learning from high-fidelity quantum mechanical data, MLPs enable efficient prediction of interatomic interactions, thereby accelerating MD simulations and materials screening with near-ab initio accuracy at a fraction of the computational cost [144].

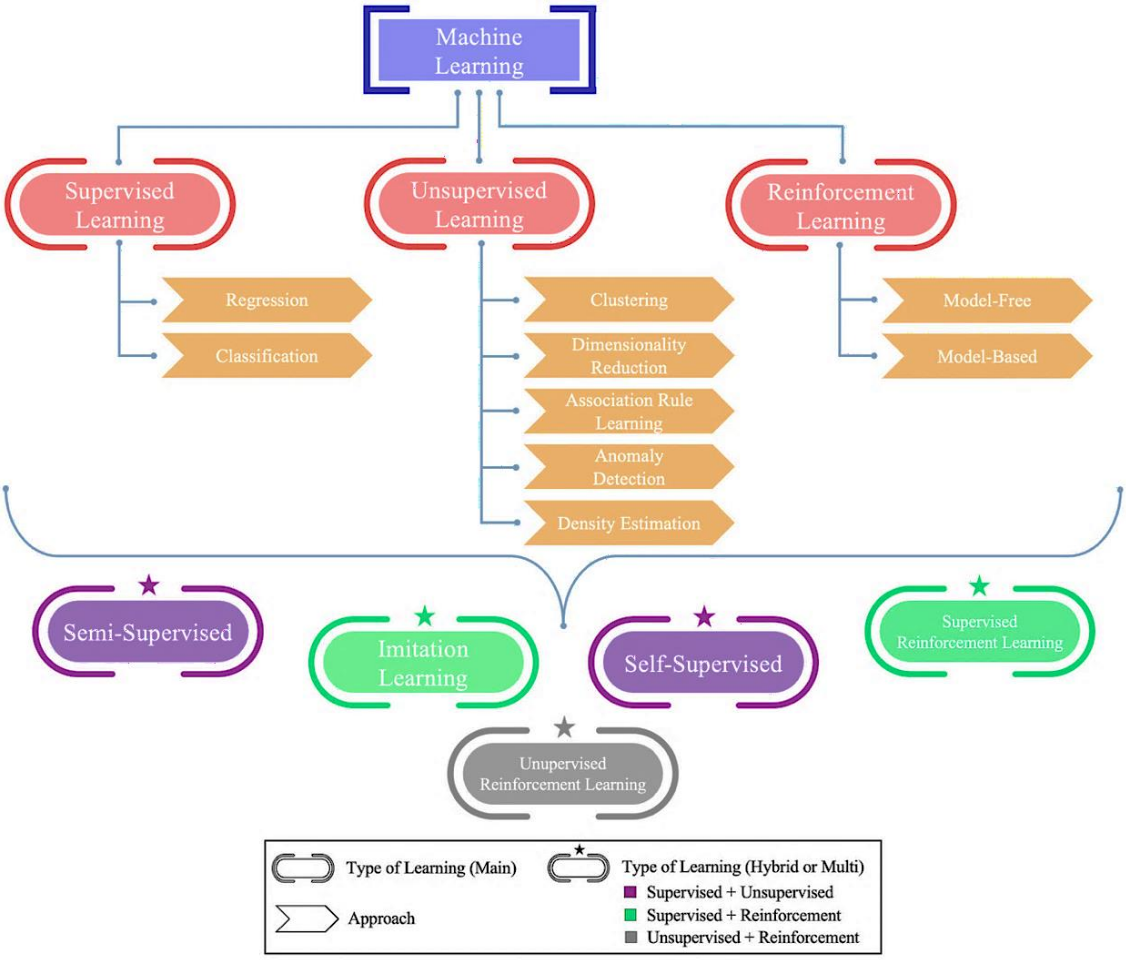


Fig. 6. Schematic overview of ML approaches applied in MOF discovery. The figure classifies ML techniques into three primary categories: SL, UL, and RL, based on the nature of the data and learning objectives. It also highlights the use of hybrid strategies and advanced emerging methodologies.

Besides hybrid learning approaches, some methods that integrate elements of ML. These methods include Social Network Analysis (SNA) [148], TM [149], Bayesian Optimization (BO) [150] and Evolutionary Algorithms (EAs), etc. Park et al. developed a text-mining algorithm to extract data from 46,701 MOFs and proposed a ML learning algorithm for synthesis prediction [151]. Qiao et al. optimized MOF separation performance using Decision Trees (DTs) and Back-Propagation Neural Networks

(BPNN) alongside a Particle Swarm Optimization (PSO) algorithm, identifying seven MOFs with superior xylene separation capabilities [152]. SNA analyzes the relationships and interactions between entities in a network to identify patterns, influences, and structure, often used to predict behaviors or properties based on these connections. Jalali et al. introduced MOFSocialNet, a SNA-based model that outperformed conventional ML tools in predicting MOF properties [153]. TL enables models to transfer knowledge from pre-trained datasets to MOF-related tasks, improving predictive performance with limited data availability. This technique is particularly valuable in deep learning models, such as Deep Neural Networks (DNNs) [154], CNNs [155], and Graph Neural Networks (GNNs) [156], where feature extraction from large chemical databases enhances the accuracy of MOF property predictions. BO is a powerful probabilistic technique used to optimize complex design spaces, such as MOF synthesis parameters. By iteratively refining predictions based on prior observations, BO efficiently identifies promising candidates with minimal computational cost. Long et al. utilized BO to enhance ML models for predicting the energy efficiency of MOFs in adsorption-driven osmotic heat engines [157]. By optimizing hyperparameters in regression and classification models, they identified key structural features influencing performance, achieving a prediction accuracy of 88.3% in classification and R^2 = 0.84 in regression. Their approach significantly accelerated the screening process compared to traditional GCMC simulations, enabling efficient identification of high-performance MOFs. EAs, including Genetic Algorithms (GA) [87] and PSO [158], simulate natural selection and swarm intelligence to optimize MOF structures. These algorithms explore large search spaces and identify optimal materials by evolving candidate solutions over multiple iterations. Pilz et al. applied a GA to optimize multiple synthesis parameters for HKUST-1 SURMOF thin films on functionalized Au substrates, improving efficiency in multi-parameter optimization [159]. By leveraging innovative ML techniques, researchers have expanded the scope of MOF discovery beyond conventional methods, integrating hybrid models, transfer learning, and optimization algorithms to enhance predictive accuracy and accelerate material screening.

4.2. ML Models Applied in MOF-Based AWH Systems

ML has significantly transformed the discovery and optimization of MOFs for diverse applications, including gas separation, catalysis, and adsorption-driven processes.

However, the design of MOFs for AWH presents unique challenges that distinguish it from other adsorption and catalytic applications. Unlike gas adsorption, where performance is primarily dictated by pore size, surface area, and functionalization, AWH efficiency is governed by a complex interplay of thermodynamic and kinetic factors such as water uptake capacity, adsorption-desorption hysteresis, and hydrolytic stability [12]. These properties are strongly influenced by framework flexibility, hydrophilicity, and dynamic host-guest interactions, making predictive modeling more intricate than in conventional gas adsorption or catalytic applications.

While previous studies have extensively explored the role of ML in screening MOFs for gas storage, separation, and general adsorption performance, the specific demands of AWH require a distinct approach. ML proves particularly powerful in predicting water adsorption isotherms, identifying key structural features influencing water capture efficiency, and streamlining the discovery of MOFs with tailored properties. Traditional trial-and-error methodologies are inherently slow and resource-intensive; in contrast, ML-driven approaches enable rapid screening of extensive MOF databases, as shown in Fig. 4, integrating both experimental and computational insights to enhance predictive accuracy. Beyond structural screening, ML is instrumental in optimizing synthesis conditions and post-synthesis characterization. By leveraging historical synthesis data, ML models can identify optimal processing parameters for producing MOFs with superior AWH performance, significantly reducing experimental costs and accelerating material development. Additionally, automated analysis of large-scale experimental datasets facilitates the identification of key stability and performance indicators, further refining the material selection process. Due to the complexities associated with AWH, ML models applied in this domain must be capable of capturing nonlinear dependencies between structural descriptors and adsorption behavior while maintaining high predictive accuracy. These models must integrate explicit descriptors such as pore geometry and chemical functionality with latent features that capture framework flexibility and water adsorption energetics. Additionally, these models should be capable of effectively working with limited data. As a result, combining these factors has significantly restricted the applicability of many commonly used ML approaches in this field. For instance, the scarcity of high-quality experimental adsorption data for MOFs under AWH conditions limits the applicability of deep learning approaches, which typically require large datasets for training. In other side, while UL techniques, have proven effective in material classification and feature extraction, their application in MOF discovery for AWH remains limited. The primary reason lies in the fundamental nature of AWH-specific MOF design, which requires precise predictive modeling of adsorption properties rather than merely identifying structural similarities. Unlike gas separation or catalytic applications, where clustering can help categorize, materials based on pore size or chemical functionality, AWH performance is highly dependent on complex thermodynamic and kinetic factors, such as water uptake capacity, adsorption-desorption hysteresis, and hydrolytic stability, which are challenging to infer from clustering alone.

Furthermore, clustering models such as K-Means [160] or Gaussian Mixture Models (GMMs) [161] rely on predefined distance metrics, which may not fully capture the nonlinear correlations between MOF structure and water adsorption behavior. Similarly, dimensionality reduction methods like Principal Component Analysis (PCA) [162] and t-Distributed Stochastic Neighbor Embedding (t-SNE) [163] focus on data visualization rather than direct property prediction, limiting their utility in optimizing MOFs for specific adsorption targets. Application-specific priorities such as accuracy, interpretability, computational cost, and data structure play a decisive role in the selection of machine learning algorithms. Each algorithm has different learning mechanisms. For example, while the SVM algorithm stands out with its high generalization capacity, ability to produce optimal solutions, and strong discriminatory power, it has limitations such as being unsuitable for large data sets, high algorithmic complexity, and memory requirements [164]. ANNs, on the other hand, are considered effective and efficient models for solving complex problems thanks to their powerful features such as self-learning, adaptability, fault tolerance, nonlinear structure, and parallel processing. However, there are still challenges to be resolved in model design, transparency, uncertainty, external data estimation, and error surface navigation [165]. In the context of the advantages and limitations of ML algorithms in MOF studies related to AWH, RF and backpropagation-based ANN models have been shown to provide successful results in the prediction of gas adsorption isotherms [166]. It has been observed that the RFR model overcomes the over-learning problem more effectively than the GBR [94]. Depending on the dataset, it is argued that while the GBRT and RF models are suitable for the prediction of CoRE-MOFs, they are not sufficient for hMOFs [167]. It has been reported that the GBR algorithm is more sensitive to the characteristics of the training set and that this may affect reliability when applied to new data [168]. Consequently, SL models, which leverage experimental adsorption data for direct property prediction, remain the dominant ML approach in AWH-driven MOF discovery. Although various ML techniques have been explored for MOF discovery, only a selected subset has been applied to AWH-related tasks, as listed in Table 1 and Fig. 7, with the top-performing models highlighted thereafter.

Given the diversity of ML approaches, no single algorithm universally outperforms others across all MOF-related tasks. Therefore, a comparative evaluation of ML models tailored to AWH applications is essential. Table 1 and Fig. 7 provide a summary of the specific ML methods applied in MOF-based water harvesting, emphasizing techniques that have shown superior performance in predicting adsorption properties and stability.

Table 1: ML Applications in the Discovery and Design of MOFs for AWH.

ML Algorithms (Best)	Explicit Descriptors	Latent Descriptors	Datasets	Applications	Ref.
DTs, SVM, GBR, RF (RF)	Metal, linker, and molar sets (chemical descriptor)	---	Burtch dataset (comprising 207 MOFs), 10 MOFs synthesized post-publication	Predict the hydrolytic stability	[169]
SVM, GP, MLP, RF, KN/KNN (RF)	LCD, PLD, LFPD, VSA, GSA, VF, PV, ρ, pore size, PV, VSA, LCD, atom density, bond density (geometrical descriptor) and metal radius, electronegativity, affinity, ionization, atomic mass, valence electron, radical electron, hybridization, oxidation state, RDKit, MACCS fingerprint (chemical descriptor)	---	Database of 1133 experimentally synthesized MOFs, ARC-MOF database (~≈280,000 MOFs)	Predict the hydrolytic stability	[170]
SVM, MLP, LR, RF, GA (RF and GA)	Pore size, SA, PV, VF, ASA (geometrical descriptor) and RACs (chemical descriptor)	T (operational descriptor)	WS24 dataset comprising 1092 experimentally MOFs, 10,000 MOFs from a design space of 485,650 hypothetical structures	Predict the hydrolytic stability	[87]
GBR, RFR, KRR (RFR)	LCD, PLD, LFPD, VSA, GSA, VF, PV, ρ, PD (geometrical descriptor) and UFF4MOF, linker chemistry (chemical descriptor)	MOFid (advanced topological descriptors) and T (operational descriptor)	285 experimentally characterized MOFs from the NIST-ISODB database and the literature articles, 8,000 computation-ready, experimental (CoRE) MOFs	Predict the water adsorption	[94]
RF, GBRT, NCA* (NCA)	LCD, PLD, VSA, VF, ρ (geometrical descriptor) and Q^{st} (energy-based descriptors)	---	6013 computation-ready experimental MOFs (CoRE-MOFs) and 137,953 hypothetical MOFs (hMOFs)	Predict the water adsorption	[167]
RF, ANN (RF)	SA, V_a, D_p (geometrical descriptor)	P (operational descriptor)	EWAID dataset	Predict the water adsorption	[171]
MACE	---	DFT (Advanced energy-based descriptors)	DFT-trained MOF structures via active learning	Predict the water adsorption	[172]

* The authors considered NCA as a standalone classifier. NCA is a metric learning method that does not directly perform classification but only improves the performance of other algorithms by transforming the data for better representation.

As demonstrated in Table 1 and Fig. 7, among the various ML models used to MOF-based AWH, a clear trend toward a few top-performing algorithms has developed. The discovery process often begins with a broad category of learning models and then narrows down to a collection of specific applicable algorithms. Comparison study demonstrates that models like Random Forest (RF), Random Forest Regression (RFR), Neighbor Component study (NCA), and Genetic Algorithms (GA) consistently have the greatest predictive accuracy. Additionally, the Machine-Learned Atomistic Cluster Expansion (MACE) algorithm (a kind of message passing neural network) has emerged as a promising deep learning approach capable of accurately predicting adsorption isotherms in flexible MOFs. Other models explored include linear based (Logistic Regression (LR)), kernel based (Kernel Ridge Regression (KRR)), tree-based (DT), nearest neighbor based (k-Nearest Neighbors (KNN)), neural networks based (Artificial Neural Network (ANN) and Multilayer Perceptron (MLP)), Support Vector Machine (SVM), Gradient Boosting Regression (GBR), Gradient-Boosted Regression Trees (GBRT), and Gaussian Process (GP). The results of the studies show a high proportion of SL models in the MOF-based AWH process. The continued refinement of these models, coupled with the expansion of experimental datasets, is expected to further enhance the predictive power of ML in this domain, ultimately accelerating the discovery of high-performance MOFs optimized for AWH. Section 6 systematically examines different ML methodologies applied to MOF discovery, emphasizing their role in addressing the specific challenges of AWH.

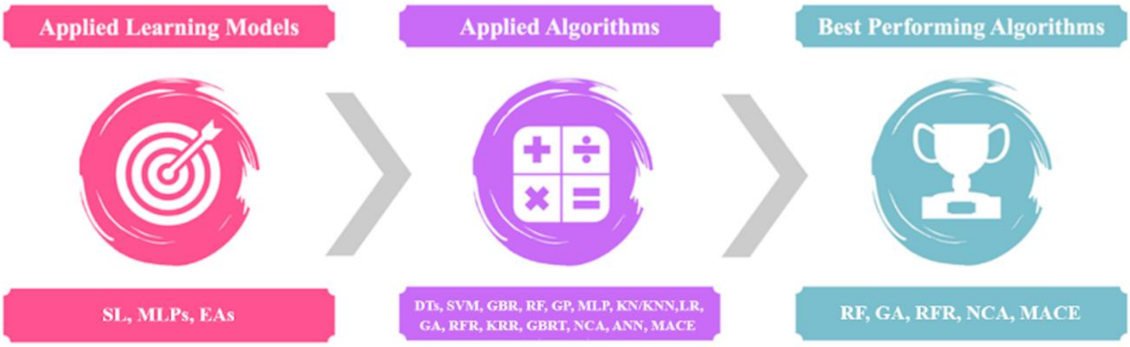


Fig. 7. Summary of ML models used in MOF-based AWH. The figure shows general categories of ML algorithms and highlights those with the best performance in predicting adsorption properties.

5. Performance Metrics

As indicated in Table 1, a range of ML algorithms have been utilized to predict adsorption properties and optimize MOF design for AWH applications. To evaluate the predictive power of these models, it is crucial to conduct thorough assessments using relevant performance metrics. This study specifically focuses on evaluating the performance of ML approaches based solely on the metrics identified in the literature, as shown in Table 2. In all metrics, y_i represents the actual values, $\hat{y}_i$ denotes the predicted values, n is the number of samples, $\bar{y}$ is the mean of the actual values, $Var(y)$ is the variance of actual values, and $\bar{\hat{y}}$ is the mean predicted values, TP is the true positive (the number of correctly predicted positive instances), FP is the false positive (the number of incorrectly predicted positive instances), FN is the false negative (the number of incorrectly predicted negative instances), TN is the true negative (the number of correctly predicted negative instances) and FPR is the false positive rate which is calculated by $FP/TN + FP.$

Table 2: Performance Metrics for Evaluating ML Models in MOF-Based AWH Systems.

Metric	Symbol	Equation	Ref
Coefficient of Determination	R²	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	[44, 94, 171]
Mean Absolute Error	MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	[94, 171]
Root Mean Square Error	RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	[171]
Accuracy*,**	ACC	$\frac{TP + TN}{TP + TN + FP + FN}$	[87, 169]
Positive Predictive Value/Precision**	PPV	$\frac{TP}{TP + FP}$	[87, 170]
True Positive Rate/Recall/Sensitivity**	TPR	$\frac{TP}{TP + FN}$	[170]
Area Under the ROC Curve**	AUC	$\int_0^1 TPR(FPR)dFPR$	[170, 87]
F1-score/F-measure**	F1	$\frac{2 * PPV * TPR}{TP + TN + FP + FN}$	[170, 87]

*We represent only the standard accuracy metrics, which are for binary classification. Various other accuracy metrics, such as balanced accuracy, overall (unweighted) accuracy and class-weighted accuracy, are commonly discussed in the corresponding articles; **In the case of multi-class classification, metrics are usually calculated by taking the average of the classes or using the total TP, FP, FN, TN values.

6. Overview of ML Applications in MOF-Based AWH Systems

ML techniques can efficiently analyze large chemical spaces and identify promising candidates for specific applications. This is especially important for MOF-based AWH, where key factors like water uptake capacity, hydrolytic stability, adsorption kinetics, and regenerability determine long-term efficiency. Recent studies have applied ML techniques to predict crucial MOF properties, including water adsorption behavior, resistance to hydrolysis, and structural stability under humid conditions. These advances not only improve our understanding of MOF-water interactions but also provide valuable insights for designing materials with enhanced durability and efficiency. Table 3 summarizes a selection of MOFs that have been discovered, optimized, or screened using ML approaches for AWH-related performance criteria, such as hydrolytic stability and water adsorption capacity. In the following subsections, we review the key ML-driven studies that led to the identification of these MOFs, highlighting their methodologies, predictive frameworks, and performance outcomes.

Table 3: MOFs Predicted by ML Models for AWH applications.

MOF Name	Main Properties	Ref.
$Ca(C_4O_4)(H_2O)$	Hydrolytic stability	[169]
AgTPB	Hydrolytic stability	[169]
$[Cd_{21}(TCA)_{14}(H_2O)_9]\cdot 21DMF\cdot 21H_2O$	Hydrolytic stability	[170]
$[Eu(pdc)(H_2O)_4]_3\cdot 3Cl$	Hydrolytic stability	[170]
N85 (Zr_6-cluster)	Hydrolytic stability	[87]
N95 (Keggin-type)	Hydrolytic stability	[87]
WAGQAY (Ni-cluster)	Water adsorption	[94]
EDUWAB (Zn-cluster)	Water adsorption	[94]
WAFKEU05 (Zn-cluster)	Water adsorption	[94]
QUTHAP	Water adsorption	[167]
Modified MOF-303	Water adsorption	[172]
Modified MOF-333	Water adsorption	[172]
Modified MOF-LA2-1	Water adsorption	[172]

6.1. Predicting MOF Hydrolytic Stabilities by ML

Evaluating the water stability of MOFs traditionally involves a time-consuming process of design, synthesis, and extensive experimental testing. This long development cycle hinders the rapid discovery of new water-stable MOFs. However, the rise of ML has opened new avenues for accelerating the design process of MOFs, including those that exhibit water stability. While conventional simulation tools like Material Studio, MATLAB, and Gaussian have been essential for MOF modeling, they come with certain limitations. These tools can sometimes lead to issues such as chemically invalid predicted ligands or erroneous MD and DFT simulations, resulting in unrealistic structures due to energy traps. To overcome these challenges, computational chemists have started leveraging ML models to predict and screen the water stability of MOFs. The fundamental difference between ML models and traditional methods is that ML models are trained using experimental data, allowing them to better align with the actual properties of MOFs. This data-driven approach, known as "real data training," enhances the accuracy and reliability of the predicted MOF properties. With the recent progress in the field, ML is becoming an increasingly powerful tool for the design of MOFs, offering the potential for a revolutionary shift in how stable materials are identified and developed. By streamlining the discovery process, ML can significantly reduce the time and resources required to find MOFs with desirable properties, such as water stability, which is crucial for MOFs used in AWH applications.

Batra et al. developed a ML framework to systematically predict the water stability of MOFs, addressing a fundamental challenge that limits their practical deployment in AWH applications [169]. They introduced the "Burtch dataset", comprising 207 MOFs categorized by 22 metal nodes and 128 ligands (Fig. 8a). The study utilized three sets of chemical descriptors: the metal set, which included information about metal nodes (such as valency, atomic radius, ionization potential, electronegativity, atomic triples and electron affinity); the linker set, which described organic ligands using quantitative QSPR and morphological data; and the molar set, which focused on the molar ratios of ligands and metal nodes, including oxo, hydroxyl, and water molecules. These descriptors enabled the unique numerical representation of each MOF, facilitating the model's ability to predict its water stability. To identify patterns within the dataset, the researchers employed various ML algorithms, including DTs, SVMs, GBR, and RF. Hyperparameter tuning and cross-validation were performed to enhance model performance. The ML models were trained using two test sets, while the Recursive Feature Elimination (RFE) was used to enhance model performance and eliminate redundant features. This feature selection method successfully reduced the feature dimension from 149 to 37 for the two-class model and to 29 for the three-class model, simplifying the learning process. The models trained on this reduced feature set were then used to classify MOFs into MOFs were classified into unstable (U), low kinetic (LK) stability, high kinetic (HK) stability, and stable (S) categories, achieving 83% accuracy for the two-class classification and 71% accuracy for the three-class classification. To evaluate the model's generalization capability, the researchers validated its predictions using 10 MOFs synthesized post-publication, obtaining 9 out of 10 correct predictions in the two-class model and 6 out of 10 correct predictions in the three-class model, demonstrating the effectiveness and reliability of their ML-based framework for predicting MOF water stability. Moreover, the model successfully identified $Ca(C_4O_4)(H_2O)$ and AgTPB (TPB: 1,2,4,5-tetra(4-pyridyl) benzene) as water-stable MOFs during screening of 88 new candidates, further demonstrating its practical applicability for discovering robust MOFs. However, comprehensive evaluation of both synthesis cost and adsorption performance is necessary to identify the most cost-effective candidates for AWH applications. Notably, the study does not address the water uptake capacity of MOFs, which remains a crucial parameter for evaluating their practical performance in AWH. The study revealed that MOFs exhibiting strong metal-ligand interactions, moderate pore sizes, and hydrophobic functional groups display enhanced water stability. Feature importance analysis further indicated that high charge density on metal centers and the presence of sterically hindered ligands significantly improved resistance to hydrolysis. Although the study considers water exposure conditions across different time scales (hours, days, and weeks), it does not explicitly assess the long-term hydrolytic stability of MOFs under continuous operational conditions relevant to real-world AWH systems. Among the tested ML models, the RF classifier demonstrated the highest predictive accuracy. Furthermore, the model's predictive capability was validated against experimentally measured stabilities of ten newly reported MOFs, achieving high agreement and showcasing its robustness and reliability. This approach provides a rapid and cost-effective alternative to experimental screening, significantly reducing the need for resource-intensive stability assessments. However, the study's reliance on a relatively small dataset (207 MOFs) may raise concerns about the generalizability of the model to more complex or diverse MOF systems, especially those with rare metal-linker combinations or underexplored topologies. Moreover, the use of solely hand-crafted descriptors, while interpretable, may limit the model's ability to capture higher-order structural and electronic interactions that contribute to hydrolytic stability. Incorporating graph-based or latent deep learning-derived features could potentially improve predictive performance and reveal hidden structure-property relationships. By integrating ML into MOF research, the study introduces an efficient predictive framework that minimizes experimental trial-and-error, guiding the rational design of robust MOFs. Furthermore, the findings emphasize that certain MOFs with high theoretical water uptake may still fail due to hydrolysis, underscoring the necessity of stability-aware screening in material selection. These insights are particularly valuable for the development of next-generation MOFs tailored for AWH applications, where long-term durability, recyclability, and hydrolytic resistance are essential. By leveraging ML-driven materials discovery, this study paves the way for more targeted and efficient development of MOFs optimized for real-world water harvesting applications.

Zhang et al. implemented a ML-driven approach to expedite the identification of water-stable MOFs, recognizing stability as a critical prerequisite for their effectiveness in AWH [170]. They constructed the largest database of 1133 experimentally synthesized MOFs with documented water stability and categorized them into four classes based on thermodynamic and kinetic stability: thermodynamically S, HK, LK, and U. To establish structure-stability relationships, MOFs were encoded using four levels of descriptors: global descriptors (LCD, PLD, LFPD, VSA, GSA, VF, PV, ρ, pore size, PV, VSA, LCD), building unit-based descriptors (metal radius, electronegativity, affinity, ionization, atomic mass, valence electron, radical electron, hybridization, oxidation state, RDKit descriptors (molecular weight, aromatic ring number, etc), MACCS fingerprint), Atom-based descriptors (Atom density (Metal atom, non-metal atom), Bond density), and other descriptors (RACs (Metal node, organic linker, functional groups)). The ML procedure involved extensive feature selection, data balancing techniques, and model optimization to develop two classification models (Classifier I and Classifier II), tailored for aqueous-phase and vapor-phase applications, respectively. Each ML model performance was optimized by tuning their respective hyperparameters. The receiver operating characteristic (ROC) was used as the primary metric for performance comparison. They reported that the combination of global and building unit-based descriptors has the best performance, while the RF classifier outperforms other algorithms (such as GP, SVM, MLP, and KN (KNN)) with the best ROC (Fig. 8a, b). To address the high dimensionality and potential redundancy of the initial descriptor set, the authors employed RFE. This systematic feature selection process allowed them to optimize the model, reducing the number of descriptors to 75 for Classifier I and 60 for Classifier II, thereby focusing only on the most impactful features. The most critical global descriptors influencing water stability included VSA, LCD, PV, and ρ. MOFs with high VSA (>3400 m^2cm^{-3}) and small LCD (<11 Å) exhibited superior water stability, as smaller pores hinder water cluster formation and prevent access to electrophilic sites (Fig. 8c). At the building unit level, the study found that hard-acid metals enhance MOF stability, as their high coordination numbers provide strong metal-ligand bonding, reducing hydrolysis susceptibility. Conversely, intermediate-acid metals exhibited diverse stability, depending on their d-electron configuration and ligand exchange rates. Among linker descriptors, aromaticity, steric bulk, and nitrogen-containing functional groups were positively correlated with stability, with imidazole-based linkers (as in ZIFs) exhibiting high resilience due to their strong coordination with metal centers. Applying the validated ML classifiers to the ARC-MOF database (≈280,000 MOFs), the study identified 129,661 water-stable candidates (≈47%) and, through multi-stability analysis, pinpointed 461 ultra-stable MOFs with concurrent water, thermal, and activation stability. Experimental validation of two predicted water-stable MOFs, $[Cd_{21}(TCA)_{14}(H_2O)_9]\cdot 21DMF\cdot 21H_2O$ and $[Eu(pdc)(H_2O)_4]_3\cdot 3Cl$, confirmed their stability under prolonged water exposure, reinforcing the ML model's reliability. However, considering the relatively high cost and limited abundance of Cd and Eu metals, the economic feasibility of using these MOFs in large-scale AWH applications remains a concern. Therefore, future research should focus on exploring more cost-effective MOFs with abundant metal clusters, balancing performance and economic viability through careful tradeoff analysis. The study highlights the transformative role of ML in accelerating the discovery of MOFs for AWH, offering a scalable and predictive framework for designing robust materials for water capture applications. Although the study benefits from a large dataset and diverse descriptors, its reliance on conventional models, like RF, limits the exploration of more advanced deep learning architectures that might better capture complex nonlinear relationships across feature spaces. While experimental validation was conducted on two MOFs, the generalizability of the model to more exotic or low-symmetry frameworks remains uncertain. Furthermore, the study overlooks the evaluation of water uptake capacity and the kinetic behavior of degradation or adsorption, both of which are crucial for understanding the material’s performance over time. Broader experimental studies, along with comparisons to generative or graph-based deep learning models and long-term cycling tests, would help strengthen both the robustness and real-world applicability of the results.

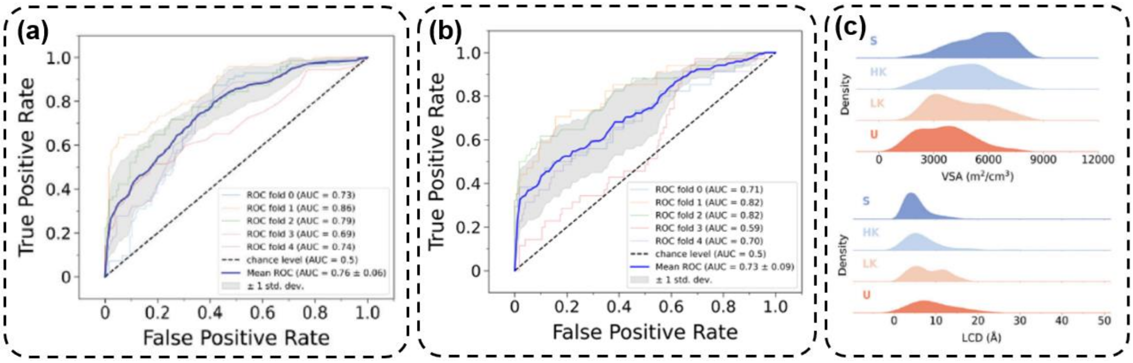


Fig. 8. (a) ROC curves of RF model based on original set in Classifier I, (b) ROC curves of RF model based on original set in Classifier I, (c) Kernel density estimations of water stability versus VSA and LCD [170]. Copyright 2024 WILEY.

Terrones et al. introduced an advanced ML methodology for classifying the water stability of MOFs, leveraging an expanded dataset and novel feature selection techniques to enhance predictive accuracy [87]. They constructed WS24, a dataset comprising 1092 experimentally characterized MOFs, significantly expanding previous datasets by over 400% through automated manuscript analysis. Each MOF was classified into one of four stability categories based on its resistance to water exposure: U, LK, HK, and S. This study included water stability and new labels for acid, base, and boiling stability, broadening the scope of the model's applicability. The authors proposed a systematic framework for MOF classification, emphasizing that an MOF reported as water-stable for only one day does not qualify for the TS class but can be categorized under the HK class. In this study, the authors employed a comprehensive set of descriptors to effectively capture the structural, geometric, and chemical characteristics of MOFs for ML-based water stability prediction. Geometric descriptors were calculated using Zeo++, encompassing parameters such as pore size, SA, PV, VF, and ASA, which collectively reflect the spatial architecture and porosity of the frameworks. To complement these macroscopic features, the authors utilized RAC descriptors, derived from both metal centers and organic linkers, to encode atomic-level information including atomic number, electronegativity, covalent radius, topological environment, and polarizability. These RAC descriptors integrate both product and difference-based metrics over various atom pairs, providing a nuanced representation of the chemical environment within the MOF structure. Additionally, descriptors were computed across different atomic scopes, such as metal-centered, linker-connecting atoms, and functional group-centered environments, thereby capturing the local chemical interactions and coordination patterns that significantly influence water stability. The use of hierarchical descriptor scopes is particularly noteworthy, as it enables the model to incorporate multi-resolution information that spans both local bonding environments and global framework geometry, an approach that remains underutilized in similar studies. By combining geometric features with RAC-based atomic descriptors, the study ensures a multi-scale representation of MOF properties, allowing the ML models to account for both physical framework attributes and intricate electronic or bonding effects that govern hydrolytic behavior. To develop a robust predictive model, they employed ML algorithms and compared multiple approaches, ultimately selecting RF method due to their superior performance over SVMs, MLPs, and LR. Two classification models were trained: a four-class model that predicted the exact stability category and a two-class model that grouped unstable and stable MOFs together to mitigate classification noise. For example, in two-class model MOFs are assigned to either the two more stable classes (HK+S) or the two less stable classes (U+LK)). The dataset exhibited class imbalance, which was addressed using synthetic minority over-sampling techniques (SMOTE) to improve model generalization. Feature selection was performed using recursive feature addition, allowing the model to retain only the most relevant descriptors while reducing complexity and overfitting. The optimized RF models achieved test set ROC-AUC scores exceeding 0.83, demonstrating their ability to accurately predict MOF stability across diverse chemical structures. The ML models identified several key structural descriptors governing MOF stability. Pore geometric features, including the largest cavity diameter, PV, and included sphere size, played a significant role in determining water uptake and hydrolytic resistance. Coordination environment descriptors, such as metal oxidation state and nuclear charge, were crucial, with higher oxidation state metals forming stronger bonds with organic linkers, thereby enhancing stability. Connectivity-based attributes, including graph-theoretic features like mc-χ-3 (metal-centered autocorrelation of electronegativity over 3-bond paths) and Dmc-χ-2 (difference-based metal-centered autocorrelation of electronegativity over 2-bond paths), provided additional insight into framework robustness, indicating that MOFs with densely connected metal centers exhibited greater water resistance. Interestingly, the study revealed that larger pores were positively correlated with acid stability, contradicting the conventional expectation that small pores provide greater protection against degradation. This observation, while significant, raises additional questions about the trade-offs between pore accessibility and chemical resilience, which the study does not fully explore. A more detailed mechanistic interpretation could clarify whether this trend is generalizable across MOF families or specific to structural subtypes. Furthermore, using the trained models, the study applied a GA to screen approximately 10,000 MOFs from a design space of 485,650 hypothetical structures, and identified 10 top-performing candidates that simultaneously exhibit high water, acid, and activation stability (composite fitness score > 0.85). Among these, the two highest-fitness hypothetical MOFs were highlighted, with detailed information on their unit cells, fitness scores, and compositional characteristics. These MOFs predominantly adopt the bcu and rtl topologies and feature highly connected SBUs, including N85 (a Zr_6 cluster derived from NU-1000) and N95 (a Keggin-type Zn-based unit found in NENU-501). While the integration of GA is an important innovation in this study, the potential limitations of the fitness function design and search space boundaries are not discussed. Including an analysis of how the GA's selection pressure may bias toward specific chemistries (e.g., Zr-based SBUs) would strengthen the generalizability of the conclusions. The findings underscore the power of ML in accelerating the discovery of stable MOFs, providing a scalable framework for screening large materials databases, optimizing experimental synthesis, and designing next-generation MOFs capable of efficient AWH. Despite offering a useful ML-based framework for predicting the water stability of MOFs, the study does not address several critical aspects required for practical AWH applications, including long-term stability under operational conditions, water uptake capacity, kinetic behavior, and experimental validation of predicted MOFs.

6.2. Predicting of water uptake on MOF by ML

While previous studies have explored the application of ML in predicting the hydrolytic stability of MOFs, an equally critical aspect of AWH performance lies in optimizing water adsorption properties. High water uptake, steep adsorption at low RH, and rapid adsorption-desorption kinetics, are key factors determining the efficiency of MOFs in AWH. ML has emerged as a powerful tool for accelerating the discovery and design of MOFs with these desirable attributes by efficiently screening large chemical spaces, identifying QSPR, and guiding the rational design of new materials. This section reviews recent advances in the use of ML techniques for predicting and optimizing MOF adsorption properties relevant to AWH. We analyze studies that leverage ML models to predict water uptake capacities, adsorption isotherms, pore characteristics, and functional group effects, aiming to uncover design principles for high-performance AWH materials.

By comparing these approaches with ML-based stability predictions, we highlight how data-driven strategies contribute to the comprehensive development of multifunctional MOFs for AWH.

Zhang et al. developed an ML-based framework to predict the water adsorption properties of MOFs and identify top-performing candidates for AWH [94]. Recognizing the limitations of traditional trial-and-error approaches in MOF discovery, the authors leveraged a data-driven methodology to systematically evaluate and optimize adsorption performance across a large dataset of materials. By leveraging both literature and the NIST ISODB database, they incorporated 344 water adsorption isotherms (198 from the NIST-ISODB database and 146 from the literature articles) from 285 experimentally characterized MOFs, enabling the extraction of key adsorption metrics, including isotherm type, K_H, α, q_{max}, and working capacities relevant for both arid ($\Delta q_{0.1\text{-}0.3}$) and humid ($\Delta q_{0.6\text{-}0.9}$) conditions. This approach highlights the value of curated isotherm databases in providing reliable input data for ML models and underscores the broader potential of resources like NIST ISODB for supporting reproducible, large-scale screening in AWH research. A high K_H indicates strong water–framework interactions, making such MOFs ideal for water harvesting in low-humidity (~20% RH) conditions. The MOFs were encoded into machine-readable descriptors, including eight structural features (LCD, PLD, LFPD, VSA, GSA, VF, PV, ρ) computed via Zeo++, 68 chemical descriptors quantifying atom types and densities using UFF4MOF field parameters, and operating conditions such as experimental temperature. Additional descriptors included PD (Zeo++ with a 1.35 Å probe), linker chemistry, and topology (MOFid), ensuring a comprehensive structure-property analysis. The integration of both microscopic (atomic-level) and mesoscopic (pore-scale) descriptors is commendable, as it enables the model to bridge local adsorption sites and global porosity, both of which are critical for water uptake performance. This level of modeling depth is often lacking in previous ML studies focused on MOFs. To develop these predictive models, the researchers employed multiple ML algorithms, including RFR, GBR, and KRR with both Gaussian and Laplacian kernels (KRR_G and KRR_L). Among these approaches, RFR demonstrated the highest accuracy, effectively capturing complex QSPR while minimizing overfitting. They featured importance analysis, based on Gini impurity, categorized key descriptors into structural (pore size, geometry), chemical (metal/nonmetal atom densities), and operational (temperature) factors. The numerical values in parentheses represent the relative importance of each feature, indicating its contribution to predictive accuracy within the ML model. They reported that K_H is primarily governed by chemical features, with metal (0.347) and nonmetal (0.283) densities playing dominant roles. Water molecules preferentially adsorb at unsaturated metal sites and hydrophilic groups, while temperature (0.102) negatively affects K_H due to the exothermic nature of adsorption. α is equally influenced by nonmetal density (0.351) and pore geometry (0.325), reflecting both local interactions and global framework effects. These findings offer a more nuanced understanding of how different design variables impact specific performance targets. This perspective not only enhances model interpretability, but also supports rational material design. However, the interpretation of trade-offs between K_H and α under practical operating conditions, such as in real AWH cycles, is not fully addressed. This limitation reduces the immediate translatability of the model outcomes to application-level optimization. They emphasized that chemical and structural features have opposite effects on K_H and α. q_{max} is structurally driven (0.515), with SA (VSA, GSA) and pore size (LFPD, PLD, LCD) exerting greater influence than PV. According to their reports, a pore size ~10 Å is identified as critical for q_{max}, while VSA > 4200 cm^2cm^{-3} leads to a sharp decrease. Adsorption uptake differences ($\Delta q_{0.1\text{-}0.3}$ and $\Delta q_{0.6\text{-}0.9}$) show a shift in dominance from chemical features at low pressure to structural factors at high pressure. Temperature has twice the importance for $\Delta q_{0.1\text{-}0.3}$ than $\Delta q_{0.6\text{-}0.9}$. To validate the robustness of their models, the authors conducted out-of-sample testing on newly synthesized MOFs, achieving strong agreement between predicted and experimental values. Although this validation strategy strengthens model credibility, the selection of validation candidates appears limited in diversity, which could restrict confidence in the generalizability of predictions across novel topologies and chemistries not well represented in the training dataset. Leveraging this validated model, they screened over 8,000 computation-ready, experimental (CoRE) MOFs to identify materials with superior AWH potential. Their analysis highlighted 149 MOFs with exceptionally high-water uptake ($q_{max} \geq 35$ $mmolg^{-1}$), 39 candidates optimized for low-humidity adsorption ($\Delta q_{0.1\text{-}0.3} \geq 10$ $mmolg^{-1}$), and 139 materials suitable for high-humidity conditions ($\Delta q_{0.6\text{-}0.9} \geq 8.7$ $mmolg^{-1}$). Among the CoRE MOFs screened, WAFKEU05 (Zn-cluster) showed the highest working capacity of $\Delta q_{0.1\text{-}0.3}$=13.82 $mmolg^{-1}$, and WAGQAY (Ni-cluster) achieved high working capacity ($\Delta q_{0.6\text{-}0.9}$=58.07 $mmolg^{-1}$). Four additional MOFs containing Zr, Cu, and Co clusters were also identified in the $\Delta q_{0.1\text{-}0.3}$ range; however, all exhibited lower uptake capacities compared to WAFKEU05. Considering that zinc is significantly more affordable and widely available than zirconium, copper, and cobalt, the Zn-based WAFKEU05 stands out as a cost-effective option that also offers the highest performance in this range. In the $\Delta q_{0.6\text{-}0.9}$ window, four MOFs with Zn and Cu clusters were identified. While WAGQAY, based on a Ni cluster, showed the highest working capacity, the Zn-based EDUWAB achieved a similarly high value of 56.06 mmolg^{-1}. Although slightly lower in uptake, the lower cost of zinc compared to nickel makes EDUWAB an economically attractive alternative. These results highlight the need to balance adsorption performance with material cost when selecting MOFs for practical AWH applications. This study establishes a scalable and data-driven roadmap for the design of next-generation MOFs, integrating structural, chemical, and adsorption descriptors to refine material selection criteria. However, one limitation not sufficiently addressed is the reliance on literature-derived isotherms, which may exhibit significant experimental inconsistencies between sources. Incorporating uncertainty quantification or model calibration against standardized benchmarks could further enhance the reliability of the predictions. By systematically uncovering the key determinants of water uptake and optimizing them through ML models, they demonstrate a powerful paradigm shift from conventional empirical screening toward predictive, ML-assisted material discovery. Their findings highlight the growing potential of ML in transforming AWH research, enabling the rapid identification of high-performance MOFs that balance adsorption capacity, stability, and regeneration efficiency. However, the study focuses on equilibrium adsorption properties but does not include kinetic analysis, which is important for evaluating real-time performance in AWH systems. Long-term stability is also not directly assessed, although K_H is used as an indirect indicator. Additionally, while model predictions are compared with experiments for seven MOFs, the limited diversity of these samples may affect generalizability. Future studies could benefit from including kinetic data, stability evaluation, and broader experimental validation.

Li et al. conducted a ML-assisted computational screening to identify high-performance MOFs for AWH [167]. This study combines Monte Carlo (MC) simulations with ML models to efficiently screen thousands of MOFs and determine their ability to selectively capture water from the atmosphere. The research focuses on 6013 computation-ready experimental MOFs (CoRE-MOFs) and 137,953 hypothetical MOFs (hMOFs), analyzing their adsorption performance based on five structural descriptors including the LCD, PLD, VSA, VF, ρ, and an energy descriptor of heat of adsorption (Q^{st}), were used to quantitatively describe the structure of the MOF. Among these, Q^{st} emerges as the most critical factor, as it directly influences the strength of water-MOF interactions. To establish predictive models, the study applies three ML algorithms: RF, GBRT, and NCA. The results are presented for the CoRE-MOFs database in Fig. 9 (a-c), and for the hMOFs database in Fig. 9 (d-f). After parameter optimization using five-fold cross-validation, NCA achieves the highest accuracy, with an R^2 of 0.97 for CoRE-MOFs and 0.86 for hMOFs, demonstrating its superior predictive capability. This highlights the efficiency of the NCA algorithm in capturing complex and diverse datasets, providing better prediction capacity and generalizability compared to conventional ML methods. Analysis of the structure-performance relationship reveals that MOFs with moderate porosity (VF between 0.2 and 0.62) and optimal pore-limiting diameters (~0.4 nm) exhibit the highest water selectivity. The study also identifies a strong correlation between Q^{st} and adsorption selectivity, confirming that higher Q^{st} values lead to superior performance in water capture. These findings, while confirming fundamental physical and chemical principles governing water adsorption, emphasize the critical role of energy interactions (Q^{st}) as the key performance factor, which has been quantitatively underexplored in many other studies. Using the trained NCA model, the authors screened the hMOF dataset and identified 10 high-performance MOFs, among which the best-performing structure was QUTHAP (i.e., a type of CoRE-MOF), exhibiting exceptionally high-water adsorption characteristics with a K_{H2O} of 2.78×10^{124} and a water selectivity of 4.14×10^{128}. Elevated K_H values reflect stronger affinities between water molecules and the MOF framework, which is crucial for effective water capture under arid conditions (~20% RH). By integrating ML with molecular simulations, this study demonstrates the potential of data-driven approaches in accelerating MOF discovery. The findings not only validate the effectiveness of computational screening for AWH applications but also provide a systematic framework for the experimental synthesis of next-generation MOFs optimized for water capture. Despite leveraging a large dataset and integrating ML with GCMC data to predict water uptake across diverse MOFs, the study does not fully address key practical concerns such as chemical stability, hydrolysis resistance, kinetic behavior, or long-term operational stability. These factors are essential for real-world AWH applications. Additionally, uncertainties regarding the diversity and representativeness of the hypothetical MOF dataset, along with the lack of experimental validation, raise questions about the model's generalizability. Nevertheless, the research presents a promising pathway toward the development of energy efficient and highly selective MOF based materials for improving water access in resource limited settings.

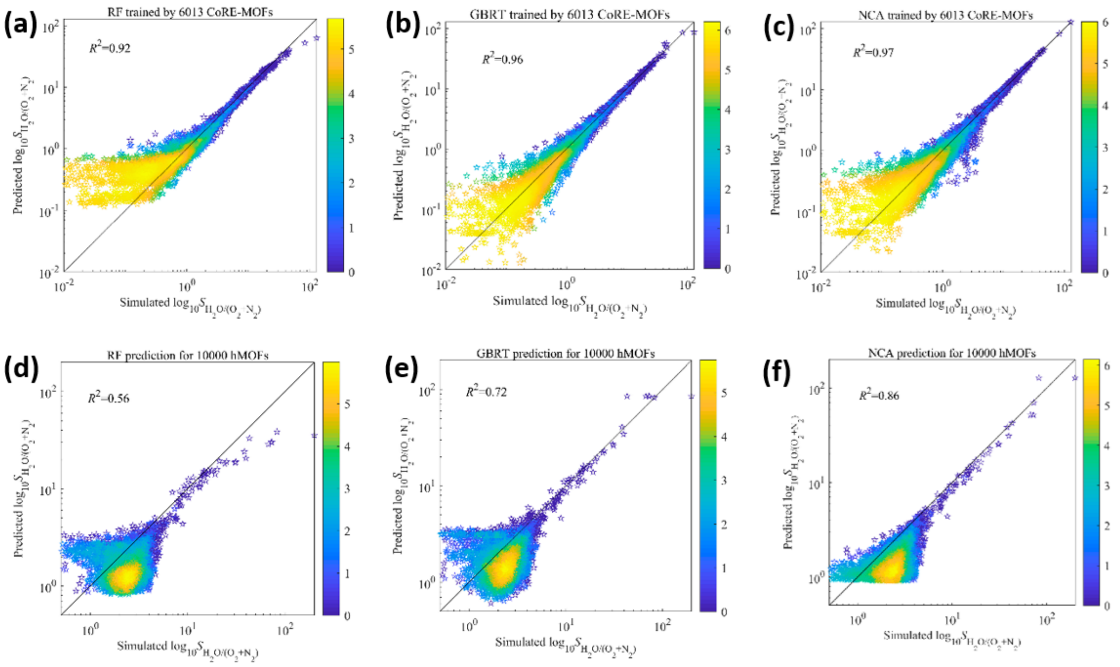


Fig. 9. Predictions of RF (a, d), GBRT (b, e), and NCA (c, f) models: training on CoRE-MOFs (a-c) and hMOFs (d-f). Color indicates the natural logarithm of the MOF counts [167]. Copyright 2022 MDPI (Open Access).

Liu et al. developed an ML-based approach to predict water adsorption isotherms, addressing an important challenge in the discovery of MOFs for AWH systems [171]. They utilized a two-step ML strategy to predict water adsorption isotherms and evaluate adsorption cooling performance. The study was based on an updated experimental water adsorption isotherm database (EWAID), which includes data on 460 nanoporous adsorbents, including MOFs, COFs, and zeolites. Descriptors like SA, V_a, D_p, and P were used for model training. Both models were implemented with Python's Scikit-learn module, optimized using grid search and cross-validation. In this study, the authors used RF and ANN models to predict water adsorption isotherms. The RF model, an ensemble of DTs, benefits from high accuracy and generalization ability. The ANN model uses backpropagation, mimics neural systems for nonlinear mapping and self-learning. Specifically, the RF model achieved an R^2 value of 0.982 with a low RMSE of 0.029 gg^{-1}, outperforming the ANN model (Fig. 10a). Feature importance analysis revealed that P had the highest impact on water uptake predictions, followed by SA, D_p, and V_a (Fig. 10b). The RF model's predicted water adsorption isotherms for various MOF adsorbents including different structural characteristics were compared to experimental data (Fig. 10c). These adsorbents exhibited four major isotherm types: type I (inverted "L" shape), type V ("S" shape), type IV, and type VI (multiple-step adsorption). As shown in Fig. 10c, the predicted curves closely matched the experimental results across all isotherm types and structural variations, highlighting the strong predictive accuracy and robustness of the RF model for water adsorption behavior. Notably, the model successfully captured variations in water adsorption behavior caused by differences in metal centers, functional groups, and framework structures. Beyond isotherm prediction, the authors applied their ML model to assess the adsorption cooling performance of adsorbent-water working pairs. They extracted key isotherm features, including q_{max}, α, and K_H, to predict specific cooling effects (SCE) and coefficient of performance for cooling (COPC). The RF model demonstrated high accuracy in predicting these performance metrics, with R^2 values exceeding 0.93. The results indicated that hydrophilic MOFs exhibiting type V (S-shaped) isotherms with high q_{max} were optimal for AWH and cooling applications. Despite these strengths, the study could benefit from a deeper discussion of potential limitations. For instance, the reliance on traditional descriptors such as SA and pore diameter may overlook more nuanced chemical or electronic factors affecting adsorption, such as metal-ligand interactions or hydrophilicity at the molecular level. Additionally, while the RF model's high accuracy is impressive, comparisons with other state-of-the-art ML architectures like GBR or GNNs could provide insights into model selection and robustness. The study also does not explicitly address the effect of framework stability under cycling conditions, a critical factor for practical AWH deployment. Overall, it demonstrated the powerful application of ML to replace costly and time-consuming experimental screening, facilitating accelerated discovery of MOFs with tailored water adsorption and cooling capabilities. The approach provides a scalable, data-driven roadmap for materials selection in AWH and related fields, marking an important step towards integrating predictive analytics with materials design.

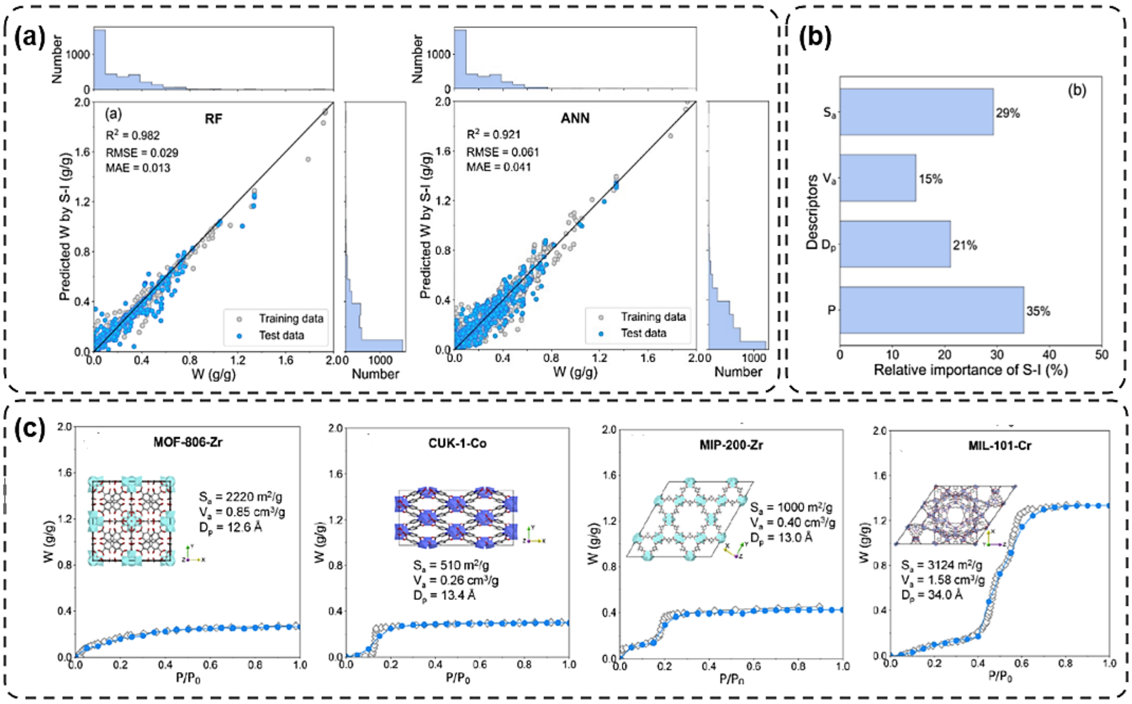


Fig. 10. (a) The predicted results of water uptake (W) by RF and ANN models, (b) The predicted results of the relative importance of descriptors for predicting water uptakes by RF, (c) The water adsorption isotherms of MOF-806-Zr, CUK-1-Co, MIP-200-Zr, and MIL-101-Cr in which the gray diamond markers represent experimental data form EWAID and the sky-blue dots represent data predicted by the RF model [171]. Copyright 2023 Royal Society of Chemistry.

Goeminne and Speybroeck developed an ML-driven methodology to predict water adsorption in flexible MOFs with chemical accuracy, addressing key limitations in conventional force-field-based simulations [172]. They emphasized that traditional GCMC simulations, which rely on classical force fields, fail to capture the complex host-guest interactions and loading-dependent deformations in MOFs, leading to inaccuracies in adsorption isotherms. To overcome this, they benchmarked the accuracy of various DFT methods by comparing water–MOF interaction energies and water cluster binding energies against high-level reference values. To accelerate the simulation of adsorption phenomena, they trained MLPs using a MACE framework based on an active learning strategy. These MLPs were trained on DFT-derived energies and forces across a wide range of water loadings and MOF configurations, with input descriptors implicitly including local atomic environments via equivariant message passing. The trained MLPs were subsequently used in Transition Matrix Monte Carlo (TMMC) simulations to accurately compute water adsorption isotherms. This allows for a comprehensive treatment of both local and global framework flexibility, an essential factor in optimizing MOF performance for AWH application. To train MLPs on selected DFT levels for water in the flexible MOF-303 framework without relying on computationally intensive DFT simulations, an active learning approach was used. Initially, an MLP was trained on a small dataset (860 structures) of guest-loaded MOF snapshots, with random cell strains and position displacements applied. Then, 5 iterations were performed, running 430 parallel MLP simulations from 430 different initial structures at 100 K to explore a small region of phase space. Low temperatures were used to avoid instability from sampling regions outside the phase space covered by the training set. The final frame from each simulation was extracted, recomputed at the reference DFT level, and added to the dataset for MLP retraining. This process was repeated with two more sets of 5 iterations: first at 300 K and then at 500 K (Fig. 11a). Through this active learning loop, simulations were performed using intermediate MLPs, with only 7310 configurations evaluated by DFT. They reported that directly running these simulations with DFT would require 32 million single-point calculations, far beyond the available computational resources. Still, the reliance on a reduced dataset raises questions about the model's applicability to more diverse MOFs or more extreme thermodynamic conditions not covered in the training set. By applying this methodology to MOF-303, a widely studied material for AWH, they achieved quantitative agreement with experimental adsorption isotherms, something not previously possible with classical force-field approaches (Fig. 11b). They showed that although adsorption isotherms vary with the level of theory, the structural features of adsorbed water remain consistent. They found that the number of hydrogen bonds between water and the framework, as well as between water molecules, showed similar trends across all theoretical methods. This indicates that even less accurate energy predictions can still provide reliable structural insights, such as interaction sites and hydrogen bonding behavior (Fig. 11c). Their simulations revealed that capturing both local distortions in adsorption pockets and global unit cell flexibility is essential for accurately describing water uptake behavior. Extending their approach to MOF-333 and MOF-LA2-1, they further demonstrated how modifications in organic linkers influence adsorption properties, providing a pathway for rational MOF design. To gain further insight into the effect of the organic linker choice on adsorption, a comparison was made among the three studied MOFs. At low relative pressures, an S-shaped adsorption isotherm was observed due to the hydrophilic PZDC linker (1-H-pyrazole-3,5-dicarboxylate) of MOF-303, whereas this effect disappeared when the less hydrophilic FDC (2,4-furandicarboxylate) linker was used in MOF-333. In this MOF, a very low uptake was recorded before a sharp increase around 20% RH. In MOF-LA2-1, a small uptake of 1 to 2 molecules per asymmetric unit cell was detected before a similar jump occurred at an RH of 26%, resembling the behavior of MOF-303.

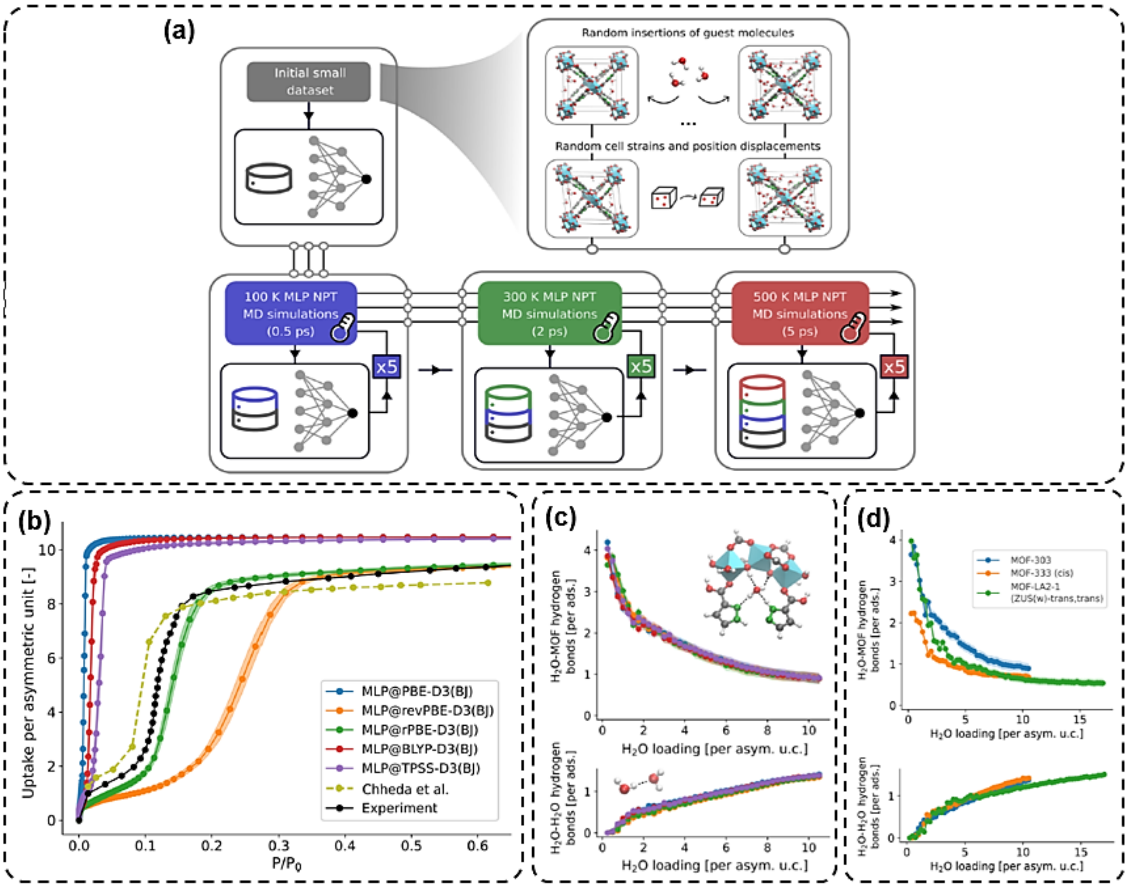


Fig. 11. (a) Summary of the Active Learning Workflow. The process begins with an initial dataset consisting of guest-loaded MOF snapshots subjected to cell and position displacements. This is followed by three rounds of active learning iterations, utilizing MLP simulations at 100, 300, and 500 K. Each round includes five iterative steps, (b) Adsorption isotherms for MOF-303 at 298 K, computed via MLP-based TMMC simulations using MLPs trained at five theoretical levels, are compared with experimental data and force-field-based Gibbs ensemble Monte Carlo simulations, (c) The computed number of MOF-H_2O hydrogen bonds and (c) H_2O-H_2O hydrogen bonds per adsorbate as a function of the water loading, (d) number of MOF-H_2O hydrogen bonds and H_2O-H_2O hydrogen bonds per adsorbate computed by DFT for the three MOFs studied [172]. Copyright 2025 ACS Publications (Open Access).

For applications, a single steep step in uptake, without significant adsorption at lower RH, as seen in MOF-333, was preferred. To explain this behavior, the number of hydrogen bonds per molecule as a function of loading was presented in Fig. 11d for all three MOFs. The study presents a powerful computational framework by combining machine learning, first-principles calculations, and Monte Carlo simulations, enabling accurate prediction of water adsorption behavior through synergy between quantum-level modeling, structural flexibility, and efficient sampling. However, kinetic performance and long-term structural stability, which are essential for practical AWH applications, were not explicitly addressed. The validation relied mainly on GCMC data, with limited comparison to experimental uptake and durability results. Including a deeper analysis of how structural features such as linker hydrophilicity translate to real-world performance, including cyclic stability and adsorption-desorption behavior, would significantly improve the practical relevance of the findings.

7. Discussion, Conclusions and Future Outlook

By leveraging data-driven models, researchers have been able to rapidly screen MOF candidates, predict adsorption properties, and optimize structural parameters with higher efficiency than conventional trial-and-error approaches. However, despite these advancements, several critical challenges and research gaps remain. While ML has significantly enhanced MOF-based AWH research, most studies have relied on SL models such as RF, GBR, NCA, RFR, GA, and MACE. Although these models have shown strong predictive performance, their accuracy and generalizability are often limited by the quality and diversity of the training datasets. Additionally, deep learning approaches, which have demonstrated success in other material science applications, remain underutilized in AWH due to the scarcity of large, high-quality datasets. Future research should explore the integration of semi-supervised learning, TL, and RL to enhance model robustness and adaptability, particularly in data-limited scenarios.

The studies about ML applications on MOF-based AWH reveal that the most practical descriptors and metrics for AWH tasks vary depending on the objective. Within the category of explicit descriptors, energy-based properties such as heat of absorption and geometric properties such as surface area are prominent in predicting water adsorption capacity, while other explicit properties, such as pore diameters and atomic characteristics, play a critical role in predicting the hydrolytic stability of the structure. Complementing these implicit descriptors, such as a model's inherent confidence scores or learned design rules derived from interpretability techniques, offer a powerful advantage. These derived properties allow researchers to go beyond mere prediction and identify the most promising candidates, leading to accelerated materials discovery by uncovering rational design principles.

When evaluating model performance, R^2 for regression tasks, and Accuracy, ROC Curve – AUC for classification, have emerged as the most reliable and widely used metrics in the literature. The most common and important metric that indicates how well a model's water retention predictions match the actual values is R^2 (for regression tasks). In the literature, almost all regression studies report high R^2 values. It is noteworthy that ROC and AUC (for classification tasks) are the primary metrics used to compare hydraulic stability prediction models. This is because AUC provides a more reliable measure of model performance than other metrics, even when the classes in the dataset are unbalanced (e.g., when the number of stable MOFs is much lower than the number of unstable ones).

A key factor influencing the success of ML models in MOF-based AWH is the selection and engineering of descriptors. While explicit descriptors such as pore size, surface area, and functional group chemistry have been extensively utilized, the incorporation of latent descriptors remains relatively unexplored. Latent descriptors derived from graph-based representations, topology analysis, and deep learning embeddings can provide a more comprehensive understanding of complex MOF-water interactions. Further research is required to develop advanced feature extraction techniques that can capture framework flexibility, water adsorption energetics, and dynamic host-guest interactions more accurately.

To further enhance the predictive accuracy and applicability of ML in MOF-based AWH, several future research directions are proposed: (1) Expansion of high-quality experimental datasets to improve ML model generalization and reliability by integrating computational and experimental data; (2) Integration of multi-scale simulations, such as DFT, MD, and GCMC, with ML models to enhance descriptor quality and predictive power; (3) Improved interpretability of ML models using techniques such as SHAP (SHapley Additive exPlanations), Gini-impurity, etc. analysis and feature importance ranking to enhance model transparency and guide experimental validation, and perform robust feature selection by identifying and removing redundant descriptors; (4) Exploration of innovative ML approaches, such as RL, BO, and GNN, to enhance MOF discovery for AWH, and (5) Real-world testing and commercialization considerations, focusing on scalability, synthesis feasibility, long-term durability, and cost-effectiveness to ensure the successful translation of ML-discovered MOFs into practical AWH technologies.

In all studies related to ML applications on MOF-based AWH, while transferability has generally been tested through out-of-sample validation using temporally distinct datasets, a systematic cross-validation between distinct chemical families (e.g., leaving out all MOFs with a specific metal center or functional group) has not been explicitly performed. Therefore, the development of robust validation strategies that explicitly test for chemical diversity, such as leave-one-group-out cross-validation in larger and more varied datasets, is a critical direction for future research to ensure the development of truly universal predictive models.

As part of this future perspective, it is also essential to acknowledge the current study's limitations. These include the dependence on limited and often heterogeneous datasets, the underutilization of deep learning models due to data scarcity, and the relatively early stage of integrating ML with experimental workflows. Moreover, although various ML algorithms have been explored, systematic comparisons across models, descriptors, and datasets remain limited. Addressing these gaps would significantly enhance the field's progress.

In conclusion, while ML has significantly advanced MOF-based AWH research, there is a pressing need for more diverse and robust modeling approaches, particularly those that leverage latent descriptors, multi-scale simulations, and generative design techniques. A greater focus on experimental validation, as well as consideration of industrial-scale challenges, will be key to transitioning from theoretical design to practical application. By addressing these challenges and embracing emerging computational methodologies, the field can accelerate the discovery of next-generation MOFs with superior water harvesting performance. The continued integration of ML with experimental validation will be pivotal in transforming AWH into a scalable and commercially viable solution for global water scarcity challenges.

Acknowledgements

This work was supported by the European Union under the Marie Skłodowska-Curie Actions (MSCA), HORIZON-MSCA-2023-PF-01-01, through the PHOTOWAT project (Grant Agreement No. 101154984, DOI: 10.3030/101154984), coordinated by University Complutense of Madrid (UCM), Spain. N. Hilal would like to thank Tamkeen for funding NYUAD Water Research Center (project CG007).

References

[1] F. Dalstein y A. Naqvi, «21st Century water withdrawal decoupling: A pathway to a more water-wise world?,» *Water Resources and Economics,* vol. 38, p. 100197, 2022.

[2] M. M. Mekonnen y A. Y. Hoekstra, «Four billion people facing severe water scarcity,» *Science Advances,* vol. 2, p. e1500323, 2016.

[3] S. Almasoudi y B. Jamoussi, «Desalination technologies and their environmental impacts: A review,» *Sustainable Chemistry One World,* vol. 1, p. 100002, 2024.

[4] M. Arjmandi, M. P. Chenar, A. Altaee, A. Arjmandi, M. Peyravi, M. Jahanshahi y E. Binaeian, «Caspian seawater desalination and whey concentration through forward osmosis (FO)-reverse osmosis (RO) and FO-FO-RO hybrid systems: Experimental and theoretical study,» *Journal of Water Process Engineering,* vol. 37, p. 101492, 2020.

[5] A. Campisano, D. Butler, S. Ward, M. J. Burns, E. Friedler, K. DeBusk, L. N. Fisher-Jeffes, E. Ghisi, A. Rahman, H. Furumai y M. Han, «Urban rainwater harvesting systems: Research, implementation and future perspectives,» *Water Research,* vol. 115, pp. 195-209, 2017.

[6] X. Liu, D. Beysens y T. Bourouina, «Water Harvesting from Air: Current Passive Approaches and Outlook,» *ACS Materials Letters,* vol. 4, p. 1003–1024, 2022.

[7] H. A. Zabidi, H. W. Goh, C. K. Chang, N. W. Chan y N. A. Zakaria, «A Review of Roof and Pond Rainwater Harvesting Systems for Water Security: The Design, Performance and Way Forward,» *Water,* vol. 12, p. 3163, 2020.

[8] W. Xu y O. Yaghi, «Metal–Organic Frameworks for Water Harvesting from Air, Anywhere, Anytime,» *ACS Central Science,* vol. 6, p. 1348–1354, 2020.

[9] H. Chen, T. Ran, Y. Gan, J. Zhou, Y. Zhang, L. Zhang, D. Zhang y L. Jiang, «Ultrafast water harvesting and transport in hierarchical microchannels,» *Nature Materials,* vol. 17, p. 935–942, 2018.

[10] M. Zhou, H. Song, X. Xu, A. Shahsafi, Y. Qu, Z. Xia, Z. Ma, M. Kats, J. Zhu, B. Ooi, Q. Gan y Z. Yu, «Vapor condensation with daytime radiative cooling,» *Applied Physical Sciences,* vol. 118, p. e2019292118, 2021.

[11] S. Laha y T. K. Maji, «Binary/Ternary MOF Nanocomposites for Multi-Environment Indoor Atmospheric Water Harvesting,» *Advanced Functional Materials,* vol. 19, p. 2203093, 2022.

[12] Y. Hu, Z. Ye y X. Peng, «Metal-organic frameworks for solar-driven atmosphere water harvesting,» *Chemical Engineering Journal,* vol. 452, p. 139656, 2023.

[13] M. Ejeian y R. Wang, «Adsorption-based atmospheric water harvesting,» *Joule,* vol. 5, pp. 1678-1703, 2021.

[14] D. Wang, J. Zhang, X. Tian, D. Liu y K. Sumathy, «Progress in silica gel–water adsorption refrigeration technology,» *Renewable and Sustainable Energy Reviews,* vol. 30, pp. 85-104, 2014.

[15] u. Hu, Z. Fang, X. Ma, X. Wan, S. Wang, S. Fan, Z. Ye y X. Peng, «CaCl2 Nanocrystals decorated photothermal Fe-ferrocene MOFs hollow microspheres for atmospheric water harvesting,» *Applied Materials Today,* vol. 23, p. 101076, 2021.

[16] Z. Zhang, H. Fu, Z. Li, J. Huang, Z. Xu, Y. Lai, X. Qian y S. Zhang, «Hydrogel materials for sustainable water resources harvesting & treatment: Synthesis, mechanism and applications,» *Chemical Engineering Journal,* vol. 439, p. 135756, 2022.

[17] M. Arjmandi, M. P. Chenar, M. Peyravi y M. Jahanshahi, «Influence of As-Formed Metal-Oxide in Non-Activated Water-Unstable Organometallic Framework Pores as Hydrolysis Delay Agent: Interplay Between Experiments and DFT Modeling,» *Journal of Inorganic and Organometallic Polymers and Material,* vol. 29, p. 178–191, 2019.

[18] E. Binaeian, N. Motaghedi, S. Maleki y M. Arjmandi, «Ibuprofen uptake through dimethyl ethylenediamine modified MOF: optimization of the adsorption process by response surface methodology technique,» *Journal of Dispersion Science and Technology,* vol. 43, pp. 1-14, 2020.

[19] M. Arjmandi y M. Pakizeh, «Effects of washing and drying on crystal structure and pore size distribution (PSD) of Zn4O13C24H12 framework (IRMOF-1),» *Acta Metallurgica Sinica (English Letters),* vol. 26, pp. 597-601, 2013.

[20] I. Ahmed, G. Lee, H. J. Lee y S. H. Jhung, «Adsorption of pharmaceuticals from water using metal-organic frameworks (MOFs), MOF-derived carbons, covalent-organic frameworks (COFs), COF-derived carbons: Comparison of the four adsorbents,» *Chemical Engineering Journal,* vol. 488, p. 151022, 2024.

[21] H. Wang, M. Wang, Y. Wang, J. Wang, X. Men, Z. Zhang y V. Singh, «Synergistic effects of COF and GO on high flux oil/water separation performance of superhydrophobic composites,» *Separation and Purification Technology,* vol. 276, p. 119268, 2021.

[22] H. L. Nguyen, «Covalent Organic Frameworks for Atmospheric Water Harvesting,» *Advanced Materials,* vol. 35, p. 2300018, 2023.

[23] K.-M. Kim, H.-T. Oh, S.-J. Lim, K. Ho, Y. Park y C.-H. Lee, «Adsorption Equilibria of Water Vapor on Zeolite 3A, Zeolite 13X, and Dealuminated Y Zeolite,» *Journal of Chemical & Engineering Data,* vol. 61, pp. 1547-1554, 2016.

[24] H. Lu, W. Shi, Y. Guo, W. Guan, C. Lei y G. Yu, «Materials Engineering for Atmospheric Water Harvesting: Progress and Perspectives,» *Advanced Materials,* vol. 34, p. 2110079, 2022.

[25] Y. Li, J. Yu, Y. Li, J. Shen, M. Du, X. Zhang, H. Zhao y J. H. Pu, «Nanoporous MOF-303 Performance for Atmospheric Water Harvesting in the Presence of Airborne Contaminants: GCMC and DFT Simulations,» *ACS Applied Nano Materials,* vol. 7, p. 23850–23859, 2024.

[26] M. J. Kalmutzki, C. S. Diercks y O. M. Yaghi, «Metal–Organic Frameworks for Water Harvesting from Air,» *Advanced Materials,* vol. 30, p. 1704304, 2018.

[27] I. I. El-Sharkawy, S. Haridy, M. Hassan, A. Radwan y M. M. Abd-Elhady, «Optimization of atmospheric water harvesting cycles for sustainable water supply in arid regions,» *International Journal of Thermofluids,* vol. 24, p. 100977, 2024.

[28] F. Luo, T. Liao, X. Liang, W. Chen, S. Wang, X. Gao, Z. Zhang y Y. Fang, «Two-linker MOFs-based glass fiber paper monolithic adsorbent for atmospheric water harvesting in arid climates,» *Journal of Cleaner Production,* vol. 373, p. 133838, 2022.

[29] S. F. Rafie, N. Abu-Zahra y R. Sabetvand, «From thin air to pure water: Enhancing atmospheric water harvesting with molecular dynamics simulations of pristine and porous zirconium-based MOF-801,» *Journal of Molecular Liquids,* vol. 422, p. 126939, 2025.

[30] C. Li, L. Bao, Y. Ji, Z. Tian, M. Cui, Y. Shi, Z. Zhao y X. Wang, «Combining machine learning and metal–organic frameworks research: Novel modeling, performance prediction, and materials discovery,» *Coordination Chemistry Reviews,* vol. 514, p. 215888, 2024.

[31] E. Aytaç, J. Contreras-Martínez y M. Khayet, «Mathematical and computational modeling of membrane distillation technology: A data-driven review,» *International Journal of Thermofluids,* vol. 21, p. 100567, 2024.

[32] E. Aytaç, «Object Detection and Regression Based Visible Spectrophotometric Analysis: A Demonstration Using Methylene Blue Solution,» *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal ,* vol. 12, p. e29120, 2023.

[33] E. Aytaç y M. Khayet, «Visual Footprint of Separation Through Membrane Distillation on YouTube,» *Data,* vol. 10, p. 24, 2025.

[34] T. Schröder y M. Schulz, «Monitoring machine learning models: a categorization of challenges and methods,» *Data Science and Management,* vol. 5, pp. 105-116, 2022.

[35] I.-T. Sung, Y.-H. Cheng, C.-M. Hsieh y L.-C. Lin, «Machine Learning for Gas Adsorption in Metal–Organic Frameworks: A Review on Predictive Descriptors,» *Industrial & Engineering Chemistry Research,* vol. 64, p. 1859–1875, 2025.

[36] I. B. Orhan, Y. Zhao, R. Babarao, A. W. Thornton y T. C. Le, «Machine Learning Descriptors for CO2 Capture Materials,» *Molecules,* vol. 30, p. 650, 2025.

[37] V. Sheth, U. Tripath y A. Sharma, «A Comparative Analysis of Machine Learning Algorithms for Classification Purpose,» *Procedia Computer Science,* vol. 215, pp. 422-431, 2022.

[38] Z. Chen, P. Li, R. Anderson, X. Wang, X. Zhang, L. Robison, L. R. Redfern, S. Moribe, T. Islamoglu, D. A. Gómez-Gualdrón, T. Yildirim, J. F. Stoddart y O. K. Farha, «Balancing volumetric and gravimetric uptake in highly porous materials for clean energy,» *Science,* vol. 368, pp. 297-303, 2020.

[39] S. M. T. Abtab, D. Alezi, P. M. Bhatt, A. Shkurenko, Y. Belmabkhout, H. Aggarwal, Ł. J. Weseliński, N. Alsadun, U. Samin, M. N. Hedhili y M. Eddaoudi, «Reticular Chemistry in Action: A Hydrolytically Stable MOF Capturing Twice Its Weight in Adsorbed Water,» *Chem,* vol. 4, pp. 94-105, 2018.

[40] N. Ko, P. G. Choi, J. Hong, M. Yeo, S. Sung, K. E. Cordova, H. J. Park, J. K. Yang y J. Kim, «Tailoring the water adsorption properties of MIL-101 metal–organic frameworks by partial functionalization,» *About Journal of Materials Chemistry A,* vol. 3, pp. 2057-2064, 2015.

[41] A. J. Rieth, S. Yang, E. N. Wang y M. Dincă, «Record Atmospheric Fresh Water Capture and Heat Transfer with a Material Operating at the Water Uptake Reversibility Limit,» *ACS Central Science,* vol. 3, p. 668–672, 2017.

[42] Z. Chen, P. Li, X. Zhang, P. Li, M. C. Wasson, T. Islamoglu, J. F. Stoddart y O. K. Farha, «Reticular Access to Highly Porous acs-MOFs with Rigid Trigonal Prismatic Linkers for Water Sorption,» *Journal of the American Chemical Society,* vol. 141, p. 2900–2905, 2019.

[43] B. Han y A. Chakraborty, «Ligand extension of aluminum fumarate metal-organic framework in transferring higher water for adsorption desalination,» *Desalination,* vol. 592, p. 118135, 2014.

[44] L. Yang, K. B. Idrees, Z. Chen, J. Knapp, Y. Chen, X. Wang, R. Cao, X. Zhang, H. Xing, T. Islamoglu y O. K. Farha, «Nanoporous Water-Stable Zr-Based Metal–Organic Frameworks for Water Adsorption,» *ACS Applied Nano Materials,* vol. 4, p. 4346–4350, 2021.

[45] W. Gong, H. Xie, K. B. Idrees, F. A. Son, Z. Chen, F. Sha, Y. C. Yan Liu y O. K. Farha, «Water Sorption Evolution Enabled by Reticular Construction of Zirconium Metal–Organic Frameworks Based on a Unique [2.2]Paracyclophane Scaffold,» *Journal of the American Chemical Society,* vol. 144, p. 1826–1834, 2022.

[46] Y. Hu, Z. Fang, X. Wan, X. Ma, S. Wang, S. Fan, M. Dong, Z. Ye y X. Peng, «Carbon nanotubes decorated hollow metal–organic frameworks for efficient solar-driven atmospheric water harvesting,» *Chemical Engineering Journal,* vol. 430, p. 133086, 2022.

[47] A. Karmakar, P. G. M. Mileo, I. Bok, S. B. Peh, J. Zhang, H. Yuan, G. Maurin y D. Zhao, «Thermo-Responsive MOF/Polymer Composites for Temperature-Mediated Water Capture and Release,» *Angewandte Chemie International Edition,* vol. 59, pp. 11003-11009, 2020.

[48] B. Han y A. Chakraborty, «Synergistic ionic liquid encapsulated MIL-101 (Cr) metal-organic frameworks for an innovative adsorption desalination system,» *Journal of Cleaner Production,* vol. 474, p. 143565, 2024.

[49] K. Yang, T. Pan, Q. Lei, X. Dong, Q. Cheng y Y. Han, «A Roadmap to Sorption-Based Atmospheric Water Harvesting: From Molecular Sorption Mechanism to Sorbent Design and System Optimization,» *Environmental Science & Technology,* vol. 55, p. 6542–6560, 2021.

[50] G. Banerjee y K. Sengupta, «Pore size optimisation of humidity sensor—a probabilistic approach,» *Sensors and Actuators B: Chemical,* vol. 86, pp. 34-41, 2002.

[51] A. LaPotin, H. Kim, S. R. Rao y E. N. Wang, «Adsorption-Based Atmospheric Water Harvesting: Impact of Material and Component Properties on System-Level Performance,» *Accounts of Chemical Research,* vol. 52, p. 1588–1597, 2019.

[52] H. Kim, S. Yang, S. R. Rao, S. Narayanan, E. A. Kapustin, H. Furukawa, A. S. Umans, O. M. Yaghi y E. N. Wang, «Water harvesting from air with metal-organic frameworks powered by natural sunlight,» *Science,* vol. 356, pp. 430-434, 2017.

[53] Q. Li, Y. Ying, Y. Tao y H. Li, «Assemblable Carbon Fiber/Metal–Organic Framework Monoliths for Energy-Efficient Atmospheric Water Harvesting,» *Materials and Interfaces,* vol. 61, p. 1344–1354, 2022.

[54] A. Terzis, A. Ramachandran, K. Wang, M. Asheghi, K. E. Goodson y J. G. Santiago, «High-Frequency Water Vapor Sorption Cycling Using Fluidization of Metal-Organic Frameworks,» *Cell Reports Physical Science,* vol. 1, p. 100057, 2020.

[55] H. Furukawa, F. Gándara, Y.-B. Zhang, J. Jiang, W. L. Queen, M. R. Hudson y O. M. Yaghi, «Water Adsorption in Porous Metal–Organic Frameworks and Related Materials,» *Journal of the American Chemical Society,* vol. 136, p. 4369–4381, 2014.

[56] M. G. Gado, M. Nasser, A. A. Hassan y H. Hassan, «Adsorption-based atmospheric water harvesting powered by solar energy: Comprehensive review on desiccant materials and systems,» *Process Safety and Environmental Protection,* vol. 160, pp. 166-183, 2022.

[57] A. J. Rieth, S. Yang, E. N. Wang y M. Dincă, «Record Atmospheric Fresh Water Capture and Heat Transfer with a Material Operating at the Water Uptake Reversibility Limit,» *ACS Central Science,* vol. 3, p. 668–672, 2017.

[58] T. Lescouet, E. Kockrick, G. Bergeret, M. Pera-Titusa y D. Farrusseng, «Engineering MIL-53(Al) flexibility by controlling amino tags,» *Dalton Transactions,* vol. 40, pp. 11359-11361, 2011.

[59] L. Wang, K. Wang, H.-T. An, H. Huang, L.-H. Xie y J.-R. Li, «A Hydrolytically Stable Cu(II)-Based Metal−Organic Framework with Easily Accessible Ligands for Water Harvesting,» *ACS Applied Materials & Interfaces,* vol. 13, p. 49509–49518, 2021.

[60] L. Yang, K. B. Idrees, Z. Chen, J. Knapp, Y. Che, X. Wang, R. Cao, X. Zhang, H. Xing, T. Islamoglu y O. K. Farha, «Nanoporous Water-Stable Zr-Based Metal–Organic Frameworks for Water Adsorption,» *ACS Applied Nano Materials,* vol. 4, p. 4346–4350, 2021.

[61] N. Hanikel, M. S. Prévo, F. Fathieh, E. A. Kapustin, H. Lyu, H. Wang, N. J. Diercks, T. G. Glover y O. M. Yaghi, «Rapid Cycling and Exceptional Yield in a Metal-Organic Framework Water Harvester,» *ACS Central Science,* vol. 4, p. 4346–4350, 2021.

[62] H. Reinsch, M. A. v. d. Veen, B. Gil, B. Marszalek, T. Verbiest, D. d. Vo y N. S. , «Structures, Sorption Characteristics, and Nonlinear Optical Properties of a New Series of Highly Stable Aluminum MOFs,» *Chemistry of Materials,* vol. 25, p. 17–26, 2013.

[63] Y. Song, N. Xu, G. Liu, H. Qi, W. Zhao, B. Zhu, L. Zhou y J. Zhu, «High-yield solar-driven atmospheric water harvesting of metal–organic-framework-derived nanoporous carbon with fast-diffusion water channels,» *Nature Nanotechnology,* vol. 17, p. 857–863, 2022.

[64] B. Li, F.-F. Lu, X.-W. Gu, K. Shao, E. Wu y G. Qian, «Immobilization of Lewis Basic Nitrogen Sites into a Chemically Stable Metal–Organic Framework for Benchmark Water-Sorption-Driven Heat Allocations,» *Advanced Science,* vol. 14, p. 2105556, 2022.

[65] N. Hanikel, M. S. Prévot y O. M. Yaghi, «MOF water harvesters,» *Nature Nanotechnology,* vol. 15, p. pages348–355, 2020.

[66] A. Khutia, H. U. Rammelberg, T. Schmidt, S. Henninger y C. Janiak, «Water Sorption Cycle Measurements on Functionalized MIL-101Cr for Heat Transformation Application,» *Chemistry of Materials,* vol. 25, p. 790–798, 2013.

[67] G. E. Cmarik, M. Kim, S. M. Cohen y K. S. Walton, «Tuning the Adsorption Properties of UiO-66 via Ligand Functionalization,» *Langmuir,* vol. 28, p. 15606–15613, 2012.

[68] D. D. Borges, G. Maurin y D. S. Galvão, «Design of Porous Metal-Organic Frameworks for Adsorption Driven Thermal Batteries,» *MRS Advances,* vol. 2, p. 519–524, 2017.

[69] N. Hanikel, X. Pei, S. Chheda, H. Lyu, W. Jeong, J. Sauer, L. Gagliardi y O. M. Yaghi, «Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting,» *Science,* vol. 374, pp. 454-459, 2021.

[70] P. Iacomi, F. Formalik, J. Marreiros, J. Shang, J. Rogacka, A. Mohmeyer, P. Behrens, R. Ameloot, B. Kuchta y P. L. Llewellyn, «Role of Structural Defects in the Adsorption and Separation of C3 Hydrocarbons in Zr-Fumarate-MOF (MOF-801),» *Chemistry of Materials,* vol. 31, p. 8413–8423, 2019.

[71] Y. Hu, Y. Wang, Z. Fang, X. Wan, M. Dong, Z. Ye y X. Peng, «MOF supraparticles for atmosphere water harvesting at low humidity,» *Journal of Materials Chemistry A,* vol. 10, pp. 15116-15126, 2022.

[72] J. Xu, T. Li, J. Chao, S. Wu, T. Yan, W. Li, B. Cao y R. Wang, «Efficient Solar-Driven Water Harvesting from Arid Air with Metal–Organic Frameworks Modified by Hygroscopic Salt,» *Angewandte Chemie International Edition,* vol. 23, pp. 5202-5210, 2020.

[73] Q. Luo, M. Chen, D. Yu, T. Zhang, J. Zhao, L. Zhang, X. Han, M. Zhou, Y. Hou y Y. Zheng, «An Atmospheric Water-Harvester with Ultrahigh Uptake-Release Efficiency at Low Humidity,» *ACS Nano,* vol. 18, p. 14650–14660, 2024.

[74] S. Yuan, L. Feng, K. Wang, J. Pang, M. Bosch, C. Lollar, Y. Sun, J. Qin, X. Yang, P. Zhang, Q. Wang, L. Zou, Y. Zhang, L. Zhang, Y. Fang, J. Li y H.-C. Zhou, «Stable Metal–Organic Frameworks: Design, Synthesis, and Applications,» *Advanced Materials,* vol. 13, p. 1704303, 2018.

[75] X. Liu, X. Wang y F. Kapteijn, «Water and Metal–Organic Frameworks: From Interaction toward Utilization,» *Chemical Reviews,* vol. 120, p. 8303–8377, 2020.

[76] J. J. Low, A. I. Benin, P. Jakubczak, J. F. Abrahamian, S. A. Faheem y R. R. Willis, «Virtual High Throughput Screening Confirmed Experimentally: Porous Coordination Polymer Hydration,» *Journal of the American Chemical Society,* vol. 131, p. 15834–15842, 2009.

[77] N. C. Burtch, H. Jasuja y K. S. Walton, «Water Stability and Adsorption in Metal–Organic Frameworks,» *Chemical Reviews,* vol. 114, p. 10575–10612, 2014.

[78] H. J. Choi, M. Dincă, A. Dailly y J. R. Long, «Hydrogenstorage in water-stable metal–organic frameworks incorporating 1,3- and 1,4-benzenedipyrazolate,» *Energy & Environmental Science,* vol. 3, pp. 117-123, 2010.

[79] K. H. Cho, D. D. Borges, U.-H. Lee, J. S. Lee, J. W. Yoon, S. J. Cho, J. Park, W. Lombardo, D. Moon, A. Sapienza, G. Maurin y J.-S. Chang, «Rational design of a robust aluminum metal-organic framework for multi-purpose water-sorption-driven heat allocations,» *Nature Communications,* vol. 11, p. 5112, 2020.

[80] K. Yanagita, J. Hwang, J. A. Shamim, W.-L. Hsu, R. Matsuda, A. Endo, J.-J. Delaunay y H. Daiguji, «Kinetics of Water Vapor Adsorption and Desorption in MIL-101 Metal–Organic Frameworks,» *The Journal of Physical Chemistry C,* vol. 123, p. 387–398, 2019.

[81] Z. Lu, H. Tan, H. Lin, X. Cai, L. Du y Q. Liu, «Stepwise Node-Locking of a Mesoporous Zirconium Metal–Organic Framework Toward Enhanced Cycle Stability for Water Adsorption,» *Chemistry of Materials,* vol. 36, p. 2652–2660, 2024.

[82] J. Lee, D. Park, E. Jin, S. Lee, J. Lee, H. Oh y W. Choe, «Programmable Merged-Net Porphyrinic Metal–Organic Frameworks for Water Sorption,» *Advanced Functional Materials,* vol. 9, p. 2413200, 2025.

[83] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp y R. Q. Snurr, «Large-scale screening of hypothetical metal–organic frameworks,» *Nature Chemistry,* vol. 4, p. 83–89, 2012.

[84] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl y R. Q. Snurr, «Computation-Ready, Experimental Metal–Organic

Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals,» *Chemistry of Materials,* vol. 26, p. 6185–6192, 2014.

[85] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl y R. Q. Snurr, «Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019,» *Journal of Chemical & Engineering Data,* vol. 64, p. 5985–5998, 2019.

[86] C. M. D. A. c. e. m.-o. f. d. w. m.-l. p. f. i. m.-p. screening, «Guobin Zhao; Logan M. Brabson; Saumil Chheda; Ju Huang; Haewon Kim; Kunhuan Liu; Kenji Mochida; Thang D. Pham; Prerna; Gianmarco G. Terrones; Sunghyun Yoon; Lionel Zoubritzky; François-Xavier Coudert; Maciej Haranczyk; Heather J. Kulik; Seyed Mohamad Moos,» *Matter,* vol. 8, p. 102140, 2025.

[87] G. G. Terrones, S.-P. Huang, M. P. Rivera, S. Yue, A. Hernandez y H. J. Kulik, «Metal–Organic Framework Stability in Water and Harsh Environments from Data-Driven Models Trained on the Diverse WS24 Data Set,» *Journal of the American Chemical Society,* vol. 146, p. 20333–20348, 2024.

[88] L. T. Glasby, K. Gubsch, R. Bence, R. Oktavian, K. Isoko, S. M. Moosavi, J. L. Cordiner, J. C. Cole y P. Z. Moghadam, «DigiMOF: A Database of Metal–Organic Framework Synthesis Information Generated via Text Mining,» *Chemistry of Materials,* vol. 35, p. 4510–4524, 2023.

[89] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward y D. Fairen-Jimenez, «Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future,» *Chemistry of Materials,* vol. 29, p. 2618–2625, 2017.

[90] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein y R. Q. Snurr, «Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery,» *Matter,* vol. 4, pp. 1578-1597, 2021.

[91] A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford y D. S. Sholl, «The Open DAC 2023 Dataset and Challenges for Sorbent Discovery in Direct Air Capture,» arXiv, 2023.

[92] J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden y T. K. Woo, «ARC–MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning,» *Chemistry of Materials,* vol. 35, p. 900–916, 2023.

[93] D. Siderius, V. Shen, R. J. III y R. v. Zee, «NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials,» National Institute of Standards and Technology, 2020.

[94] Z. Zhang, H. Tang, M. Wang, B. Lyu, Z. Jiang y J. Jiang, «Metal–Organic Frameworks for Water Harvesting: Machine Learning-Based Prediction and Rapid Screening,» *ACS Sustainable Chemistry & Engineering,* vol. 11, p. 8148–8160, 2023.

[95] K. M. Jablonka, A. S. Rosen, A. S. Krishnapriyan y B. Smit, «An Ecosystem for Digital Reticular Chemistry,» *ACS Central Science,* vol. 9, p. 563–581, 2023.

[96] E. I. Ioannidis, T. Z. H. Gani y H. J. Kulik, «molSimplify: A toolkit for automating discovery in inorganic chemistry,» *Journal of Computational Chemistry,* vol. 37, pp. 2106-2117, 2016.

[97] L. Ward, A. Dunn, A. Faghaninia, N. E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster y A. Jain, «Matminer: An open source toolkit for materials data mining,» *Computational Materials Science,* vol. 152, pp. 60-69, 2018.

[98] I.-T. Sung, Y.-H. Cheng, C.-M. Hsieh y L.-C. Lin, «Machine Learning for Gas Adsorption in Metal–Organic Frameworks: A Review on Predictive Descriptors,» *Industrial & Engineering Chemistry Research,* vol. 64, p. 1859–1875, 2025.

[99] T. F. Willems, C. H. Rycroft, M. K. J. C. M. y M. Haranczyk, «Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials,» *Microporous and Mesoporous Materials,* vol. 149, pp. 134-141, 2012.

[100] J. T. Gostick, Z. A. Khan, T. G. Tranter, M. D. Kok, M. Agnaou, M. Sadeghi y R. Jervis, «PoreSpy: A Python Toolkit for Quantitative Analysis of Porous Media Images,» *Journal of Open Source Software,* vol. 4, p. 1296, 2019.

[101] C. H. Rycroft, «VORO++: A three-dimensional Voronoi cell library in C++,» *Chaos,* vol. 19, p. 041111 , 2019.

[102] L. Sarkisov, R. Bueno-Perez, M. Sutharson y D. Fairen-Jimenez, «Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database,» *Chemistry of Materials,* vol. 32, p. 9849–9867, 2020.

[103] M. Arjmandi y M. Pakizeh, «AN EXPERIMENTAL STUDY OF H2 AND CO2 ADSORPTION BEHAVIOR OF C-MOF-5 AND T-MOF-5: A COMPLEMENTARY STUDY,» *Brazilian Journal of Chemical Engineering,* vol. 33, pp. 225-233, 2016.

[104] A. Ö. Yazaydın, R. Q. Snurr, T.-H. Park, K. Koh, J. Liu, M. D. LeVan, A. I. Benin, P. Jakubczak, M. Lanuza, D. B. Galloway, J. J. Low y R. R. Willis, «Screening of Metal−Organic Frameworks for Carbon Dioxide Capture from Flue Gas Using a Combined Experimental and Modeling Approach,» *Journal of the American Chemical Society,* vol. 131, p. 18198–18199, 2009.

[105] R. Banerjee, H. Furukawa, D. Britt, C. Knobler, M. O'Keeffe y O. M. Yaghi, «Control of Pore Size and Functionality in Isoreticular Zeolitic Imidazolate Frameworks and their Carbon Dioxide Selective Capture Properties,» *Journal of the American Chemical Society,* vol. 131, p. 3875–3877, 2009.

[106] A. S. Krishnapriyan, M. Haranczyk y D. Morozov, «Topological Descriptors Help Predict Guest Adsorption in Nanoporous Materials,» arXiv, 2020.

[107] I.-T. Sung y L.-C. Lin, «In Silico Study of Metal–Organic Frameworks for CO2/CO Separation: Molecular Simulations and Machine Learning,» *The Journal of Physical Chemistry C,* vol. 127, p. 13886–13899, 2023.

[108] S. Kim y J.-H. Lee, «Data-driven discovery of novel metal organic frameworks with superior ammonia adsorption capacity,» *Materials Today Advances,* vol. 23, p. 100510, 2024.

[109] E. Gupta, D. Verma, S. Bhardwaj y S. M. N. Islam, «Detection of Open Metal Sites in Metal–Organic Frameworks Using Machine Learning,» de *International Conference on Innovative Computing and Communications*, 2023.

[110] A. Ahmed y D. J. Siegel, «Predicting hydrogen storage in MOFs via machine learning,» *Patterns,* vol. 2, p. 100291, 2021.

[111] K. R. Briling, Y. C. Alonso, A. Fabrizio y C. Corminboeuf, «SPAHM(a,b): Encoding the Density Information from Guess Hamiltonian in Quantum Machine Learning Representations,» *Journal of Chemical Theory and Computation,* vol. 20, p. 1108–1117, 2024.

[112] C. Altintas, O. F. Altundal, S. Keskin y R. Yildirim, «Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation,» *Journal of Chemical Information and Modeling,* vol. 61, p. 2131–2146, 2021.

[113] S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit y H. J. Kulik, «Understanding the diversity of the metal-organic framework ecosystem,» *Nature Communications,* vol. 11, p. 4068 , 2020.

[114] T. H. Nguyen, K. M. Le, L. H. Nguyen y T. N. Truong, «Atom-Based Machine Learning Model for Quantitative Property–Structure Relationship of Electronic Properties of Fusenes and Substituted Fusenes,» *ACS Omega,* vol. 8, p. 38441–38451, 2023.

[115] Y. Liu, T. Zhao, W. Ju y S. Shi, «Materials discovery and design using machine learning,» *Journal of Materiomics,* vol. 3, pp. 159-177, 2017.

[116] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji y T. K. Woo, «Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO2 capture,» *The Journal of Physical Chemistry Letters,* vol. 5, p. 3056–3060, 2014.

[117] H. Lin, Y. Yang, Y.-C. Hsu, J. Zhang, C. Welton, I. Afolabi, M. Loo y H.-C. Zhou, «Metal–Organic Frameworks for Water Harvesting and Concurrent Carbon Capture: A Review for Hygroscopic Materials,» *Advanced Materials,* vol. 21, p. 2209073, 2024.

[118] «RDKit: Open-Source Cheminformatics Software,» 2016. [En línea]. Available: http://www.rdkit.org.

[119] T. Shastry, Y. Basdogan, Z.-G. Wang, S. K. Kumar y M. R. Carbone, «Machine learning-based discovery of molecular descriptors that control polymer gas permeation,» *Journal of Membrane Science,* vol. 697, p. 122563, 2024.

[120] Y. Yang, Z. Yu y D. S. Sholl, «Machine Learning Models for Predicting Molecular Diffusion in Metal–Organic Frameworks Accounting for the Impact of Framework Flexibility,» *Chemistry of Materials,* vol. 35, p. 10156–10168, 2023.

[121] M. J. M. d. Oca-Estévez y R. Prosmiti, «Automated learning data-driven potential models for spectroscopic characterization of astrophysical interest noble gas-containing NgH2+ molecules,» *Artificial Intelligence Chemistry,* vol. 2, p. 100059, 2024.

[122] A. Chakraborty, B. B. Saha, K. C. Ng, S. Koyama y K. Srinivasan, «Theoretical Insight of Physical Adsorption for a Single Component Adsorbent + Adsorbate System: II. The Henry Region,» *Langmuir,* vol. 25, p. 7359–7367, 2009.

[123] Y. Lee, R. Poloni y J. Kim, «Probing gas adsorption in MOFs using an efficient ab initio widom insertion Monte Carlo method,» *Journal of Computational Chemistry,* vol. 37, pp. 2808-2815, 2016.

[124] B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri y R. Q. Snurr, «Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks,» *Molecular Systems Design & Engineering,* vol. 4, pp. 162-174, 2019.

[125] K. Shi, Z. Li, D. M. Anstine, D. Tang, C. M. Colina, D. S. Sholl, J. I. Siepmann y R. Q. Snurr, «Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials,» *Journal of Chemical Theory and Computation,* vol. 19, p. 4568–4583, 2023.

[126] Y. Yang, S. Guo, S. Li, Y. Wu y Z. Qiao, «Topological Data Analysis Combined with High-Throughput Computational Screening of Hydrophobic Metal–Organic Frameworks: Application to the Adsorptive Separation of C3 Components,» *Nanomaterials ,* vol. 14, p. 298, 2024.

[127] S. M. Moosavi, H. Xu, L. Chen, A. I. Cooper y B. Smit, «Geometric landscapes for material discovery within energy–structure–function maps,» *Chemical Science,* vol. 11, pp. 5423-5433, 2020.

[128] B. J. Bucior, A. S. Rosen, M. Haranczyk, Z. Yao, M. E. Ziebel, O. K. Farha, J. T. Hupp, J. I. Siepmann, A. Aspuru-Guzik y R. Q. Snurr, «Identification Schemes for Metal–Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis,» *Crystal Growth & Design,* vol. 19, p. 6682–6697, 2019.

[129] B. C. X. M. J. C. M. M. Carlos Nieto-Draghi, «Potential Energy Surface-Based Descriptors for Nanoporous Materials and its Applications to Classification and CO2 Gas Adsorption into Zeolites,» *ACS Applied Engineering Materials,* vol. 2, p. 478–491, 2024.

[130] Z. Deng y L. Sarkisov, «Engineering Machine Learning Features to Predict Adsorption of Carbon Dioxide and Nitrogen in Metal–Organic Frameworks,» *The Journal of Physical Chemistry C,* vol. 128, p. 10202–10215, 2024.

[131] I. B. Orhan, T. C. Le, R. Babarao y A. W. Thornton, «Accelerating the prediction of CO2 capture at low partial pressures in metal-organic frameworks using new machine learning descriptors,» *Communications Chemistry,* vol. 6, p. 214, 2023.

[132] C. Dösinger, T. Hammerschmidt, O. Peil, D. Scheiber y L. Romaner, «Descriptors based on the density of states for efficient machine learning of grain-boundary segregation energies,» *Computational Materials Science,* vol. 247, p. 113493, 2025.

[133] C. Lu, X. Wan, X. Ma, X. Guan y A. Zhu, «Deep-Learning-Based End-to-End Predictions of CO2 Capture in Metal–Organic Frameworks,» *Journal of Chemical Information and Modeling,* vol. 62, p. 3281–3290, 2022.

[134] J. Noh, J. Kim, H. S. Stein, B. Sanchez-Lengeling, J. M. Gregoire, A. Aspuru-Guzik y Y. Jung, «Inverse Design of Solid-State Materials via a Continuous Representation,» *Matter,* vol. 1, pp. 1370-1384, 2019.

[135] T.-H. Hung, Z.-X. Xu, D.-Y. Kang y L.-C. Lin, «Chemistry-Encoded Convolutional Neural Networks for Predicting Gaseous Adsorption in Porous Materials,» *The Journal of Physical Chemistry C,* vol. 126, p. 2813–2822, 2022.

[136] T. Xie y J. C. Grossman, «Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties,» arXiv, 2018.

[137] R. Wang, Y. Zou, C. Zhang, X. Wang, M. Yang y D. Xu, «Combining crystal graphs and domain knowledge in machine learning to predict metal-organic frameworks performance in methane adsorption,» *Microporous and Mesoporous Materials,* vol. 331, p. 111666, 2022.

[138] Y.-H. Cheng, I.-T. Sung, C.-M. Hsieh y L.-C. Lin, «Module-based machine learning models using sigma profiles of organic linkers to predict gaseous adsorption in metal-organic frameworks,» *Journal of the Taiwan Institute of Chemical Engineers,* vol. 165, p. 105728, 2024.

[139] Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr y A. Aspuru-Guzik, «Inverse design of nanoporous crystalline reticular materials with deep generative models,» *Nature Machine Intelligence,* vol. 3, p. 76–86, 2021.

[140] B. Oral, E. Can y R. Yildirim, «Analysis of photoelectrochemical water splitting using machine learning,» *International Journal of Hydrogen Energy,* vol. 47, pp. 19633-19654, 2022.

[141] E. Aytaç, F. E. Ahmed, F. Aziz, M. Khayet y N. Hilal, «A metadata survey of photothermal membranes for solar-driven membrane distillation,» *Separation and Purification Technology,* vol. 364, p. 132565, 2025.

[142] X. Zhou y M. Belkin, «Chapter 22 - Semi-Supervised Learning,» de *Academic Press Library in Signal Processing*, Elsevier, 2014, pp. 1239-1269.

[143] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue y G. Zhang, «Transfer learning using computational intelligence: A survey,» *Knowledge-Based Systems,* vol. 80, pp. 14-23, 2015.

[144] K. Zhu y Z. Zhang, «Equivariance is essential, local representation is a need: A comprehensive and critical study of machine learning potentials for tobermorite phases,» *Computational Materials Science,* vol. 246, p. 113363, 2025.

[145] V. V. Korolev, A. Mitrofanov, E. I. Marchenko, N. N. Eremin, V. Tkachenko y S. N. Kalmykov, «Transferable and Extensible Machine Learning-Derived Atomic Charges for Modeling Hybrid Nanoporous Materials,» *Chemistry of Materials,* vol. 32, p. 7822–7831, 2020.

[146] Z. Wang, Y. Zhou, T. Zhou y K. Sundmacher, «Identification of optimal metal-organic frameworks by machine learning: Structure decomposition, feature integration, and predictive modeling,» *Computers & Chemical Engineering,* vol. 160, p. 107739, 2022.

[147] R. Ma, Y. J. Colón y T. Luo, «Transfer Learning Study of Gas Adsorption in Metal–Organic Frameworks,» *ACS Applied Materials & Interfaces,* vol. 12, p. 34041–34048, 2020.

[148] M. Nurek y R. Michalski, «Combining Machine Learning and Social Network Analysis to Reveal the Organizational Structures,» *Applied Sciences,* vol. 10, p. 1699, 2020.

[149] H. Yan, M. Ma, Y. Wu, H. Fan y C. Dong, «Overview and analysis of the text mining applications in the construction industry,» *Heliyon,* vol. 8, p. e12088, 2022.

[150] R. Moriconi, M. P. Deisenroth y K. S. S. Kumar, «High-dimensional Bayesian optimization using low-dimensional feature spaces,» *Machine Learning,* vol. 109, p. 1925–1943, 2020.

[151] H. Park, Y. Kang, W. Choe y J. Kim, «Mining Insights on Metal–Organic Framework Synthesis from Scientific Literature Texts,» *Journal of Chemical Information and Modeling,* vol. 62, p. 1190–1198, 2022.

[152] Z. Qiao, Y. Yan, Y. Tang, H. Liang y J. Jiang, «Metal–Organic Frameworks for Xylene Separation: From Computational Screening to Machine Learning,» *The Journal of Physical Chemistry C,* vol. 125, p. 7839–7848, 2021.

[153] M. Jalali, M. Tsotsalas y C. Wöll, «MOFSocialNet: Exploiting Metal-Organic Framework Relationships via Social Network Analysis,» *Nanomaterials,* vol. 12, p. 704, 2022.

[154] A. Subasi, «Chapter 3 - Machine learning techniques,» de *Practical Machine Learning for Data Analysis Using Python*, 2020, pp. 91-202.

[155] E. Maggiori, Y. Tarabalka, G. Charpiat y P. Alliez, «Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification,» *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, pp. 645 - 657, 2016.

[156] A. Keramatfar, M. Rafiee y H. Amirkhani, «Graph Neural Networks: A bibliometrics overview,» *Machine Learning with Applications,* vol. 10, p. 100401, 2022.

[157] R. Long, X. Xia, Y. Zhao, S. Li, Z. Liu y W. Liu, «Screening metal-organic frameworks for adsorption-driven osmotic heat engines via grand canonical Monte Carlo simulations and machine learning,» *iScience,* vol. 24, p. 101914, 2021.

[158] M. Benmalek y A. Seddiki, «Particle swarm optimization-enhanced machine learning and deep learning techniques for Internet of Things intrusion detection,» *Data Science and Management,* vol. In Press, 2025.

[159] L. Pilz, M. Koenig, M. Schwotzer, H. Gliemann, C. Wöll y M. Tsotsalas, «Enhancing the Quality of MOF Thin Films for Device Integration Through Machine Learning: A Case Study on HKUST-1 SURMOF Optimization,» *Advanced Functional Materials,* vol. 24, p. 2404631, 2024.

[160] K. Rawal, A. Parthvi, D. K. Choubey y V. Shukla, «Chapter 16 - Prediction of leukemia by classification and clustering techniques,» de *Machine Learning, Big Data, and IoT for Medical Informatics*, 2021, pp. 275-295.

[161] S. Chander y P. Vijaya, «3 - Unsupervised learning methods for data clustering,» de *Artificial Intelligence in Data Mining*, 2021, pp. 41-64.

[162] F. Kherif y A. Latypova, «Chapter 12 - Principal component analysis,» de *Machine Learning*, 2020, pp. 209-225.

[163] M. A. Anvari, D. Rahmati y S. Kumar, «7 - t-Distributed stochastic neighbor embedding,» de *Dimensionality Reduction in Machine Learning*, 2025, pp. 187-207.

[164] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua y A. Lopez, «A comprehensive survey on support vector machine classification: Applications, challenges and trends,» *Neurocomputing,* vol. 408, pp. 189-215, 2020.

[165] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed y H. Arshad, «State-of-the-art in artificial neural network applications: A survey,» *Heliyon,* vol. 4, p. e00938, 2018.

[166] Z. Liu, D. Shen, S. Cai, Z. Tu y S. Li, «Machine learning-assisted prediction of water adsorption isotherms and cooling performance,» *Journal of Materials Chemistry A,* vol. 11, pp. 19455-19464, 2023.

[167] L. Li, Z. Shi, H. Liang, J. Liu y Z. Qiao, «Machine Learning-Assisted Computational Screening of Metal-Organic Frameworks for Atmospheric Water Harvesting,» *Nanomaterials,* vol. 12, p. 159, 2022.

[168] G. Barletta, S. Moitra, S. Derrible, A. Mathew, A. M. Nair y C. M. Megaridis, «Exploring machine learning models to predict atmospheric water harvesting with an ion deposition membrane,» *Journal of Water Process Engineering,* vol. 72, p. 107476, 2025.

[169] R. Batra, C. Chen, T. G. Evans, K. S. Walton y R. Ramprasad, «Prediction of water stability of metal–organic frameworks using machine learning,» *Nature Machine Intelligence,* vol. 2, p. 704–710, 2020.

[170] Z. Zhang, F. Pan, S. A. Mohamed, C. Ji, K. Zhang, J. Jiang y Z. Jiang, «Accelerating Discovery of Water Stable Metal−Organic Frameworks by Machine Learning,» *Small,* vol. 20, p. 2405087, 2024.

[171] Z. Liu, D. Shen, S. Cai, Z. Tu y S. Li, «Machine learning-assisted prediction of water adsorption isotherms and cooling performance,» *Journal of Materials Chemistry A,* vol. 11, pp. 19455-19464, 2023.

[172] R. Goeminne y V. V. Speybroeck, «Ab Initio Predictions of Adsorption in Flexible Metal–Organic Frameworks for Water Harvesting Applications,» *Journal of the American Chemical Society,* vol. 147, p. 3615–3630, 2025.