ChemComm

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2026, **62**, 3233

Machine-learning-guided design of MOF-based electrocatalysts for sustainable ammonia production

Bo Han,[a] Marcus de Carvalho,[b] Jie Zhang,*[b] Hui Mao *[c] and Qingyu Yan *[ad]

Metal–organic frameworks (MOFs) are highly versatile materials known for their exceptional properties, including high specific surface area, tunable pore structures, and multifunctionality, making them invaluable in fields of electrochemical ammonia synthesis. Over the past decade, MOF-based systems have evolved from initial N_2-reduction prototypes into advanced catalysts that can convert nitrates and nitrites into ammonia. This offers a broader perspective on the electrochemical nitrogen cycle. However, their structural complexity poses significant challenges to traditional design and optimization approaches. This review first critically summarizes the recent advances in MOF development for electrochemical ammonia synthesis. Next, this review systematically explores the transformative role of machine learning (ML) in advancing MOF research. It also addresses the 8 major challenges and limitations currently facing this intersection, including data scarcity, model interpretability, and inverse design. To address these challenges and guide future progress, the review proposes 4 clear and practical future directions: leveraging explicable AI to improve model transparency, integrating active learning with automated platforms to optimize experimental workflows, exploring the synergy between ML and quantum computing to simulate complex structures, and fostering multidisciplinary collaboration for holistic innovation. By bridging computational intelligence and materials discovery, this work underscores the central role of ML in shaping the next generation of MOF-based technologies for sustainable energy and environmental applications.

Received 15th December 2025,
Accepted 15th January 2026

DOI: 10.1039/d5cc07118f

rsc.li/chemcomm

1. Introduction

Ammonia is at the heart of modern industry and agriculture, supporting the production of fertilizers, the storage of hydrogen, and carbon-free energy strategies.[1] However, the conventional Haber–Bosch synthesis process remains one of the most energy- and carbon-intensive chemical operations, consuming nearly 2% of global energy and emitting vast amounts of CO_2.[2] The push towards sustainability has rekindled interest in electrocatalytic nitrogen fixation, where renewable electricity, water and nitrogen-bearing feedstocks replace fossil fuels.[3]

The electrochemical nitrogen reduction reaction offers the potential to produce ammonia under mild conditions. However, the immense bond energy of N_2 (941 kJ mol^{-1}) and its limited solubility result in sluggish kinetics and low Faradaic efficiencies.[4] Meanwhile, electrocatalytic nitrate/nitrite reduction offers a more thermodynamically favorable route while addressing nitrate pollution from fertilizers and industrial waste.[5] Together, these approaches form the basis of a new electrochemical nitrogen economy.[6]

Metal–organic frameworks (MOFs) are highly versatile materials with exceptional properties such as chemical multifunctionality, tunable pore structure, and high specific surface area.[7] The modular architecture of MOFs, combining metal nodes with organic linkers, allows for customized functionality tailored to specific applications.[8–10] These properties have positioned MOFs as critical candidates for electrochemical ammonia synthesis.[11]

Despite their enormous potential, MOFs exhibit structural complexity due to several factors that make their design and optimization challenging for traditional approaches: (i) diversity of building blocks.[12] MOFs are composed of many metal nodes and organic linkers, allowing for an almost infinite number of structural combinations. This structural diversity results in a high-dimensional design space. (ii) Tunable properties.[13] MOFs can be tailored at multiple levels, such as pore size, surface area, topology, and functionalization, greatly increasing the number of potential variants. (iii) High crystallographic symmetry.[14] Many MOFs have intricate geometries

[a] *Energy Research Institute@NTU (ERI@N), Nanyang Technological University, Singapore 637141, Singapore. E-mail: alexyan@ntu.edu.sg*
[b] *College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore. E-mail: zhangj@ntu.edu.sg*
[c] *College of Chemical Engineering, Sichuan University, Chengdu 610065, P. R. China. E-mail: rejoice222@163.com*
[d] *School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore*

with high symmetry and complex connectivity patterns, making their characterization and computational modeling difficult. (iv) Dynamic behavior.[15] MOFs can exhibit flexibility, breathing, and other dynamic properties that add a layer of complexity to understanding and predicting their behavior under different conditions. (v) Defects and disorder.[16] The presence of defects, partial occupancies, and framework disorder in many MOFs can affect their performance but are difficult to quantify and model accurately. (vi) Interplay of chemical interactions.[17] The performance of MOFs is often influenced by a combination of chemical interactions, such as van der Waals forces, hydrogen bonding, and π–π interactions, which require advanced computational tools to model effectively. (vii) Multifunctionality.[18] Many MOFs are designed for multifunctional applications that require simultaneous optimization of multiple, often competing, properties such as stability, adsorption capacity, and selectivity. (viii) Scalability and reproducibility.[19] Scaling up MOF synthesis while maintaining

Bo Han

Bo Han is currently a Research Assistant Professor (RAP) at Nanyang Technology University (NTU). He received his BS degree (2016) and PhD degree (2021) from the Mechanical and Aerospace Engineering (MAE) Department of NTU, Singapore. His research focuses on developing nano-porous materials for applications in adsorption, energy storage, membranes and electrocatalysts. His research topics cover metal–organic framework (MOF) materials, adsorption assisted cooling/heat pump/desalination/water harvesting, metal ion separation, and electrochemical catalytic activity. Dr Han has published >25 papers and 3 book chapters as the first-author, and his citations received are >1000 (H-index = 21, i10-index = 24). Dr Han is also actively involved in services to the global scientific community, he has been invited to give oral presentations at over 10 international conferences, and he has also been an invited reviewer for several journals.

Marcus de Carvalho

Marcus de Carvalho is a researcher at Nintendo European Research and Development, where he manages the exploration, development, and optimization of machine learning algorithms. He received his PhD in 2023 from the College of Computing and Data Science at Nanyang Technological University, Singapore. His doctoral research focused on transfer learning, continual learning, and real-time learning, with an emphasis on deploying adaptive models in dynamic, resource-constrained environments. Building on this foundation, he now contributes to projects in computer vision and reinforcement learning, investigating how perception and decision-making can be tightly integrated to create more responsive, data-efficient systems. Across both academia and industry, his work centers on practical, deployable AI that can learn from limited data, adapt over time, and deliver robust performance in real-world, interactive applications.

Jie Zhang

*Jie Zhang is a Professor in the College of Computing and Data Science at NTU Singapore, where he leads the Artificial Intelligence Group. He received his PhD from the Cheriton School of Computer Science, University of Waterloo, earning the Alumni Gold Medal in 2009. He joined NTU as an Assistant Professor and became an Associate Professor in 2015. From 2017 to 2018, he was a Tan Chin Tuan Exchange Fellow at New York University. Since 2020, he has also been an Adjunct Fellow at SIMTech, A*STAR. His research focuses on user modeling and intelligent agents that provide personalized recommendations on which people to interact with, with notable contributions to trust and preference modeling that carefully integrate user subjectivity and social trust relationships.*

Hui Mao

Hui Mao is currently an associate professor in the college of chemical engineering at Sichuan University. She obtained her PhD. from the Biomass Chemistry and Engineering Department of Sichuan University. She joined the School of Materials Science and Engineering of Nanyang Technological University as a postdoctoral research fellow. She joined Sichuan University as an associate professor in 2020. Mao's research focuses on advanced electrode materials based on biomass and functional utilization of biomass.

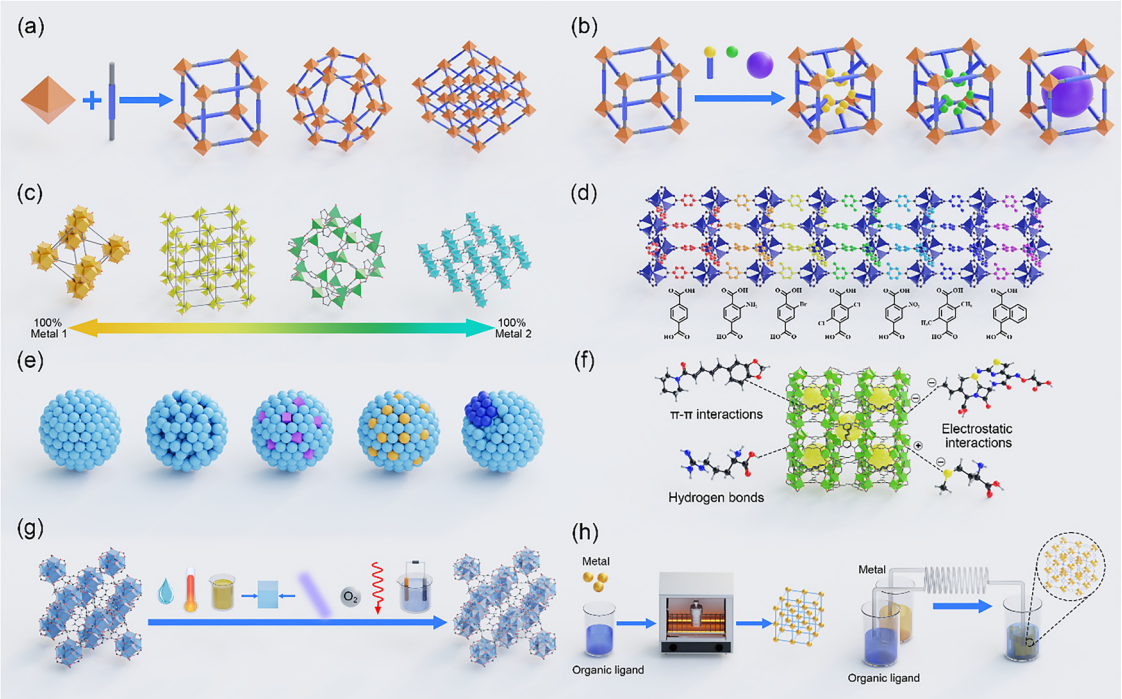

Fig. 1 The structure complexity of MOFs. (a) Theoretical infinite frameworks of MOFs by direct synthesis, (b) various post-synthesis methods of MOFs, (c) various metal/metal clusters, (d) various organic linkers, (e) defects and disorder of MOFs, (f) multiple chemical interactions between MOFs and the surrounding environment, (g) structure variation of MOFs under complicated conditions, and (h) structural variation of MOFs under continuous synthesis flow.

precise structural and property control adds practical complexity to their design and application. The structural complexity of MOFs is shown schematically in Fig. 1, which raises the motivation to create advanced techniques for precise design and development of MOFs.

The rapid growth of data generation in materials science, coupled with advances in machine learning (ML) algorithms, has made ML a transformative approach to solving complex scientific challenges.[20] ML excels at unraveling non-linear relationships in large data sets, enabling prediction of properties and optimization of materials with unprecedented accuracy and speed. This capability is particularly relevant for MOFs, whose diverse structural and chemical compositions require sophisticated tools to uncover and exploit structure–property relationships.[21] By integrating ML, researchers have significantly accelerated the processes of screening, developing, and refining MOFs for cutting-edge applications, bridging the gap between computational efficiency and experimental feasibility.[22]

Machine learning (ML) provides a data-driven framework that can identify complex, nonlinear relationships between MOFs and their catalytic performance. These relationships are often inaccessible through conventional, intuition-based design. In the context of electrochemical ammonia synthesis, ML can be integrated into MOF research by learning from existing datasets that correlate metal nodes, ligand chemistry, defect structures and electronic descriptors with activity towards N_2 activation and nitrate or nitrite reduction. High-dimensional optimization enables ML to rapidly predict key properties such as adsorption energies, charge transfer characteristics, conductivity pathways and intermediate binding energetics. These insights can then be used to guide the rational design of MOF architectures with active sites that are

Qingyu Yan

Qingyu Yan is currently a professor in the School of Materials Science and Engineering at Nanyang Technology University. He obtained his BS in Materials Science and Engineering, from Nanjing University. He finished his PhD at the Materials Science and Engineering Department of State University of New York at Stony Brook. After that, he joined the Materials Science and Engineering Department of Rensselaer Polytechnic Institute as a postdoctoral research associate. He joined the School of Materials Science and Engineering of Nanyang Technological University as an assistant professor in early 2008 and became a professor in 2018. He is currently the vice chair of the Electrochemical Society, Singapore Section. He has been a fellow of the Royal Society of Chemistry since 2018. He has published >300 papers in the research areas of: (1) battery development; (2) electronic waste recycling, (3) thermoelectric materials and (3) electrocatalytic process for energy conversion.

tailored to specific applications, coordination environments that are optimized, and electronic structures that can be tuned as required. Furthermore, ML can support the optimization of synthesis by identifying the optimal solvothermal parameters, linker–metal combinations and defect densities that maximize catalytic efficiency. By integrating ML into the research process, the discovery and refinement of MOF electrocatalysts is accelerated, reducing reliance on trial-and-error experimentation and enhancing the precision and scalability of catalyst development.

This review explores the intersection of ML and MOF research, highlighting the transformative role of ML in revolutionizing the design and development of MOFs. The review identifies and critically examines 8 key challenges that currently impede the broader adoption of ML in MOF research, such as the scarcity of high-quality data, challenges in interpreting complex ML models, and the inverse design and experimental validation. To address these barriers, this work proposes 4 actionable strategies: increasing model transparency through explicable AI, integrating active learning with automated experimentation, leveraging quantum computing to simulate complex material behavior, and fostering interdisciplinary collaboration to drive innovation. By addressing these issues, this review aims to highlight pathways for harnessing the full potential of ML to advance MOF-based technologies for energy and environmental sustainability.

2. Structural strategies for MOF design

The EN_2RR is a sustainable alternative to the energy-intensive Haber–Bosch process, as it enables ammonia synthesis under ambient conditions using renewable electricity. In the EN_2RR process, molecular nitrogen is activated at the catalyst surface and subsequently reduced through multiple proton–electron transfer steps to form ammonia.[23] However, the reaction remains challenging from a kinetic perspective due to the strong triple bond in N_2, the low solubility of N_2 in aqueous electrolytes and the hydrogen evolution reaction, which often dominates under typical electrochemical conditions. Effective electrocatalysts must therefore provide coordinatively accessible active sites that can absorb and polarize N_2, stabilize key intermediates such as *NNH and facilitate selective protonation pathways, while suppressing hydrogen adsorption.[24] MOFs have therefore emerged as promising candidates, as they can be tailored to enhance N_2 activation and electron transfer through their tunable metal nodes, adjustable ligand environments, and customizable pore microenvironments. Moving beyond N_2 towards nitrate and nitrite electroreduction broadens the concept of the nitrogen cycle. Unlike inert N_2, nitrates (NO_3^-) and nitrites (NO_2^-) are soluble, reactive and prevalent in the environment, often found in agricultural runoff or wastewater.[25,26] Electrochemical conversion of these compounds to NH_3 yields a valuable product and mitigates eutrophication and groundwater contamination. MOFs have proven to be ideal hosts for these transformations.[27] Their open metal nodes catalyze multi-step proton–electron transfer sequences while confining intermediates such as NO_2, NO and NH_2OH. Therefore, understanding and engineering the mechanistic landscape of ENRR is essential for developing efficient and selective MOF-based electrocatalysts for ammonia production.[28,29] This chapter categorizes the main structural strategies for MOF designs and critically analyses the recent advances in MOF-based catalysts for ammonia synthesis.

2.1. Metal node engineering

Metal node engineering has emerged as one of the most powerful strategies for improving the performance of MOFs in ammonia synthesis, whether from inert nitrogen or from nitrate and nitrite. By tailoring the identity, oxidation state, coordination geometry, and electronic environment of the metal centers, MOFs can host active sites that can selectively absorb, activate, and reduce nitrogen-containing species *via* multistep proton–electron transfer pathways. The incorporation of secondary metals or the formation of bimetallic nodes enables tandem or synergistic catalysis, whereby different metals specialize in complementary elementary steps. This lowers reaction barriers and improves selectivity towards NH_3. Overall, metal node engineering provides a straightforward and flexible approach to modulating the intrinsic catalytic activity and reaction microenvironment of MOFs, establishing it as a fundamental design principle for advanced electrocatalysts in sustainable ammonia production.

Yan *et al.*[30] recently reported one representative example of this phenomenon. Their work describes a bimetallic Cu–Co MOF (Cu_xCo_y-DHTA) that acts as a catalyst for the highly efficient electrochemical reduction of nitrates to ammonia. Fig. 2(a) illustrates the hydrothermal synthesis of Cu_xCo_y-DHTA and its electrochemical reconstruction during nitrate reduction. When a cathodic potential is applied, some of the Cu species in the bimetallic MOF are reduced to form finely dispersed Cu(0) nanoparticles. Meanwhile, the Co-MOF framework remains largely intact, resulting in a composite catalyst with embedded Cu sites. Fig. 2(b) presents the proposed dual-site mechanism. The Cu(0) nanoparticles catalyze the reduction of NO_3^- to NO_2^- preferentially, while the adjacent Co sites in the MOF promote the water dissociation and hydrogen donation steps required for the subsequent hydrogenation of NO_2^- to NH_3. The MOF pores enable the efficient transfer of intermediates between these sites, thereby supporting a tandem Cu–Co relay pathway for the selective conversion of nitrates to ammonia. Consequently, the optimized $Cu_{1/2}Co_{1/2}$-DHTA achieves Faradaic efficiencies for NH_3 of over 95% within a broad potential range of −0.8 to −1.0 V *vs.* RHE, with a maximum NH_3 production rate of 20.02 mg h^{-1} cm^{-2}, surpassing the performance of most previously reported MOF-based eNO_3RR catalysts. For the mechanism study, Fig. 2(c) further illustrates the Cu–Co dual-site tandem mechanism, in which Cu(0) drives the initial NO_3^--to-*NO_2 step, while the Co sites govern the subsequent hydrogenation steps towards NH_3. Furthermore, Fig. 2(d–f) summarize DFT results comparing monometallic Cu-DHTA with bimetallic $Cu_{1/2}Co_{1/2}$-DHTA.

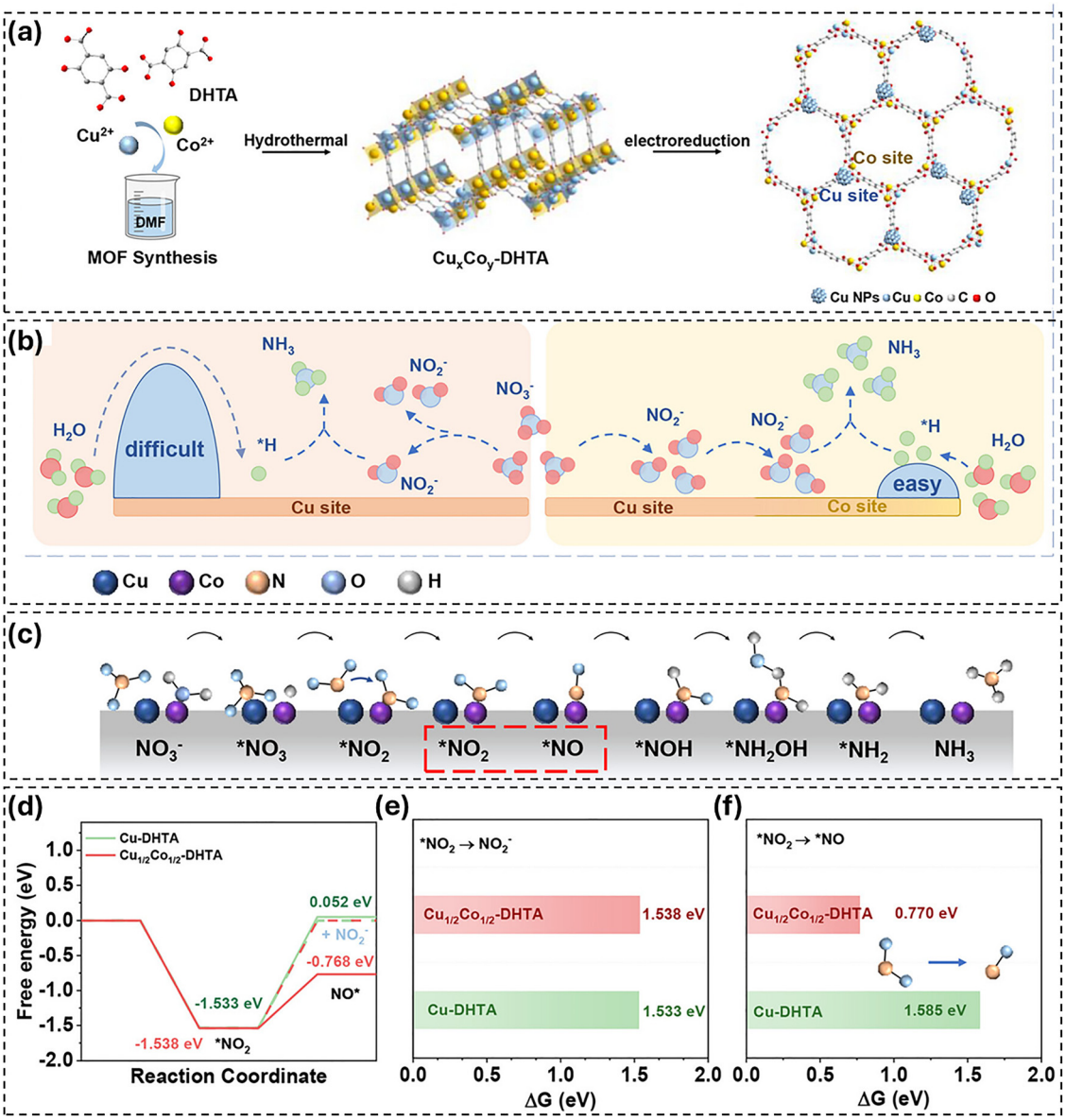

Fig. 2 (a) Schematic of Cu_xCo_y-DHTA preparation and electro-reduction. (b) Mechanism diagram of CuCo dual-site electrocatalytic nitrate reduction. (c) Schematic of the CuCo bimetallic Dual-Site tandem catalysis mechanism. (d) Gibbs free energy diagrams of NO_3RR over Cu-DHTA and $Cu_{1/2}Co_{1/2}$-DHTA. (e) Gibbs free energy diagrams of $*NO_2$ to NO_2^- over Cu-DHTA and $Cu_{1/2}Co_{1/2}$-DHTA. (f) Gibbs free energy diagrams of $*NO_2$ to *NO over Cu-DHTA and $Cu_{1/2}Co_{1/2}$-DHTA. Reproduced with permission.[30] Copyright 2025, Wiley.

Although the initial reduction of NO_3^- to $*NO_2$ is thermodynamically favorable on both catalysts, hydrogenation of $*NO_2$ is energetically hindered on Cu-DHTA, resulting in desorption as NO_2^-. In contrast, the Co sites in the bimetallic system exhibit significantly lower free energy barriers for $*NO_2$ hydrogenation, facilitating the retention and continued reduction of intermediates to NH_3. These calculations provide mechanistic support for the experimentally observed tandem catalysis. Overall, this study provides convincing evidence that bimetallic MOF precursors that form dual-site architectures offer a powerful structural approach to efficient nitrate-to-ammonia electrocatalysis. However, it also highlights the need for a deeper understanding of the reconstruction dynamics, durability and scalability required for real-world applications.

In another study, Sun *et al.*[31] reported a bimetallic structural approach involving the immobilization of cobalt clusters within the cavities of Cu-based MOF-818 to form a Cu–Co dual-site catalyst that promotes the electrocatalytic reduction of nitrate to ammonia. Mechanistically, the Cu sites preferentially adsorb and reduce NO_3^- to NO_2^-, while the encapsulated Co clusters catalyze the kinetically demanding NO_2^--to-NH_3 steps. This enables a synergistic tandem pathway that markedly lowers the free energy barrier of the rate-determining $*NO_2 \rightarrow *NO$ transformation. The MOF-818 (Cu)–Co catalyst benefits from this dual-site cooperation, achieving NH_3 Faradaic efficiencies above 90% across −1.3 to −1.8 V (*vs.* Ag/AgCl), with a peak FE of ~99.5% at −1.5 V and an NH_3 yield rate of 35.0 mol h^{-1} $g_{Cu,Co}^{-1}$. This significantly outperforms both pristine MOF-818

(Cu) and Co nanoparticles. This work critically demonstrates that embedding secondary metal clusters within MOF cages effectively couples adsorption and reduction functions within confined, nanoscale reactors. However, this success relies heavily on controlled metal dispersion and stable MOF architectures, which could limit the strategy's scalability and applicability to other MOF systems.

Recently, Wang *et al.*[32] presented on a phosphate-assisted iron-modified MOF-808 (Fe@MOF-808-P) catalyst. This catalyst uses metal-center engineering, specifically the incorporation of Fe(III) into the Zr-based MOF-808 framework, combined with phosphate ligand modulation, to boost nitrate electroreduction to ammonia. The iron nodes act as the primary active sites for nitrate adsorption and stepwise hydrogenation, while the coordinated hydrogen phosphate species accelerate proton transfer and stabilize reaction intermediates, collectively lowering the energy barriers for nitrate to ammonia conversion. Mechanistic analysis shows that the Fe(III) centers promote NO_3^- chemisorption and facilitate electron-proton coupling, while the phosphate groups create a more favorable microenvironment for the hydrogenation steps, thereby enhancing selectivity. Electrochemical measurements demonstrate a Faradaic efficiency of 90.8% ± 0.5% at −0.9 V *vs.* RHE, with an NH_3 yield rate of 420.4 μmol h^{-1} cm^{-2} and a specific activity of 1035.9 mg NH_3 h^{-1} mg Fe. The system exhibits excellent durability over repeated cycles. The structural strategy of dual incorporation of Fe(III) and phosphate proves effective because it couples precise active-site design with microenvironment engineering, enabling high intrinsic activity of the Fe catalytic center and accelerated proton supply. However, the reduced surface area after P/Fe loading may limit scalability and the system still relies on relatively high NO_3^- concentrations for peak performance. This indicates room for optimization in terms of mass transport and active-site exposure.

2.2. MOF-derived composite

MOF-derived composites have emerged as a highly effective class of electrocatalyst for the sustainable synthesis of ammonia from nitrogen and nitrate sources. This is due to their unique ability to convert the ordered atomic structure of MOFs into functional metal/metal oxide–carbon hybrids through pyrolysis. During this thermal transformation process, the metal nodes, which are uniformly dispersed within the MOF framework, convert into catalytically active nanoparticles. Meanwhile, the organic linkers generate conductive, porous, heteroatom-doped carbon matrices that enhance electron transport, expose active sites and stabilize reaction intermediates. Despite the reconstruction that occurs during pyrolysis, which can limit atomic precision, MOF-derived composites remain a powerful and versatile strategy for engineering efficient active sites and conductive reaction microenvironments. This makes them one of the most promising platforms for electrocatalytic ammonia production under ambient conditions.

Hu *et al.*[33] reported a bimetallic oxide composite derived from a MOF, in which Cu-doped Co_3O_4 nanosheets are grown on three-dimensional nickel foam (Cu–Co_3O_4/NF). This composite was constructed using Cu-ZIF-67 as a precursor, followed by calcination, to create an electrocatalyst that can reduce nitrates to ammonia at ultra-low concentrations. This structural approach combines MOF-derived spinel Co_3O_4 with the incorporation of copper and a conductive nickel foam (NF) scaffold. This enhances the density of active sites, improves charge transport and suppresses the hydrogen evolution reaction (HER). Here, Fig. 3 systematically compares the electrocatalytic performance of Cu–Co_3O_4/NF and Co_3O_4/NF towards nitrate reduction. Fig. 3(a) shows that upon the addition of NO_3^-, both electrodes exhibit enhanced cathodic currents from approximately −0.2 V *vs.* RHE, indicating the onset of nitrate reduction. However, Cu–Co_3O_4/NF consistently displays higher current densities than Co_3O_4/NF, reflecting superior conductivity and catalytic activity. Fig. 3(b) and (c) show the potential-dependent NH_3 yield rate and Faradaic efficiency, respectively, in 2 mM NO_3^-. The Cu-doped catalyst achieves a maximum NH_3 yield of 0.137 mmol h^{-1} cm^{-2} at −0.83 V *vs.* RHE and a peak FE of 92.4% at −0.53 V *vs.* RHE. These values are significantly higher than those of the Co_3O_4/NF catalyst, confirming beneficial Cu–Co synergy. Fig. 3(d) shows the effect of nitrate concentration on the NH_3 yield rate. It demonstrates that the yield rate increases with increasing NO_3^- concentration up to ∼20 mM. Notably, the FE remains high and still reaches 92.4% at only 2 mM NO_3^-, highlighting the catalyst's suitability for low-concentration conditions. Fig. 3(e) shows good durability, with both the NH_3 yield rate and the FE fluctuating only slightly over six consecutive electrolysis cycles at −0.53 V *vs.* RHE. Finally, the Nyquist plots in Fig. 3(f) show that Cu–Co_3O_4/NF has much smaller charge-transfer resistance than Co_3O_4/NF. This indicates faster interfacial electron transfer, which corroborates the enhanced kinetic performance of the MOF-derived Cu-doped composite. Mechanistic studies based on *in situ* FTIR and DFT calculations suggest that the reduction of NO_3^- proceeds *via* an H-assisted surface pathway. This pathway involves the following sequence of steps: NO_3^- → *NO_3 → *NO_2 → *NO_2H → *NO → *NOH → *N → *NH → *NH_2 → *NH_3. Meanwhile, Cu doping modulates the electronic structure of the Co sites, strengthens NO_3^- adsorption and lowers the free-energy barrier of the key deoxygenation steps. This significantly disfavors the HER. In addition, the authors also emphasized that preparing MOF-derived Cu-doped Cu-Co_3O_4/NF is simple and inexpensive.

In another study, Chen *et al.*[34] developed an MOF-derived composite electrocatalyst in which an adenine-based cobalt-bio-MOF undergoes pyrolysis to form cobalt/nitrogen-doped carbon nanospheres (Co@NC). This process enables the uniform dispersion of metallic cobalt nanoparticles within an N-rich carbon matrix, facilitating the efficient conversion of nitrates to ammonia. This structural approach uses the nitrogen-containing organic ligand to create an abundance of pyridinic, pyrrolic and graphitic nitrogen sites during carbonization. These sites coordinate with the cobalt and modify the electronic structure, facilitating selective NO_x hydrogenation. Mechanistically, the Co nanoparticles act as primary active centers for NO_3^- adsorption and multi-electron reduction *via*

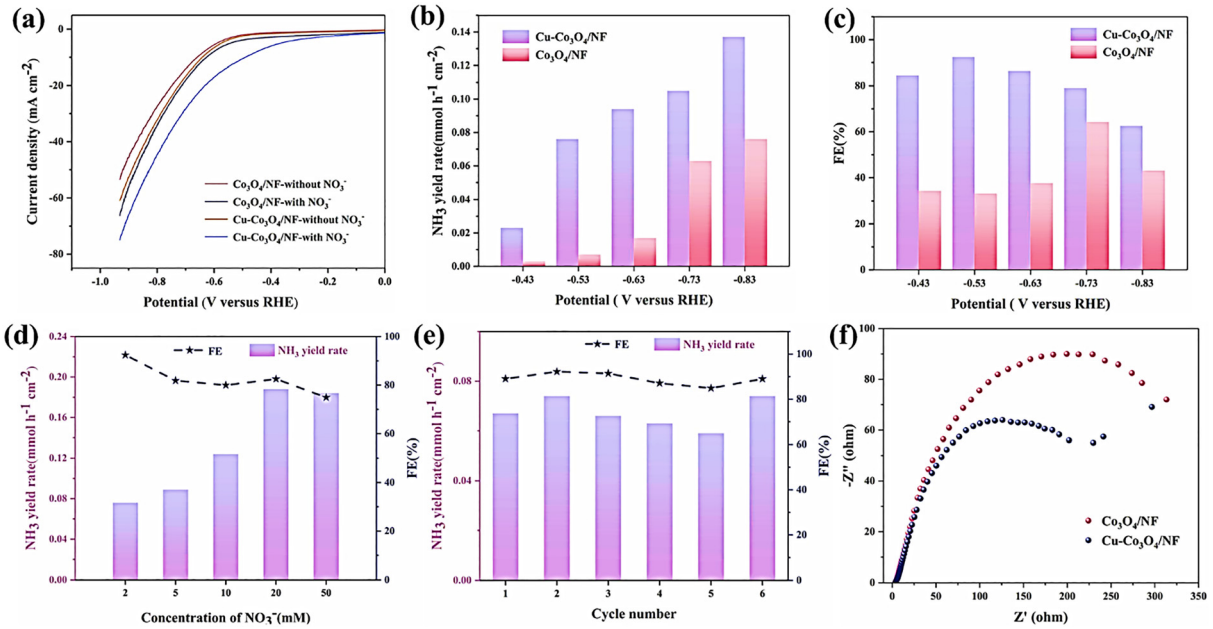

Fig. 3 Electrocatalytic performances. (a) LSV curves of Cu–Co_3O_4/NF and Co_3O_4/NF in Ar-saturated 0.2 M K_2SO_4 with and without 2 mM NO_3^-. The comparisons of the (b) NH_3 yield rate and (c) FE between Cu–Co_3O_4/NF and Co_3O_4/NF in contained 2 mM NO_3^- electrolyte under a range of potentials for 30 min. (d) The NH_3 yield rate and FE of Cu–Co_3O_4/NF at −0.53 V *vs.* RHE in electrolytes with various concentrations of NO_3^- for 30 min. (e) The NH_3 yield rate and FE of cycling tests for Cu–Co_3O_4/NF at −0.53 V *vs.* RHE for 30 min. (f) The Cu–Co_3O_4/NF and Co_3O_4/NF Nyquist plots. Reproduced with permission.[33] Copyright 2024, Elsevier.

$*NO_3$, $*NO_2$, *NO and $*NH_x$ intermediates. Meanwhile, the N-doped carbon shell enhances electron transport and stabilizes reactive species, collectively suppressing competing HERs. Consequently, Co@NC exhibits a Faradaic efficiency of 96.5% and an exceptionally high NH_3 yield rate of 758.0 μmol h^{-1} mg_{cat}^{-1} in an alkaline electrolyte, alongside excellent cycling durability. The proposed MOF-derived approach is effective because it yields atomically coupled Co–N–C motifs and high conductivity within a robust carbon matrix. However, its structural precision is limited by uncontrollable reconstruction during pyrolysis and the catalyst's performance still relies on relatively negative potentials. This suggests that further optimization of active-site exposure and electronic structure is required for practical advancement.

2.3. Defect engineering and conductivity enhancement

Defect engineering[35] and conductivity enhancement[36,37] are two highly synergistic strategies for activating intrinsically inert MOFs for efficient electrochemical ammonia synthesis from nitrogen and nitrate sources. Introducing coordinatively unsaturated metal sites, linker vacancies, or extended ligands generates new electronic states near the Fermi level. Concurrently, conductivity-oriented modifications, such as introducing mixed-valence centers, embedding conductive metal clusters or coupling MOFs with π-conjugated carbon domains, facilitate rapid charge transport and suppress competitive hydrogen evolution. Together, these approaches transform MOFs from insulating porous hosts into electronically active catalytic platforms and have been gradually applied for MOF structural modification.

For example, Han *et al.*[38] investigated the use of a main-group-element-based aluminum fumarate (Al-Fum) MOF as an electrocatalyst for the reduction of nitrogen to ammonia. They demonstrated that defect engineering *via* solvent-assisted linker exchange (SALE) is an effective structural strategy for activating otherwise inert aluminum sites. In this approach, pristine Al-Fum, constructed from Al–O chains and fumarate linkers, is partially subjected to linker extension. This generates coordinatively unsaturated and electronically modified Al defect sites, while preserving the parent framework topology largely intact. It is found that only the defective Al-Fum exhibits significant N_2RR activity. It achieves a maximum NH_3 production rate of 53.9 μg(NH_3) h^{-1} mg_{cat}^{-1} in 0.4 M K_2SO_4, with a Faradaic efficiency of 73.8% in 0.1 M K_2SO_4 at −0.15 V *vs.* RHE under ambient conditions. From a theoretical study, Fig. 4 shows how Al-site defects affect the electronic structure and N_2RR pathway. The projected density of states (DOS) for Al in pristine Al-Fum (Fig. 4(a)) reveals no notable states at the Fermi level. This suggests that the Al centers are electronically 'silent' and unable to engage effectively in N_2 adsorption or activation. By contrast, the DOS of Al in defective Al-Fum (Fig. 4(b)) exhibits a pronounced peak at the Fermi level, indicating increased electronic activity and the formation of defective Al sites that can interact with N_2 and N_2RR intermediates. Fig. 4(c) shows the free-energy diagram mapping the N_2RR pathway on a defective Al site at 0 V *vs.* SHE. This reveals that the initial N_2 → *NNH conversion is the potential-determining step, with a free-energy increase of 1.815 eV. Energetically accessible protonation steps then occur along an alternating associative mechanism. In addition, Fig. 4(d) shows side and top views of the

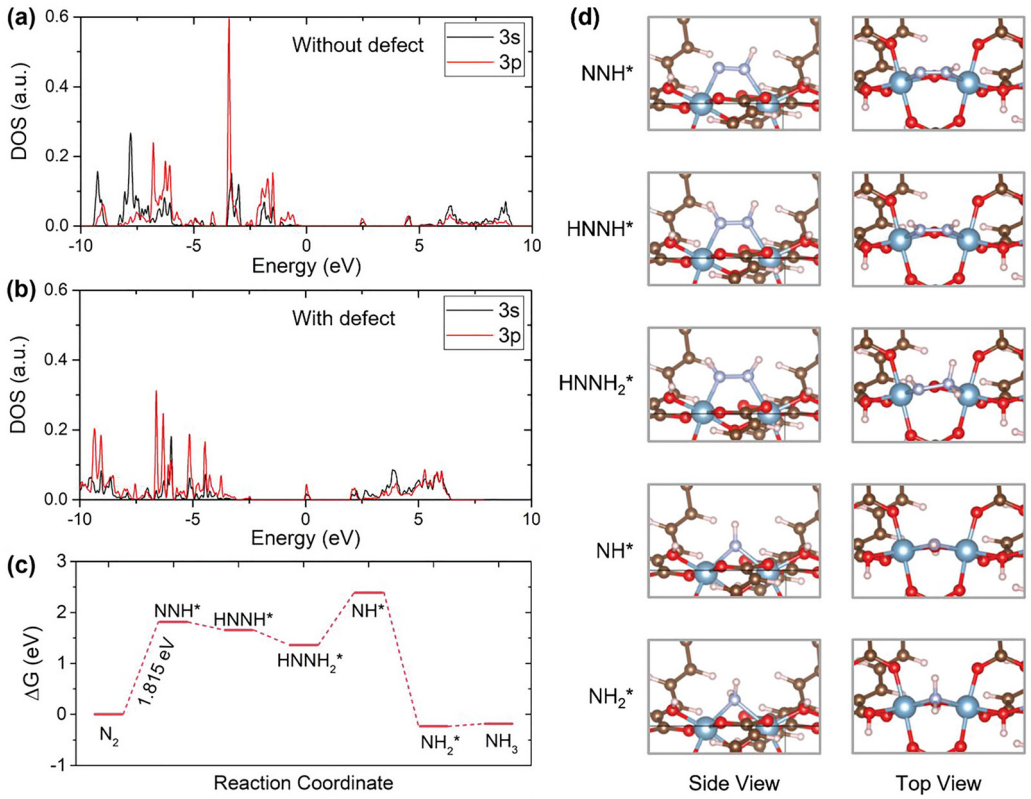

Fig. 4 DFT calculations. DOS of the Al atom for Al-Fum MOF (a) without and (b) with defects. (c) Free-energy diagrams for N_2 reduction on the defect site of Al-Fum at 0 V *versus* SHE. (d) Side and top views of atomic configurations for N_2 reduction reaction intermediates adsorbed on the defect site of Al-Fum. Reproduced with permission.[38] Copyright 2023, Wiley.

adsorbed intermediates. N-containing species, such as *NNH, *HNNH, and *$HNNH_2$, bridge two adjacent defective Al centers, while *NH and *NH_2 bind to a bridging position between the same pair of Al atoms. These configurations demonstrate how pairs of defective aluminum sites, coordinated by surrounding oxygen atoms, can stabilize key intermediates cooperatively and support the continuous formation of N–H bonds along the N_2RR pathway. Overall, this study provides compelling evidence that linker-extension-induced defect engineering is an effective method of activating main-group MOFs for the N_2RR. However, it also emphasizes the necessity of enhancing conductivity and achieving greater control over defect chemistry to maximize the potential of this approach.

In terms of the conductivity enhancement, Zhu *et al.*[39] presented a structural strategy focused on conductivity, in which ultrafine metallic Cu clusters (approximately 1.5 nm in size) are generated *in situ* and tightly confined within the mesoporous channels of the conductive CuHHTP MOF. This forms a Cu@CuHHTP hybrid that maximizes electronic transport while preventing cluster aggregation. The catalytic mechanism is governed by metallic Cu sites, where a higher d-band center and an 'accept–donate' charge-transfer pathway on Cu(111) surfaces enhance the adsorption and activation of NO_3^- intermediates. This lowers the overall reaction barriers relative to pristine MOF sites. Consequently, the catalyst exhibits impressive ammonia-synthesis metrics from nitrate reduction, including 85.81% nitrate conversion, 96.84% NH_3 selectivity, an NH_3 yield of 1.84 mg h^{-1} cm^{-2} and a Faradaic efficiency of 67.55%. This work proved the pore-filling approach successfully couples MOF conductivity with highly active Cu sites.

2.4. 2D MOF and MOF nanosheets

2D MOFs and MOF-derived nanosheets have emerged as a particularly powerful structural platform for electrochemical ammonia synthesis from nitrogenous pollutants due to their ultrathin architectures, maximized exposure of active metal sites, and intrinsically shortened charge-transport pathways. Their planar morphology eliminates diffusion bottlenecks and facilitates rapid access of N_2, NO_x^-, or NO molecules to catalytic centers, while the high ratio of unsaturated surface atoms enhances adsorption and activation of key intermediates. Collectively, 2D MOFs offer a structurally precise and catalytically advantageous platform for high-performance ammonia electrosynthesis, particularly in nitrate/nitrite reduction where adsorption strength and charge-transfer efficiency are rate-determining.

For example, Zhu *et al.*[40] investigated the use of a 2D Co-MOF nanosheet as a scaffold for embedding *p*-mercaptobenzoic acid-protected iridium nanoclusters (Ir NCs@Co-MOF) in order to facilitate the electrocatalytic reduction of nitrate to ammonia. Fig. 5(a) illustrates the solvothermal synthesis of Ir NCs@Co-MOF on Ni foam. Co^{2+}, terephthalic acid (TPA) and p-MBA-protected iridium nanoclusters assemble, whereby the –SH

group binds to the iridium and the terminal –COOH groups coordinate with the cobalt, enabling the iridium nanoclusters to be chemically integrated into the growing Co-MOF nanosheets. Fig. 1(b) shows an SEM image of the pure Co-MOF, which displays dense, layered 2D nanosheets that cover the Ni foam uniformly. After the incorporation of the Ir NCs (Fig. 1(c)), the electrode retains the same nanosheet morphology. This indicates that the embedding of the Ir NCs does not disrupt the 2D Co-MOF structure. TEM imaging of the Ir NCs@Co-MOF composite (Fig. 1(d)) confirms that the Ir nanoclusters, which have an average size of approximately 1.8 nm, are finely and uniformly dispersed on the Co-MOF. The lattice spacing is 0.225 nm, which corresponds to the Ir(111) plane. Elemental mapping (Fig. 1(e)) further demonstrates the homogeneous distribution of C, O, Co, and Ir throughout the nanosheets, verifying the successful embedding of the low loading of Ir (0.39 at%) into the 2D MOF framework. Electrochemically, Ir NCs@Co-MOF exhibits strong NO_3^--to-NH_3 performance in mildly saline electrolyte (0.1 M Na_2SO_4, 50 mg L^{-1} NO_3^-–N). At −0.8 V *vs.* RHE, the catalyst achieves a nitrate conversion rate of 92.5%, an NH_4^+ selectivity of 81.4%, and an NH_4^+ yield of 230.1 μg h^{-1} cm^{-2}, with Faradaic efficiency peaking at around 31.4%. It is reported that this design maximizes exposure of active sites, enhances electron transfer *via* Ir, and stabilizes the nanoclusters against aggregation, thereby improving NO_3^--to-NH_3 conversion under ambient conditions.

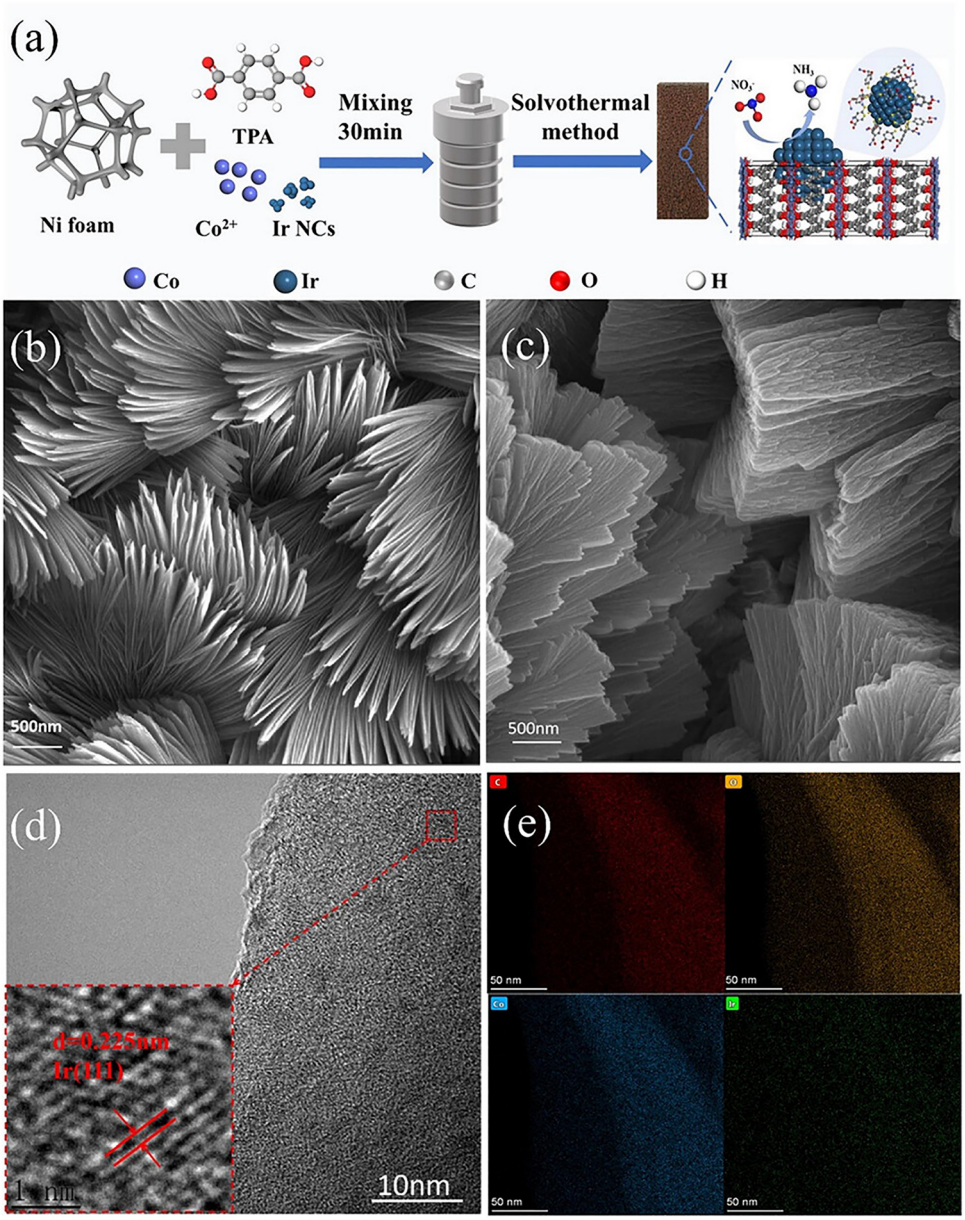

Fig. 5 Synthesis process and morphology characterization. (a) Schematic illustration of the synthesis of Ir NCs@Co-MOF *in situ* grown on Ni foam by a solvothermal method. (b) SEM image of Co-MOF. (c) SEM image of Ir NCs@Co-MOF. (d) TEM of Ir NCs@Co-MOF. (e) Element mapping images of Ir NCs@Co-MOF. Reproduced with permission.[40] Copyright 2025, American Chemical Society.

In another study, Pan *et al.*[41] employed a 2D, ultrathin Ni-MOF nanosheet strategy to enhance the electrochemical production of ammonia from nitrate. This strategy exploits the inherently high surface exposure, the abundance of coordinatively unsaturated nickel sites, and the improved electron transport properties of 2D MOFs. Combining experimental analysis and DFT calculations, mechanistic studies reveal that hydroxyl-ligated Ni sites possess stronger nitrate adsorption and lower activation barriers than carboxyl-ligated ones. This enables a more favorable $NO_3 \rightarrow NO_2 \rightarrow NO \rightarrow NH_x$ hydrogenation pathway. Under −1.4 V, the Ni-MOF nanosheets achieve 96.4% nitrate conversion, ∼80% NH_3 selectivity and an NH_3 yield of 110.13 μg $h^{-1}cm^{-2}$, outperforming pristine Ni foam. This nanosheet-based structural design is effective because thinning the MOF to a few nanometers maximizes the exposure of catalytically active Ni centers, shortens the diffusion lengths of electrons and ions, and modulates the local microenvironment to stabilize key intermediates.

2.5. Ligand and topology engineering

Ligand and topology engineering are two highly effective strategies for tailoring MOF electrocatalysts to enhance their catalytic performance.[42,43] Unlike metal-node modification, which primarily adjusts the intrinsic activity of catalytic centers, ligand and topology engineering alter the electronic structure, mass transport pathways and adsorption microenvironment of the framework.[44] This influences the reaction kinetics and selectivity at the system level. Ligand engineering involves modifying the organic linkers to adjust the coordination environment and electronic properties of the metal nodes.[45] Functional groups such as $–NH_2$, –OH and $–SO_3H$, as well as heterocyclic moieties, can introduce electron-donating or electron-withdrawing effects, regulate local charge density and stabilize key intermediates. For example, Zhang *et al.*[46] explored the fundamental roles of functionalized ligands in Pt@MOF catalysts. This work presents systematic studies on Pt@UiO-66 derivatives bearing $–NH_2$ and $–SO_3H$ groups. These studies demonstrate that functional groups anchored on organic linkers significantly increase steric hindrance within MOF channels, which enhanced steric confinement and restricts substrate orientation and diffusion, thereby improving shape- and site-selective catalysis. It is reported that ligand functionalization introduces multi-pathway electron-transfer channels among metal nanoparticles, metal nodes and organic linkers. XPS analysis revealed that $–NH_2$ groups promote electron withdrawal from Pt nanoparticles, leading to increased Pt binding energies. In contrast, $–SO_3H$ groups induce reverse electron transfer, enriching electron density on Pt. Taken together, these findings demonstrate that ligand functionalization reshapes the physical confinement and electronic landscape of MOF catalysts, enabling precise control over the activity–selectivity balance and providing a general design principle for high-performance MOF-based electrocatalysts.[46]

Topology engineering, on the other hand, involves altering the global connectivity and pore network of MOFs, including framework dimensionality, channel orientation, and connectivity density.[47,48] For example, Jiang *et al.*[49] reported the topology functionalization of a robust chiral Zr-MOF for enantioselective hydrogenation. In this study, reticular and modulated synthesis strategies were employed to generate frameworks with either FLU or ITH network topologies from identical tetratopic biphenol-derived linkers. Despite being constructed from the same building blocks, the two topologies exhibited strikingly different catalytic performance, which highlights the decisive role of framework architecture in determining reactivity and selectivity. In detail, the flu topology was found to feature large octahedral cages interconnected by open one-dimensional channels, with the dihydroxy groups of the biphenol ligands oriented towards spacious cavities. This configuration enables the successful post-synthetic grafting of monophosphoramidite ligands and the subsequent immobilization of iridium (Ir) complexes. This yields highly active, recyclable heterogeneous catalysts for asymmetric hydrogenation with up to 98% enantiomeric excess. In contrast, the *i*th topology features triple-helicate cages with significantly smaller diameters and more confined triangular channels. These findings demonstrate that topology engineering is a crucial control parameter for catalytic performance, not just a structural design parameter. Together, these structural strategies show that MOFs can be custom-designed reactors where atomic structure and electrochemical function are seamlessly integrated. Although these structural strategies offer an essential foundation for understanding how MOF architectures influence catalytic performance, the rapid integration of ML provides a powerful means of accelerating their design. This is achieved by predicting structure–property relationships and guiding the discovery of next-generation MOFs for efficient electrocatalysis.

3. ML fundamentals in MOF research

ML methodologies have broad applicability across many MOF research domains. In this review, however, we focus specifically on their role in guiding the design and optimization of MOF-based electrocatalysts for sustainable ammonia production. In particular, ML is increasingly being used to predict key performance indicators of the EN_2RR and ENO_3RR, such as Faradaic efficiency, NH_3 yield rate, selectivity against the HER and catalyst durability. This directly bridges computational intelligence with experimental ammonia synthesis.

3.1. Overview of ML techniques

Traditional computational methods such as Grand Canonical Monte Carlo (GCMC),[50] Molecular Dynamics (MD),[51] and Density Functional Theory (DFT)[52] have long been used to predict the performance of MOFs in gas adsorption, separation, and stability applications. While these methods provide accurate insight, they are computationally intensive and limited in scalability. High-throughput computational screening (HTCS) addresses these limitations by enabling rapid evaluation of large MOF libraries, significantly accelerating material discovery.[53]

The integration of ML with HTCS further enhances this process by enabling predictive modeling and pattern

recognition, thereby reducing the reliance on exhaustive simulations.[54] Recent advances, particularly in Graph Neural Networks (GNNs), have revolutionized the screening of MOF databases by capturing complex structural and chemical relationships.[55] This approach allows rapid identification of MOFs with tailored properties for specific applications.

In addition, the synergy between ML and automated experimental platforms has transformed high-throughput MOF research. These platforms seamlessly integrate ML-guided synthesis with experimental validation, closing the loop between computational predictions and laboratory results.[56] This iterative cycle accelerates the development of advanced materials and bridges the gap between theory and practice.

Supervised learning is one of the most widely used ML paradigms in MOF research. It involves training models on labeled data sets to predict specific outcomes, such as gas adsorption capacities, catalytic efficiencies, or electronic properties. This paradigm is particularly useful for performance prediction and optimization tasks.[57] Common models used in supervised learning include Support Vector Machines (SVM),[58] Random Forests (RF),[59] Decision Trees (DT),[60] Gradient Boosting Machines (GBM),[61] Artificial Neural Networks (ANN),[62] k-Nearest Neighbors (k-NN)[63] and Bayesian algorithms such as Bayesian Networks (BN).[64] These algorithms excel at identifying structure–property relationships and optimizing synthesis conditions to improve material performance.

Unsupervised learning, on the other hand, focuses on discovering hidden patterns in data without requiring labeled outputs.[65] This approach is ideal for exploratory tasks, such as clustering MOFs based on structural similarities or identifying trends in performance data sets. Popular unsupervised learning models include K-means,[66] hierarchical clustering,[67] Gaussian mixture models (GMM),[68] and density-based clustering (DBSCAN).[69] In addition, dimensionality reduction algorithms such as principal component analysis (PCA),[70] t-distributed stochastic neighbor embedding (t-SNE),[71] and autoencoders[72] can be used to summarize the data by describing it using less information, which is helpful in both supervised and unsupervised learning. These techniques are instrumental in understanding the underlying relationships within complex MOF datasets and generating new hypotheses.

In the field of MOF research for sustainable ammonia production, both supervised and unsupervised learning play complementary yet fundamentally different roles. Supervised learning is primarily used for performance prediction. Regression models such as Random Forest or gradient boosting, for example, are trained using descriptors including metal-node electronic properties, defect density and pore geometry, to predict Faradaic efficiency, NH_3 yield rate and selectivity over the HER for EN_2RR or ENO_3RR systems. These models allow for the rapid virtual screening of thousands of hypothetical MOFs, enabling the identification of promising candidates prior to experimental validation.

In contrast, unsupervised learning is mainly used for structural discovery and pattern recognition. Clustering algorithms and dimensionality reduction techniques, such as k-means or t-SNE, can group MOFs based on latent structural similarities. This reveals hidden correlations between framework topology, functional group distribution and catalytic behaviour, without the need for labelled performance data. Such analyses help to uncover previously overlooked structure–activity trends and guide the exploration of new MOF families with desirable catalytic microenvironments.

Semi-supervised learning bridges the gap between supervised and unsupervised methods by using both labeled and unlabeled data.[73] This technique is particularly advantageous in MOF research, where labeled data (*e.g.*, experimental adsorption capacities) can be scarce and expensive to obtain. Models such as self-training[74] and co-training[75] are commonly used. In addition, transfer learning enables the use of knowledge gained from one prediction task (*e.g.*, MOFs containing metals A, B, and C) to improve model performance on a related but different task (*e.g.*, MOFs containing metals D, E, and F), thereby enhancing generalization and reducing the need for extensive new training data.[76] Multi-task learning, on the other hand, facilitates the simultaneous optimization of multiple interrelated goals by leveraging shared representations.[77] For example, in MOF-based electrocatalysis, a model could be trained to predict both high Faradaic efficiency (FE) and high NH_3 yield (task 1), while simultaneously enforcing constraints related to structural topology (task 2), ultimately leading to a more holistic and efficient design strategy. The integration of these techniques could significantly improve the predictive capabilities of ML models in MOF research, especially for applications that require trade-offs between multiple performance criteria. These methods improve the predictive power of machine learning models while reducing the dependence on large, labeled datasets.

Reinforcement learning (RL) represents a dynamic approach where models learn optimal actions through iterative interactions with an environment.[78] In MOF research, RL is applied to guide the discovery of synthesis pathways, optimize material designs, and improve decision-making in experimental setups. Key algorithms in reinforcement learning include Q-Learning,[79] Deep Q-Networks (DQN),[80] Policy Gradient Methods,[81] Actor-Critic Methods,[82] and Monte Carlo Tree Search (MCTS).[83] These algorithms enable the systematic exploration of vast design spaces and facilitate the identification of optimal solutions for complex problems.

3.2. ML performance metrics

Evaluating the performance of machine learning models in MOF research requires robust metrics to ensure reliability and accuracy of predictions. Common performance metrics include

Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of determination (R^2), and Pearson entropy. MAE measures the average size of errors in predictions and provides a simple interpretation of prediction accuracy.[84] It is calculated as the average of the absolute differences between predicted and observed values ($\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$, where y_i and $\hat{y}_i$ are the true and predicted

values, respectively, and n is the number of samples). MSE quantifies the mean squared difference between predicted and actual values $\left(\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)$. It serves as the primary metric for assessing prediction accuracy, with smaller MSE values indicating higher accuracy. By penalizing larger errors more severely, MSE provides a robust measure for error analysis, making it particularly effective in identifying models with significant outliers.[85] On the other hand, RMSE provides insight into the standard deviation of prediction errors, penalizing larger errors more than smaller ones $\left(\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\right)$.[86] This metric is particularly useful for evaluating models where large errors are critical. In addition, R^2, also known as the coefficient of determination $\left(R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}\right)$ (where $\bar{y}$ is the mean of the observed values), measures the proportion of variance in the dependent variable that is explained by the model.[87] Ranging from 0 to 1, R^2 values closer to 1 indicate that the model captures a substantial amount of variance, reflecting its predictive power. This metric is essential for understanding the explanatory power of a model and determining its fitness to the data. Furthermore, the Pearson correlation coefficient evaluates the linear relationship between two variables and ranges from -1 to 1. A coefficient closer to 1 or -1 indicates a strong positive or negative correlation, respectively, while a value closer to 0 indicates a weak or no linear correlation.[88] This metric is critical in determining the strength and direction of relationships within data, providing insight into underlying patterns. In summary, MAE, MSE and RMSE follow different principles in error evaluation, and their selection depends on the characteristics of the data set, especially the presence of outliers. MAE is less sensitive to extreme values, making it more suitable for datasets with many outliers, while also being easier to interpret. In contrast, MSE and RMSE are preferable when the dataset contains few or no outliers, as they penalize larger errors more severely, thereby emphasizing substantial deviations. Among the two, RMSE is often preferred for interpretability due to its consistency with the original data units. In addition, Mean Absolute Percentage Error (MAPE), although less commonly used in the field, is usually a good metric for summarizing regression predictions as a percentage, which can be useful for explaining the performance of the model to non-specialists (thanks to its high interpretability value). These error metrics provide different insights, with lower error values indicating higher predictive accuracy. When used together, these metrics provide a comprehensive framework for evaluating and improving predictive models, ensuring their robustness, accuracy, and applicability in MOF research contexts.

3.3. Data acquisition and curation

In the field of MOF research, the acquisition and curation of high-quality data is critical to advancing materials discovery and application. Publicly available databases serve as foundational resources, providing rich datasets that facilitate computational and experimental investigations.[89] Notable examples include (i) hMOF.[90] Contains 137 953 hypothetical MOFs generated from a library of building blocks, with data on properties such as surface area, pore size distribution, and methane storage capacity. (ii) CoRE MOF.[91] Contains 14 142 computationally ready, experimentally derived MOFs designed for simulations. This database is widely used for various computational studies. (iii) Cambridge Structural Database (CSD) MOF Subset.[92] Provides 69 666 MOFs from the CCDC with calculated properties such as pore volume, surface area, and density. (iv) Boyd Materials Cloud.[93] Includes 324 426 hypothetical MOFs screened for CO_2/N_2 selectivity and work capacity, ensuring accurate electrostatic potential representation. (v) Quantum MOF (QMOF).[94] Provides calculated quantum chemical properties for 15 713 experimentally synthesized MOFs, including band gaps, energies, and density of states. (vi) Open DAC Dataset (ODAC).[95] Includes extensive DFT calculations for over 8412 MOF materials targeted for direct air capture applications. (vii) ARC-MOF.[96] A diverse database of 279, 610 MOFs with detailed geometric, physical, and chemical properties derived from multiple sources. (viii) DigiMOF.[97] Generated using a chemistry-aware natural language processing tool that captures over 43 000 MOF-related journal articles and 52 680 associated properties such as synthesis methods and topologies.

In summary, MOF data could be obtained from four main sources: materials databases, scientific literature, theoretical calculations, and experimental results. These databases aggregate data from various sources, including experimental measurements, computational simulations, and scientific literature, providing a comprehensive foundation for ML applications in MOF research. The quality of the data from these repositories is paramount. Experimental data, such as adsorption capacities and catalytic performances, provide empirical insights but can vary due to differences in measurement conditions and techniques.[98] Computational simulation data, including DFT calculations, provide theoretical predictions that are consistent within defined parameters, but may not always capture real-world complexities. Therefore, it is essential to ensure data quality through careful curation. This process includes data cleaning to remove inaccuracies, standardization to ensure consistency across datasets, and validation against known benchmarks. Such rigorous curation increases the reliability of ML models trained on these datasets, resulting in more accurate and generalizable predictions.

3.4. Descriptors and feature engineering and selection

In the study of MOF design and optimization, the selection and extraction of appropriate descriptors is essential to create a simple yet comprehensive representation of MOF structures.[99] Key features such as pore size, surface functionalization, and node-ligand relationships are critical for understanding and predicting MOF performance.[100] Gross-level descriptors, including geometric properties such as void fraction, and volumetric surface area, are widely used due to their reliability

and strong correlation with adsorption and separation performance.[101] At the molecular fragment level, descriptors focus on specific components such as ligands or secondary building units (SBUs) which provide insight into the chemical environment, while subatomic level descriptors, including quantum mechanical properties such as electron density and charge distribution, offer high accuracy at low computational cost. In addition, experimental descriptors derived from synthesis and characterization data integrate real-world performance metrics into ML models.[99]

In practice, the construction of descriptors for MOFs should follow a hierarchical strategy that starts with physically meaningful building blocks and progressively incorporates data-driven screening. At the atomic level, descriptors can be derived from elemental properties, such as electronegativity, oxidation state, d-band filling and coordination number, to capture the intrinsic activity of metal nodes and ligand environments. At the pore level, geometric descriptors such as pore-limiting diameter, accessible surface area, void fraction and pore connectivity encode the effects of mass transport and confinement that govern adsorption and catalytic kinetics. At the topology level, graph-based descriptors reflecting node degree, ring statistics and framework connectivity provide a compact representation of long-range structural organization.

These descriptors, driven by domain knowledge, should be then complemented by data-driven feature engineering. Model-agnostic feature importance analysis (*e.g.* RF importance or SHAP values) and dimensionality reduction methods (*e.g.* PCA or autoencoders) can be employed to identify redundant or weakly informative variables. Additionally, graph neural networks facilitate the direct learning of latent structural descriptors from atomic graphs, effectively capturing intricate interactions that are challenging to encode manually. By combining physical intuition with statistical screening iteratively, a reduced set of interpretable and predictive descriptors can be constructed. This improves model generalizability and enables rational, performance-oriented MOF design.

Feature importance analysis, performed using models such as RF or GB, identifies the most influential descriptors for predicting MOF properties.[99] These two models are particularly well-suited to tabular data. However, when working with high-dimensional data, such as structured data, more advanced models are required, such as GNNs. Another useful procedure in feature engineering is dimensionality reduction (PCA, t-SNE and autoencoders), as this removes duplicated or non-informative features. These analyses often reveal that geometric descriptors such as pore size and void fraction dominate adsorption predictions, while chemical and energy descriptors play complementary roles, especially under variable conditions.[102] Understanding these relationships guides rational MOF design by prioritizing structural features that directly influence desired properties.

Feature engineering plays a vital role in bridging the gap between raw MOF data and high-performance ML models. It can be divided into three sequential stages: feature construction,[103] feature transformation,[104] and feature selection.[103] The first stage, feature construction,[105] involves generating new descriptors from raw inputs. For example, elemental properties of metal nodes can be combined with coordination numbers to derive effective catalytic site descriptors. Alternatively, pore size distribution, void fraction, and surface chemistry can be aggregated to describe mass-transport behavior relevant to electrocatalysis. The aim of feature transformation is to normalize or rescale heterogeneous descriptors (*e.g.* *z*-score normalization of electronic properties[106]) to reduce bias during model training. Dimensionality reduction techniques, such as principal component analysis or autoencoders, further compress correlated features into a compact latent space while preserving key structure–property relationships.

Feature selection then identifies the most informative subset of descriptors for model construction. This can be achieved using filter methods (*e.g.* correlation screening[107]), wrapper methods (*e.g.* recursive feature elimination[108]) or embedded methods based on model-specific importance scores, such as Random Forest feature importance or SHAP values. In MOF-based electrocatalysis, this process enables the isolation of the most relevant descriptors for ammonia production metrics (*e.g.* metal-node electronic configuration, defect density, and local coordination environment) while discarding redundant geometric variables. This structured feature engineering pipeline enables ML models to achieve improved accuracy, interpretability and generalization, thereby facilitating more reliable prediction and rational design of MOF electrocatalysts.

In addition, graph model-based feature representation further enhances the understanding of structure–property relationships in MOFs. By representing MOF structures as graphs, where atoms serve as nodes and chemical bonds as edges, advanced techniques such as GNNs capture intricate structural and chemical relationships.[101] This representation facilitates the extraction of meaningful features, such as connectivity patterns and node-ligand interactions, that are difficult to identify using traditional methods.[109] Through this graph-based approach, ML models can uncover correlations between structure and performance, enabling the design of MOFs optimized for specific applications. This integration of descriptor selection and graph modeling represents a powerful strategy for advancing MOF research.

3.5. Prediction of MOF properties using ML

ML enables accurate prediction of various MOF properties, supporting the rational design of materials for various applications. ML models predict mechanical and electrical conductivity by correlating structural features such as density, connectivity, and topology with performance metrics essential for electronic and sensing applications.[110,111] Adsorption capability is predicted using descriptors such as pore size, surface area, and functional groups,[112] with deep neural networks improving prediction accuracy. In addition, synthesis conditions, including temperature and solvent parameters, are optimized to streamline experimental reproducibility.[113] Stability and guest accessibility, critical for separation and catalysis, are modeled using metrics such as thermal resilience and pore

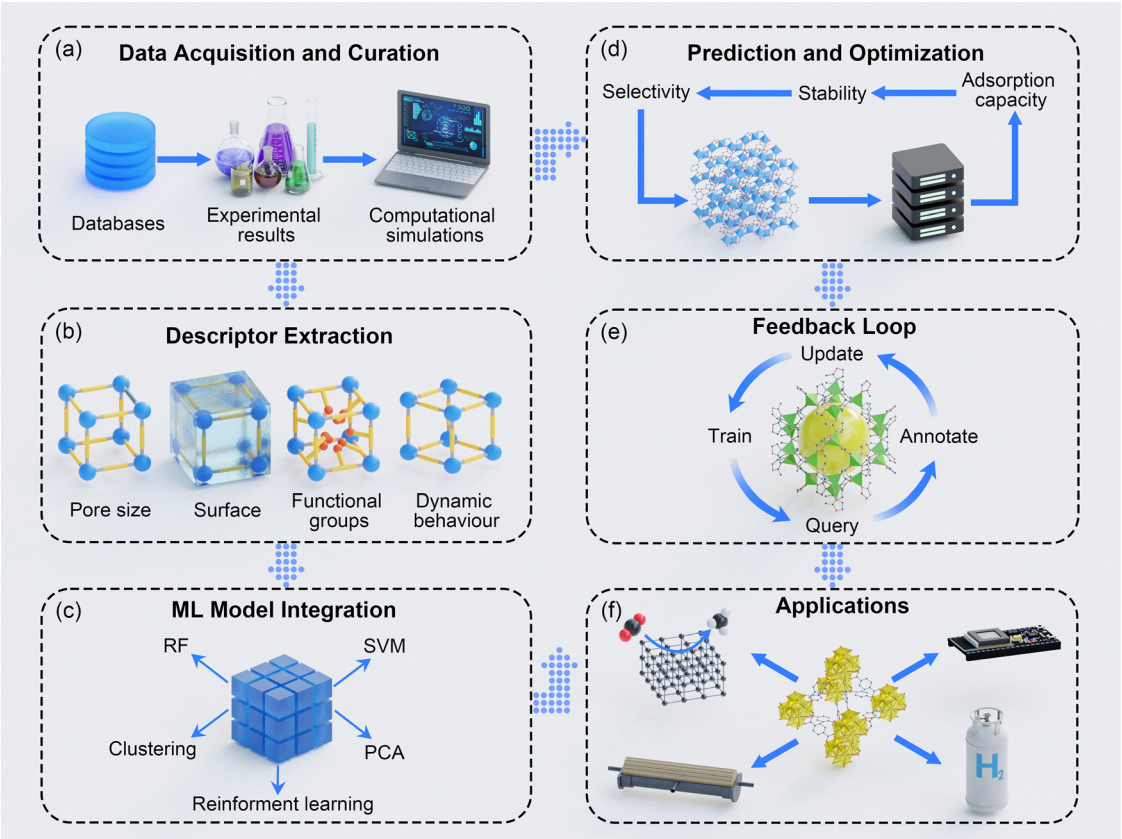

Fig. 6 Schematic of the workflow of ML in the design of MOFs for various applications. (a) Data acquisition and curation, (b) descriptor extraction, (c) ML model integration, (d) prediction and optimization, (e) feedback loop, and (f) applications.

energy descriptors.[113] Furthermore, structural properties, including scaffold flexibility,[114] elastic moduli,[115] and mechanical stability,[115] are predicted for high-pressure and high-temperature suitability. In addition, thermal transport properties, critical for heat exchange and thermal management, are modeled by linking scaffold topology to heat transfer efficiency.[116,117] Besides the properties mentioned above, morphology analysis predicts MOF crystal properties, such as size and shape, which affect material scalability. Advanced ML techniques, including GNNs, RF, and deep learning, enable accurate property modeling that accelerates material discovery and provides insight into structure–property relationships. Here, a schematic workflow of ML in the design of MOFs for various applications is shown in Fig. 6.

3.6. Linking ML fundamentals to applications

In the context of electrochemical ammonia synthesis, ML-assisted MOF design focuses on correlating structural and electronic descriptors with experimentally measurable metrics of NH_3 production, including Faradaic efficiency, NH_3 yield rate, turnover frequency and operational stability. For EN_2RR systems, descriptors such as the d-band centre of the metal node, the N_2 adsorption energy, the density of states near the Fermi level induced by defects, and the local coordination geometry are commonly used to train supervised learning models that predict catalytic activity and selectivity. For nitrate and nitrite reduction, ML models also incorporate descriptors relating to NO_3^-/NO_2^- adsorption strength, proton-coupled electron transfer energetics and suppression of competing HER pathways. This enables the targeted screening of MOF architectures with enhanced ammonia selectivity.

ML serves as a transformative tool to bridge fundamental research with practical applications in MOF design. Using techniques such as classification, regression, clustering, and optimization, ML enables a deeper understanding of structure–property relationships and drives the rational design of MOFs for diverse applications.[118] For example, ML-guided property predictions enable researchers to optimize MOFs for gas storage and separation, including membrane technique,[119–122] CO_2 capture,[123] adsorption heat transformation,[124–126] CH_4 adsorption,[127] hydrogen storage,[128] water harvesting[129,130] and gas separation.[131] Similarly, ML facilitates catalyst development by predicting reactivity and stability, while also supporting the design of MOFs for water treatment,[132] sensing,[133,134] adsorption heat transformation,[135,136] and drug delivery.[110] The theoretical foundation provided by ML methods integrates seamlessly with application-driven research, accelerating material discovery cycles and providing targeted solutions to real-world challenges. Techniques such as GNNs and deep learning have enabled accurate predictions and insights into key properties such as thermal conductivity, mechanical stability, and adsorption capacities, providing actionable data for performance optimization. Here, ML-assistance in MOF

design and synthesis for these applications is illustrated in Fig. 7. Some representative studies are elaborated in Section 3.6 (Linking ML fundamentals to applications) for comprehensive understanding. The intersection between fundamental ML techniques and application-oriented research sets the stage for Section 4 (Challenges and suggested directions), where we explore specific practical implementations of ML in MOF design.

Recently, these ML-guided MOFs showed remarkable potential in a wide range of emerging fields. For example, Wang *et al.*[137] applied ML to design 2D MOFs with high magnetic anisotropy energy (MAE) for magnetic storage. Using the Transition-Metal Interlink Neural Network (TMINN) model, it screened 1440 MOFs and predicted MAE for 2583 additional candidates, identifying 28 high-MAE structures, including 11 novel ferromagnetic MOFs validated by DFT. ML enabled efficient high-throughput screening and achieved accurate predictions with an RMSE of 11.36 meV per atom, demonstrating its potential to accelerate MOF discovery despite reduced accuracy for low-MAE structures. In another study, Wang *et al.*[138] focused on the use of ML to predict the drug loading capacity and cytotoxicity of MOFs, specifically UiO-66-NH_2 and MIL-125-NH_2, for biomedical drug delivery applications. In this study, ML plays a critical role in streamlining candidate selection, improving performance prediction, and reducing reliance on trial-and-error experimental methods. Furthermore, Anadebe *et al.*[139] investigated the application of Co-MOF (ZIF-67) as a corrosion inhibitor for X65 steel in CO_2-saturated environments, addressing critical industrial challenges in preventing steel degradation. The study achieves an impressive corrosion inhibition efficiency of up to 97%, which is attributed to the formation of a dense protective film on the steel surface. ML, specifically ANFIS, played a critical role in predicting inhibition performance, streamlining the analysis of experimental electrochemical data, and optimizing the protective properties of the material.

Specifically, ML-guided MOFs have been widely used in electrochemical catalysis (*e.g.*, CO_2 reduction[140] and nitrogen reduction reactions), photocatalysis (*e.g.*, hydrogen evolution

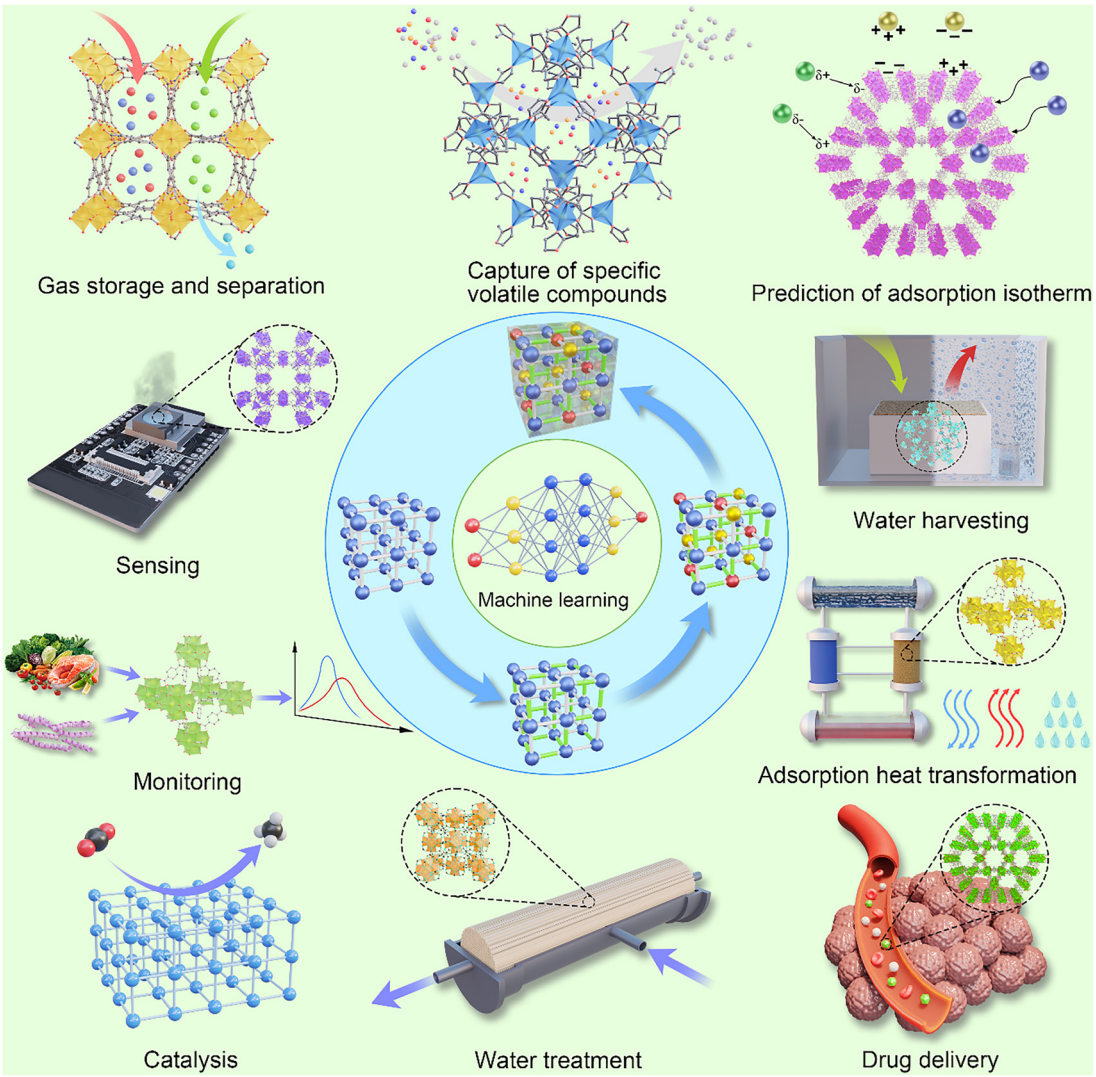

Fig. 7 Illustration of ML-assistance in MOF design and synthesis for versatile applications.

and pollutant degradation), and biocatalysis (*e.g.*, enzyme immobilization and drug synthesis).[38] These applications address critical global challenges such as clean energy production, greenhouse gas mitigation, and sustainable chemical manufacturing.[141] ML is accelerating advances in MOF-based catalysis by enabling the prediction of active sites, reaction pathways, and catalytic performance. ML models can identify structure–activity relationships, optimize reaction conditions, and design novel MOFs with improved catalytic efficiencies. In addition, ML facilitates the integration of MOFs into hybrid systems, such as mixed-matrix membranes and composite catalysts, enabling scalable and efficient solutions for industrial catalysis. By leveraging ML, researchers can unlock the full potential of MOFs for catalytic applications, driving advances in sustainable and innovative chemical technologies.

For example, McCullough *et al.*[142] used NU-1000 MOFs as catalytic supports and employed high-throughput experimentation (HTE) and data-driven algorithms to optimize catalytic performance for propyne dimerization. Fig. 9(a) shows an integrated experimental and ML framework for hexadiene yield optimization of over 23 catalytically active metals, totaling 652 experiments. The pie chart categorizes the distribution of metals tested, providing a clear overview of the experimental diversity. The methodology follows a dual model scheme: Principal Component Analysis (PCA) and *k*-nearest neighbors (KNN) are first used to optimize conditions for the Cu catalyst. XGBoost Random Forest Regression is then applied to predict catalytic performance across different metals, leveraging the insights gained from the Cu optimization. In addition, Fig. 8(a) also highlights the highest hexadiene yields achieved for each metal (represented by black circles) and their corresponding product selectivity. Cu shows the most significant improvement in hexadiene yield under optimized conditions, validating the ML-guided approach. In summary, Bayesian optimization guided experiments for different metals, with Cu-deposited NU-1000 showing significant yield improvements from 0.4% to 24.4% hexadiene. The comprehensive evaluation provides a robust pipeline for rational catalyst design using ML.

Another representative example was recently reported by Bai *et al.*[143] They investigated the HTS of MOFs as catalysts for CO_2 cycloaddition reactions using an explicable ML model. The MOFs include 12 415 hypothetical structures and 100 reported MOFs, with experimental validation of MOF-76(Y). Here, Fig. 8(b) evaluates the performance of 16 ML models for predicting CO_2 cycloaddition activity, focusing on metrics such as training score, test score, precision, recall, and F1 score calculated at the 75% quantile. Among these models, RF emerges as the most effective, achieving the highest scores across all metrics. The consistency between training and test scores highlights the model's ability to balance bias and variance, reducing overfitting while ensuring accurate predictions. Fig. 8(c) shows the structures of highly active MOFs identified by high-throughput screening, where Fig. 8(c.1)–(c.8) represent hypothetical MOFs generated by ToBaCCo calculations and Fig. 8(c.9) and (c.10) represent experimentally reported MOFs from the CoRE MOF 2019 database. The selected MOFs exhibit diverse structural and electronic features, including different pore sizes, surface areas, and metal–ligand combinations, which contribute to their catalytic efficiency. This work highlights the potential of combining data-driven methodologies with computational tools for the rapid and accurate discovery of efficient catalysts for CO_2 cycloaddition reactions. Further experimental validation of hypothetical MOFs could expand their applicability and address potential limitations of computational assumptions.

4. Challenges and suggested directions

4.1. Data limitations

Data limitation challenges mainly include: (i) data scarcity. Despite the growing number of MOF databases, the availability of high-quality datasets remains limited. Many datasets focus on specific properties such as adsorption or catalytic activity, leaving significant gaps in less explored properties such as mechanical stability, thermal transport, or electronic conductivity. In addition, experimental data, which are often more accurate than simulated data, are scarce due to the costly and time-consuming nature of MOF synthesis and testing. (ii) Inconsistent quality. The lack of uniformity in data collection methods further complicates ML applications in MOF research. Variations in experimental conditions, simulation parameters, and reporting standards lead to inconsistencies that reduce the robustness of ML models. For example, adsorption measurements can vary significantly depending on the temperature, pressure, and purity of the gases used, which are often not reported consistently across data sets. (iii) Lack of standardization. The lack of standardized protocols for data curation and sharing creates barriers to integrating datasets from multiple sources. Differences in data formats, descriptor definitions, and units of measurement make it difficult to harmonize datasets, limiting their usability for training generalized ML models.

To overcome these limitations and improve the reliability of ML models, several strategies can be implemented: (i) data augmentation. Techniques such as mix-up,[144] which generate new training samples by interpolating between existing data points, can improve model generalization and robustness. This method introduces controlled variability while preserving meaningful relationships in the data. (ii) Data synthesis. Generative models such as Generative Adversarial Networks (GANs),[145] Autoencoders (AEs),[145] Variational Autoencoders (VAEs),[146] and Diffusion Models[147] can be used to generate synthetic MOF data by introducing controlled noise. These models enable the creation of diverse yet realistic MOF structures and properties, filling gaps in underrepresented data distributions. Although generative models are commonly introduced as tools for data augmentation, they have a more transformative capability in inverse materials design. In the context of MOF-based ammonia electrocatalysts, variational autoencoders, generative adversarial networks and graph-

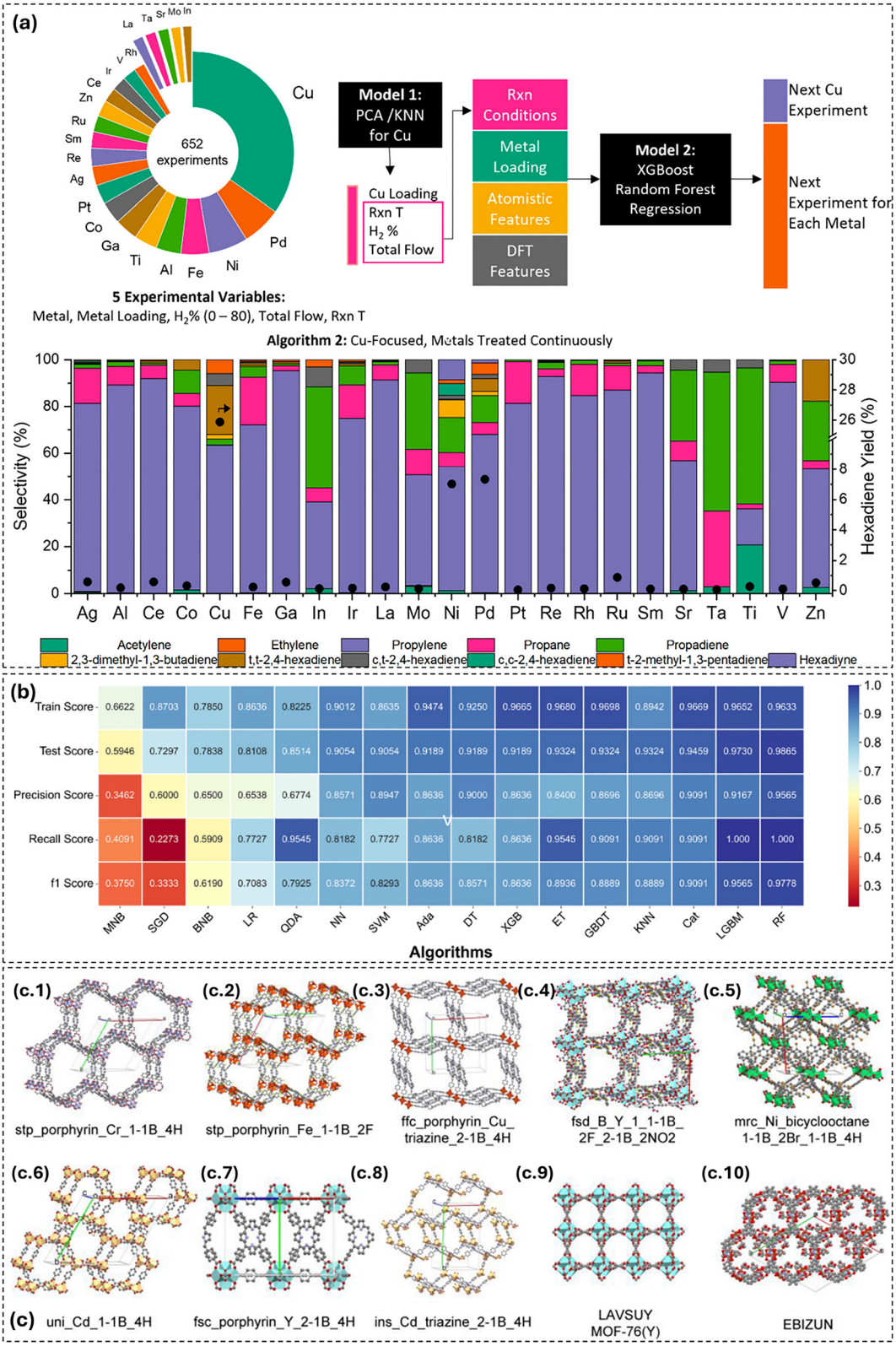

Fig. 8 (a) Summary of the second campaign to optimize hexadiene yield over 23 different catalytically active metals. Reproduced with permission.[142] Copyright 2023, American Chemical Society. (b) Evaluation of 16 ML models (75% quantile) with training score, test score, precision score, recall score and f1 score. (c) Structures of representative highly active MOFs predicted by HTS of 12 515 MOFs. Reproduced with permission.[143] Copyright 2025, KeAi.

based generative models can be trained using existing MOF databases that are annotated with NH_3 production metrics, to learn the latent structure–property manifold. Once trained, these models can generate new MOF candidates directly, based on desired performance targets such as high Faradaic efficiency, strong suppression of the HER or enhanced structural stability. In practical workflows, the generated hypothetical MOF structures are first filtered using rapid ML predictors for key descriptors, followed by high-throughput DFT screening to validate the adsorption energetics of N_2, NO_3^- and reaction intermediates. This closed-loop generative–predictive framework transforms ML from a passive prediction tool into an active discovery engine, enabling the rational exploration of vast chemical spaces that would be inaccessible through trial-and-error experimentation alone. (iii) Collaborative databases. Establishing open-access, standardized repositories that integrate experimental and computational data from multiple research groups will improve data accessibility and interoperability. Transfer learning techniques, in particular domain adaptation, can be used to enhance the utility of these databases. Domain adaptation allows models trained in data-rich domains to be adapted for use in data-poor domains, provided there is a meaningful relationship between the two. This approach allows insights from well-characterized MOF properties (*e.g.*, adsorption data) to inform predictions for underrepresented properties (*e.g.*, electronic conductivity or mechanical stability). Incorporation of the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles will further facilitate seamless data integration, improve ML model performance and accelerate materials discovery. (iv) Quality control measures. Rigorous data validation protocols, including systematic comparisons between experimental and computational data, can ensure consistency and reliability. The use of anomaly detection algorithms can help identify outliers and reduce noise in training data sets, resulting in more robust ML models. (v) Standardization initiatives. The development of community-driven standards for data collection, reporting, and sharing is essential for seamless dataset integration. Establishing consistent descriptor definitions, measurement protocols, and data formats will facilitate greater interoperability and improve model generalizability.

4.2. Model interpretability

The interpretability of ML models is a major challenge in MOF research, especially for complex architectures such as DNNs. These models often operate as ''black boxes'', making it difficult to understand how specific input features influence the predicted results, especially when capturing structure–property relationships in MOFs. One of the main challenges is to disentangle the contributions of different structural descriptors, such as pore size, surface functionalization, and node-ligand interactions,[148] to the predicted properties. For example, while a DNN may accurately predict gas adsorption capacities, it may not provide insight into which structural features drive performance. This lack of transparency hinders the rational design of MOFs and reduces confidence in ML-guided discoveries.

Several solutions have been proposed to address these challenges. One promising approach is to incorporate graph attention mechanisms into GNNs. These mechanisms assign varying degrees of importance to different nodes (atoms) and edges (chemical bonds) in the graph representation of MOFs, providing a clearer picture of the key structural features that influence model predictions. By highlighting critical interactions within the MOF framework, graph attention mechanisms improve interpretability and guide researchers in identifying design rules for targeted applications. Another promising approach to improve interpretability is knowledge distillation,[149] a transfer learning technique in which a complex deep learning model (*e.g.*, DNN) serves as a teacher to train a more interpretable model, such as a DT. While DT are limited in their ability to fully capture the complex knowledge embedded in a DNN, they can extract simplified representations that provide valuable insights into key decision patterns.[150] Although this method does not provide a complete explanation of the model's reasoning, it increases transparency by identifying influential structural features and trends, thereby aiding in the rational design of MOFs. In addition, techniques such as SHAP and Local Interpretable Model-agnostic Explanations (LIME) can be applied to *post hoc* analysis of ML models. These methods decompose predictions into the contributions of individual descriptors, providing insight into the underlying decision process of the model. Visualization tools, such as feature importance plots and attention heatmaps, further aid in understanding the relationship between MOF structures and their properties.

4.3. Transferability and generalization

Transferability and generalization of ML models remain significant challenges in MOF research, especially when applying models across different MOF systems. The ability of an ML model to accurately predict properties for novel or unseen MOFs is critical for advancing the field but is often limited by the variety and size of the training data. First, many ML models in MOF research are developed and trained on specific datasets, resulting in high performance within the confines of those datasets but poor generalization to new MOF chemistries, topologies, or application domains. For example, a model trained on datasets optimized for hydrogen storage may underperform when applied to predict properties relevant to CO_2 capture or catalytic activity. This lack of transferability is often due to model overfitting specific features or biases inherent in the training data. Second, the vast diversity of MOF structures, including variations in pore size, functional groups, and metal centers, poses additional challenges to model generalization.[151] Capturing this diversity requires data sets that represent the full range of MOF chemistries, which are often unavailable or incomplete.

To improve the generalization capability of ML models, four strategies are proposed for future improvement: (i) transfer learning. Transfer learning encompasses several techniques designed to improve model performance by leveraging knowledge from related tasks. One approach is domain adaptation, where a model is initially trained on a dataset with many

samples (source domain) and simultaneously adapted to a dataset with fewer samples (target domain) to improve generalization in data-scarce scenarios. Another widely used method is fine-tuning, in which a pre-trained model is not developed from scratch but refined by adjusting specific parameters to adapt it to a new but related task. These transfer learning strategies enable more efficient knowledge reuse, reducing the need for extensive new training data while improving model robustness and applicability to diverse MOF systems. (ii) Available data. Expanding datasets to include a broader range of MOF structures and properties to better represent the chemical and structural diversity in MOF systems. Synthetic data generation through high throughput simulations can also complement experimental data sets. (iii) Model regularization. Incorporation of techniques such as dropout, L1 (Lasso regularization) and L2 (Ridge regularization), and ensemble learning to prevent overfitting and improve robustness across different data sets. (iv) Benchmarking and validation. Regularly testing models on independent datasets or using cross-validation techniques to ensure their applicability to new MOF systems. This includes benchmarking against experimental results or datasets from different sources.

4.4. Integration challenges

The integration of ML with experimental studies in MOF research presents several challenges, largely due to the inherent variability and complexity of experimental workflows. A major issue is the presence of experimental errors (which is called ''noise'' in ML terms), which can undermine the accuracy of ML predictions. Variations in synthesis conditions, measurement techniques, and environmental factors often lead to discrepancies between predicted and observed results. For example, adsorption capacity measurements can vary significantly depending on gas purity, pressure, or temperature, which can be inconsistent with the assumptions made in ML models. In addition, data variability due to different MOF structures, synthesis methods, and test protocols makes it difficult to generate consistent data sets for model training. The lack of standardized protocols for reporting experimental results exacerbates this problem, further reducing the reliability and generalizability of ML models. Finally, the process of validating ML predictions through experiments is resource-intensive, requiring significant time and effort to synthesize and characterize new MOFs, often leaving a gap between theoretical predictions and real-world feasibility.

To address these integration challenges, several strategies can be implemented to bridge the gap between ML predictions and experimental workflows. Developing robust ML models that account for experimental variability, such as incorporating uncertainty quantification and error correction mechanisms, can improve prediction reliability (also known as ''robust against noise''). Establishing standardized protocols for data collection, reporting, and validation can reduce variability and improve consistency, allowing for better agreement between experimental results and ML predictions. High-throughput experimental setups integrated with ML-guided predictions can accelerate the validation process by testing multiple MOFs under controlled conditions. Iterative learning loops, where experimental results are fed back into the ML model for refinement, can ensure continuous improvement in prediction accuracy. Improving the interpretability of ML models using tools such as SHAP scores can provide actionable insights into predictions, reducing trial-and-error efforts in experimental design. By implementing these strategies, the synergy between ML and experimental studies can be strengthened, leading to faster and more reliable progress in MOF research.

4.5. Energy and computational costs

The application of ML in MOF research often requires significant computational resources, especially when using deep learning techniques and high-throughput simulations. In general, training ML models involves extensive computation, especially for large MOF datasets with high-dimensional features. High-throughput computing (HPC) screening amplifies these requirements by evaluating thousands of MOF candidates simultaneously. Deep learning models, while powerful, often require the use of GPUs or specialized accelerators due to their parallel processing computing capabilities, contributing to increased energy consumption and operating costs. As a result, the energy-intensive nature of ML contributes to a larger carbon footprint, raising concerns about the environmental sustainability of computational research. As the size of MOF datasets and the complexity of ML models grow, addressing these sustainability issues becomes increasingly important.

To address these challenges, possible solutions include: (i) knowledge distillation. Simplifying complex models into smaller, more efficient versions without significant loss of accuracy. This approach reduces computational complexity while maintaining predictive performance, making ML models more accessible to resource-constrained researchers. (ii) Model compression. Techniques such as pruning, quantization, and low-rank factorization can reduce the size and complexity of ML models. These methods improve the efficiency of both training and inference processes, reducing hardware requirements and power consumption. It is also noted that model compression and knowledge distillation are usually in hand-to-hand mode. (iii) Transfer learning. Reusing pre-trained models for related tasks minimizes the need for extensive retraining, saving computational resources and time. (iv) Few-shot and zero-shot learning. Few-shot learning allows ML models to learn from a minimal number of data points, reducing the dependence on large data sets and extensive computing resources. Zero-shot learning, on the other hand, allows models to make predictions on unseen data by leveraging cross-domain knowledge, such as textual descriptions or structural similarities. These techniques improve model adaptability while reducing training time and energy consumption. (v) Efficient algorithms. Developing lightweight algorithms optimized for MOF datasets, such as sparse matrix operations or low-complexity neural architecture, can further reduce computational requirements.

4.6. Inverse design and experimental validation

Inverse design represents a transformative approach to MOF research, enabling the synthesis of materials tailored to achieve specific performance goals. Unlike traditional trial-and-error methods, inverse design uses ML to infer the optimal MOF structure from the desired properties – such as high adsorption capacity, thermal stability, or catalytic efficiency. This strategy not only accelerates discovery but also provides a systematic path to rational material design. Despite its potential, inverse design in MOFs faces significant challenges. The first is the inherent complexity of MOF structures, which are characterized by a large and high-dimensional design space that includes variations in metal nodes, organic linkers, and functional groups. Navigating this space to identify feasible candidates that match the desired properties requires robust and accurate ML models. In addition, ML-driven designs often require experimental validation to confirm their feasibility, creating additional resources and time constraints. Discrepancies between computational predictions and experimental results, due to limitations in the predictive accuracy of ML models or unaccounted for real-world variables, add further complexity.

In addition to inverse design, experimental validation remains a critical step to ensure the reliability and practicality of inverse design results. This involves synthesizing the proposed MOFs, characterizing their properties, and comparing them to predicted values. However, the iterative nature of this process can be resource intensive, especially for complex MOF designs.[152] High-throughput experimental techniques, automated synthesis platforms, and advanced characterization methods are being developed to mitigate these challenges and enable faster and more efficient validation cycles.

To address these limitations, future work could focus on (i) reinforcement learning for synthesis routes. Incorporating reinforcement learning into ML models can help identify feasible synthesis routes, making experimental realization of designed MOFs more practical. (ii) Active learning approaches. Iterative integration of experimental results into ML models can refine predictions and reduce the number of synthesis trials required. (iii) Hybrid models. Combining data-driven ML approaches with physics-based simulations can improve the accuracy of predictions by incorporating fundamental material properties and interactions. (iv) Collaborative validation platforms. Establishing shared resources for synthesis and characterization, supported by a network of research laboratories, can streamline the experimental validation process.

4.7. Long-term model accuracy and stability

Ensuring the long-term accuracy and stability of ML models for MOF research is critical, especially since these models are used to predict key aspects such as MOF stability and synthesizability. While ML has significantly accelerated the discovery of new materials, challenges arise in maintaining model performance over time and across evolving datasets. In terms of model accuracy challenges, ML models often rely on datasets that may become outdated or insufficient as new MOFs are developed. For example, models trained on older data may fail to predict properties for new MOFs with unconventional structures or compositions. In addition, the accuracy of ML predictions, especially for complex properties such as stability under operating conditions, can degrade if the training data lacks diversity or contains biases.

One of the most critical applications of ML in MOF research is the prediction of material stability under various conditions, such as thermal, chemical, and mechanical stress.[153] Stability predictions often require models to analyze structural features such as metal–ligand coordination strength, pore size, and framework flexibility. However, these predictions are challenging due to the lack of comprehensive experimental stability data and the need to account for dynamic interactions within the MOF framework. Another focus is on predicting the synthesizability of MOFs – whether a theoretically designed structure can be practically synthesized. This involves evaluating factors such as reaction conditions, precursor availability, and the likelihood of achieving the desired topology. ML models trained in synthesis data can provide valuable guidance but can be inaccurate due to missing or inconsistent experimental data.

Here are some practical strategies for improving long-term accuracy: (i) continuous model updating. Ensuring that models remain accurate over time requires continuous updating with new and diverse data. This can be achieved through lifelong learning, a subfield of ML that focuses on incrementally adding knowledge to an existing model without forgetting previously learned patterns. In addition, machine learning operations (MLOps) provide a structured approach to deploying and maintaining ML models, enabling automated retraining and evaluation as new experimental and computational data become available. Integrating high-throughput experimental and computational pipelines into MLOp frameworks can further enhance model robustness and adaptability. (ii) Transfer learning. This broad category of techniques leverages existing data and models to improve performance for specific MOF-related tasks. One commonly used method is fine-tuning, where a pre-trained model is adapted to a new application by adjusting a subset of its parameters. More generally, transfer learning enables knowledge transfer from well-studied domains to understudy MOF applications, reducing the need for extensive training data while improving model accuracy and generalization. (iii) Synthetic data. Generating synthetic data, such as through simulations or generative models, can expand training data sets and fill gaps in experimental data, especially for rare or complex MOF properties. (iv) Active learning. Active learning improves model efficiency by interactively selecting the most informative data points for human or expert annotation. Instead of passively training on static data sets, the model ''asks questions'' by identifying uncertain predictions and requesting validation from an oracle (*e.g.*, a researcher or an automated high-fidelity simulation). This targeted approach prioritizes data collection for uncertain regions, significantly improving model accuracy with minimal additional data. (v) Validation with experimental feedback. Integrating iterative

feedback loops between ML predictions and experimental results helps identify and correct inaccuracies, ensuring that models remain aligned with real-world performance.

4.8. Lack of the universal ML workflows

One of the major challenges in applying ML to MOF research is the lack of standardized workflows that guide the entire process, from data collection and preprocessing to model development, validation, and publication. This lack of universal workflows leads to inefficiencies, inconsistencies, and difficulty in reproducing results across research groups. Current ML workflows for MOF research are often fragmented, with individual researchers or groups developing *ad hoc* processes tailored to their specific datasets or applications. This variability leads to a lack of comparability across studies, making it difficult to benchmark model performance or build on previous work. In addition, the workflow often stops at model prediction, leaving gaps in the integration of experimental validation and interpretation.

Specifically, there are a few key elements missing from current workflows. First, there are the data pipeline issues. Many studies lack a clear and standardized process for collecting, cleaning, and curating data sets. Issues such as inconsistent formats, missing metadata, and unverified data quality can undermine the reliability of ML models. Second, researchers often use different algorithms and hyperparameter tuning methods for model selection and training without a consistent standard, leading to inconsistent results for similar problems. Third, for validation and reporting, few workflows emphasize rigorous validation using independent data sets or cross-validation techniques. In addition, there is a lack of transparency in reporting key metrics and model limitations. Next, integration with experimental studies is still not fully explored in the reported literature. The lack of a feedback loop between ML predictions and experimental results limits the ability to refine models and validate their applicability in the real world. Finally, there is no standard approach to presenting ML workflows and results in publications, leading to gaps in reproducibility and knowledge dissemination.

To address the above issues, a comprehensive and universal ML workflow for MOF research should begin with standardized protocols for data collection and curation, ensuring that datasets are well documented, consistently formatted, and annotated with metadata to enhance reproducibility. Data preprocessing should include automated tools for cleaning, normalization, and augmentation to address missing values, outliers, and inconsistencies. Feature engineering is critical and should include guidelines for selecting and generating descriptors, such as structural, chemical, and functional properties of MOFs, to improve model performance. The model development process should rely on standardized frameworks for algorithm selection, model training, and hyperparameter optimization, with transparent criteria to guide decisions. Rigorous validation techniques, including cross-validation, independent dataset testing, and uncertainty quantification, are essential, along with publication of validation metrics and benchmark results for reproducibility. Integration with experimental studies should include iterative feedback loops in which experimental results refine model predictions to ensure practical relevance. In addition, interpretability and explainability must be prioritized using tools such as SHAP scores or attention maps to increase confidence in ML models. Finally, the workflow should emphasize a standardized format for reporting methods, datasets, and results in publications, with open access repositories facilitating collaboration, reproducibility, and further innovation.

5. Emerging trends and future directions

5.1. Explainable AI in MOF design

As the application of ML in MOF research grows, there is an increasing need for explicable AI approaches to demystify the decision-making processes of complex models. Tools such as GNNs are being developed to provide deeper insights into how specific molecular structures within MOFs contribute to their performance. By highlighting critical features such as pore size, surface functionalization, and node-ligand interactions, Explainable AI enables researchers to identify structure–property relationships with greater clarity. This transparency not only builds confidence in ML models but also provides actionable guidance for the rational design of new MOFs with optimized properties for applications such as gas storage, catalysis, and sensing.

5.2. Active learning and autonomous research

Active learning, a process in which ML models identify and prioritize the most informative data points for acquisition, is emerging as a key strategy to address data scarcity in MOF research. By integrating active learning with automated experimental platforms, researchers can streamline the optimization of material properties while minimizing the need for exhaustive experimentation. These autonomous systems can iteratively refine ML models by feeding experimental results back into the training process, enabling rapid exploration of the vast MOF design space. The potential for fully autonomous materials design, where AI independently generates, tests, and validates hypotheses, represents a transformative leap for the field, accelerating discovery cycles and reducing research costs.

5.3. Integration with quantum computing

Quantum computing is poised to revolutionize the simulation of complex MOF structures by overcoming the limitations of classical computing, particularly in modeling molecular interactions and reaction pathways. The synergy between quantum computing and ML offers new ways to predict the electronic, thermal, and mechanical properties of MOFs with unprecedented accuracy. Quantum algorithms can improve ML models by providing accurate quantum mechanical descriptors, while ML can help reduce the computational burden of quantum simulations. This integration has the potential to unlock new

insights into the behavior of MOFs under extreme conditions, paving the way for innovative applications in energy storage, catalysis, and beyond.

5.4. Multi-disciplinary collaboration

The advancement of ML in MOF research requires robust interdisciplinary collaborations, drawing on expertise from chemistry, materials science, computational science, and data engineering. Collaborative efforts can bridge knowledge gaps by bringing together theoretical insights, experimental validation, and computational innovations. Cross-disciplinary resources such as shared databases, joint research initiatives, and collaborative platforms can accelerate breakthroughs in MOF design and ML applications. By fostering a culture of openness and shared learning, the community can tackle complex challenges, from designing novel MOF architectures to addressing global issues such as climate change and resource sustainability. Here, the emerging trends and future directions for ML-based MOF design and application exploration are schematically shown in Fig. 9.

6. Conclusions

6.1. Summary of key insights

ML has emerged as a transformative tool in MOF research, bridging the gap between fundamental theory and practical electrochemical ammonia synthesis. By enabling high-throughput screening, advanced property prediction, and data-driven optimization, ML has accelerated the discovery of novel MOFs tailored for diverse applications such as gas storage, catalysis, sensing, water treatment, and energy systems. Fundamental advances in ML methodologies, including the use of graph neural networks, explicable AI approaches, and active learning, have provided deeper insights into structure–property relationships, allowing researchers to accurately identify key design principles. In addition, the integration of ML with experimental workflows and the establishment of standardized datasets have significantly improved the reliability and scalability of materials design. These contributions underscore the pivotal role of ML in transforming MOF research, enabling researchers to navigate the vast chemical space of MOFs with unprecedented efficiency and accuracy.

6.2. Outlook

Looking ahead, the future of ML in MOF research has immense potential. Enhancing the generalization capabilities of ML models remains a critical focus, with strategies such as transfer learning and diverse data augmentation expected to play a key role. Improving model interpretability through advanced visualization techniques and explicable AI tools will further increase confidence in ML predictions and guide the rational design of materials. The integration of quantum computing with ML offers exciting opportunities to simulate complex molecular interactions and advance our understanding of MOF behavior under extreme conditions. Active learning combined with autonomous experimental platforms is expected to streamline the materials discovery process, enabling fully automated cycles of prediction, synthesis, and validation. Addressing challenges such as energy consumption and environmental sustainability will be critical as the field moves toward more efficient and greener computational approaches. By embracing interdisciplinary collaboration and leveraging emerging technologies, ML-driven MOF research is poised to address global

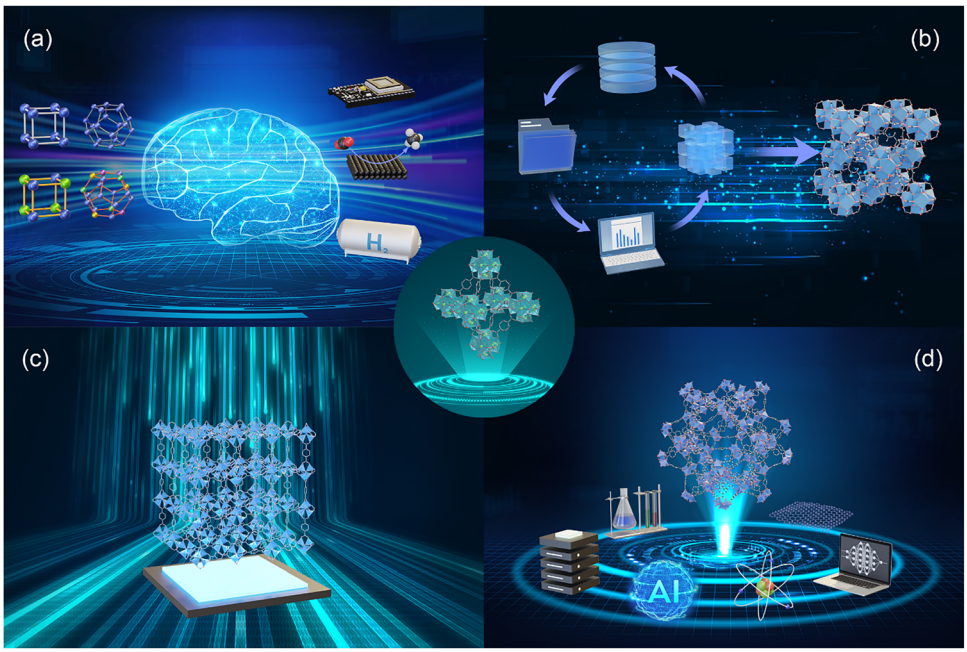

Fig. 9 Emerging trends and future directions for ML-based MOF design and application exploration. (a) Explainable AI in MOF design, (b) active learning and autonomous research, (c) integration with quantum computing, and (d) multi-disciplinary collaboration.

challenges and open new frontiers in materials science, shaping a sustainable and innovative future.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

φ	Porosity
ρ	Density
ADCC	Attention-densely connected conventional network
ANFIS	Adaptive neuro-fuzzy inference systems
APS	Adsorbent performance score
ASA	Accessible surface area
AUC	Area under the curve
AV	Accessible volume
BPNN	Back propagation neural network
CFFNN	Cascade feedforward neural networks
CFIDs	Force-filed-inspired descriptors
CGCNN	Crystal graph convolutional neural network
CMIS	Committee machine intelligence system
CoRE	Computation-ready, experimental
ELM	Extreme learning machine
ERT	Extremely randomized trees
ET	Extra trees
ETR	Extra trees regressor
DCNN	Deep convolutional neural network
DFT	Density functional theory
DNNs	Deep, feed-forward neural networks
DOE	Department of energy
DT	Decision tree
FAIR	Findable, accessible, interoperable, and reusable
FNN	Feedforward neural networks
GANs	Generative adversarial networks
GBDT	Gradient boosting decision trees
GBRT	Gradient-boost regression tress
GCMC	Grand canonical Monte Carlo
GM	Volumetric uptake using geometric data
GNN	Graph neural networks
GRNN	General regression neural networks
GPR	Gaussian process regression
GSA	Gravimetric surface area
GUM	Predicts gravimetric uptake
HTCS	High-throughput computational screening
HTE	High-throughput experimentation
KNN	*k*-Nearest neighbors
LAMOFS	Long-arm metal–organic frameworks
LCD	Largest cavity diameter
LDA	Linear discriminant analysis
LDH	Layered double hydroxide
LGBM	Light gradient boosting machines
LIME	Local interpretable model-agnostic explanations
LSSVM	Least squares support vector machines
MAE	Mean absolute error
MCGCNN	Modified crystal graph convolutional neural network
MCTS	Monte Carlo tree search
MD	Molecular dynamics
ML	Machine learning
MLOps	Machine learning operations
MLP	Multilayer perceptron
MLPNN	Multilayer perceptron neural networks
MOF	Metal–organic framework
MSE	Mean squared error
NCA	Neighbor component analysis
NGM	Volumetric uptake without geometric data
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PLD	Pore-limiting diameter
PSO	Particle swarm optimization
Q_{st}	Heat of adsorption
RAE	Relative absolute error
RDF	Radial distribution function
RF	Random forest
RFC	Random forest classifiers
RL	Reinforcement learning
RMSE	Root mean squared error
RNN	Recurrent neural network
RR	Ridge regression
RRSE	Root relative squared error
SBUs	Secondary building units
SHAP	Shapley additive explanation
SVM	Support vector machine
TMINN	Transition-metal interlink neural network
TPOT	Tree-based pipeline optimization tool
UMAP	Unified manifold approximation and projection
VCD	Volumetric deliverable capacity
VF	Void fraction
VOC	Volatile organic compounds
VSA	Volumetric surface area
XAI	Explainable artificial intelligence
XGB	Extreme gradient boost

Data availability

This article is a review and does not report new experimental or computational data. All data discussed in this work are derived from previously published studies, which are cited throughout the manuscript. No new datasets were generated for this review.

Acknowledgements

The authors acknowledge funding support from Singapore MOE AcRF Tier 1 grant no. 2020-T1-001-031 and RT6/22. The English writing was grammatically checked by using the AI-assisted tool DeepL.

References

1 D. Ye and S. C. E. Tsang, Prospects and challenges of green ammonia synthesis, *Nat. Synth.*, 2023, **2**(7), 612–623.
2 C. M. Goodwin, P. Lömker, D. Degerman, B. Davies, M. Shipilin, F. Garcia-Martinez, S. Koroidov, J. Katja Mathiesen, R. Rameshan and G. L. Rodrigues, Operando probing of the surface chemistry during the Haber–Bosch process, *Nature*, 2024, **625**(7994), 282–286.
3 W. Li, K. Li, Y. Ye, S. Zhang, Y. Liu, G. Wang, C. Liang, H. Zhang and H. Zhao, Efficient electrocatalytic nitrogen reduction to ammonia with aqueous silver nanodots, *Commun. Chem.*, 2021, **4**(1), 10.
4 H.-j Chen, G.-r Deng, Z.-s Feng, Z.-q Xu, M.-y Yang, Y. Huang, Q. Peng, T. Li and Y. Wang, Enhanced electrocatalytic performance of TiO2 nanoparticles by Pd doping toward ammonia synthesis under ambient conditions, *Chem. Commun.*, 2022, **58**(19), 3214–3217.
5 J. Liu, R. Duan, Y. Xu, C. Zhang, C. Lv, E. Hu, J. Gao, B. Han, C. Lee and Z. Liu, Decoupled Control of CO2 and Nitrate Reduction Intermediates to Enable Efficient Tandem Urea Electrosynthesis, *ACS Nano*, 2025, **19**(32), 29646–29656.
6 X. Zhang, Y. Wang, Y. Wang, Y. Guo, X. Xie, Y. Yu and B. Zhang, Recent advances in electrocatalytic nitrite reduction, *Chem. Commun.*, 2022, **58**(17), 2777–2787.
7 D.-W. Lim and H. Kitagawa, Proton transport in metal–organic frameworks, *Chem. Rev.*, 2020, **120**(16), 8416–8467.
8 S. Zhao, C. Tan, C.-T. He, P. An, F. Xie, S. Jiang, Y. Zhu, K.-H. Wu, B. Zhang and H. Li, Structural transformation of highly active metal–organic framework electrocatalysts during the oxygen evolution reaction, *Nat. Energy*, 2020, **5**(11), 881–890.
9 S. Kitagawa, Metal–organic frameworks (MOFs), *Chem. Soc. Rev.*, 2014, **43**(16), 5415–5418.
10 B. Han and A. Chakraborty, *Thermodynamic trends for carbon dioxide and hydrogen adsorption on MOFs, in Gas-Solid Adsorption Thermodynamics*, Elsevier, 2026, pp. 93–122.
11 Y. Zou, Y. Yan, Q. Xue, C. Zhang, T. Bao, X. Zhang, L. Yuan, S. Qiao, L. Song and J. Zou, MOF-on-MOF Heterostructured Electrocatalysts for Efficient Nitrate Reduction to Ammonia., *Angew. Chem., Int. Ed.*, 2024, **63**(41), e202409799.
12 S. L. James, Metal-organic frameworks, *Chem. Soc. Rev.*, 2003, **32**(5), 276–288.
13 M. Bonneau, C. Lavenn, J.-J. Zheng, A. Legrand, T. Ogawa, K. Sugimoto, F.-X. Coudert, R. Reau, S. Sakaki and K.-I. Otake, Tunable acetylene sorption by flexible catenated metal–organic frameworks, *Nat. Chem.*, 2022, **14**(7), 816–822.
14 J.-H. Dou, M. Q. Arguilla, Y. Luo, J. Li, W. Zhang, L. Sun, J. L. Mancuso, L. Yang, T. Chen and L. R. Parent, Atomically precise single-crystal structures of electrically conducting 2D metal–organic frameworks, *Nat. Mater.*, 2021, **20**(2), 222–228.
15 A. Koutsianos, R. Pallach, L. Frentzel-Beyme, C. Das, M. Paulus, C. Sternemann and S. Henke, Breathing porous liquids based on responsive metal-organic framework particles, *Nat. Commun.*, 2023, **14**(1), 4200.
16 S. Dissegna, K. Epp, W. R. Heinz, G. Kieslich and R. A. Fischer, Defective metal-organic frameworks., *Adv. Mater.*, 2018, **30**(37), 1704501.
17 M. Sadakiyo, T. Yamada, K. Honda, H. Matsui and H. Kitagawa, Control of crystalline proton-conducting pathways by water-induced transformations of hydrogen-bonding networks in a metal–organic framework, *J. Am. Chem. Soc.*, 2014, **136**(21), 7701–7707.
18 C. Koschnick, M. W. Terban, R. Frison, M. Etter, F. A. Böhm, D. M. Proserpio, S. Krause, R. E. Dinnebier, S. Canossa and B. V. Lotsch, Unlocking new topologies in Zr-based metal–organic frameworks by combining linker flexibility and building block disorder, *J. Am. Chem. Soc.*, 2023, **145**(18), 10051–10060.
19 M. Rubio-Martinez, T. D. Hadley, M. P. Batten, K. Constanti-Carey, T. Barton, D. Marley, A. Mönch, K. S. Lim and M. R. Hill, Scalability of continuous flow production of metal–organic frameworks, *ChemSusChem*, 2016, **9**(9), 938–941.
20 Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent and Z. W. Seh, Machine learning for a sustainable energy future, *Nat. Rev. Mater.*, 2023, **8**(3), 202–215.
21 A. Nandy, C. Duan and H. J. Kulik, Using machine learning and data mining to leverage community knowledge for the engineering of stable metal–organic frameworks, *J. Am. Chem. Soc.*, 2021, **143**(42), 17535–17547.
22 S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen and V. Van Speybroeck, Machine learning potentials for metal-organic frameworks using an incremental learning approach, *npj Comput. Mater.*, 2023, **9**(1), 1–8.
23 Y. Zhu, H. Ji, T. Huang, Y. Sun and H. Pang, Research Progress on the Application of MOF and MOF-Based Materials in Nitrogen Reduction, *Adv. Sustainable Syst.*, 2024, **8**(10), 2400225.
24 T.-H. Yang, S. Yue, W.-W. Gong, R.-Q. Wang, W.-D. Hu, X.-P. Liu, P.-Z. Gao, H. Qin, W.-M. Guo and H.-N. Xiao, Mechanism of MOF-derived Fe-doped TiO2 material enhanced by the AC magnetic field on electrochemical nitrogen reduction reaction (eNRR), *Int. J. Hydrogen Energy*, 2024, **91**, 447–457.
25 Y. Hu, J. Liu, C. Lee, W. Luo, J. Dong, Z. Liang, M. Chen, E. Hu, M. Zhang and X. Y. Debbie Soo, Balanced NOx and Proton Adsorption for Efficient Electrocatalytic NOx to NH3 Conversion, *ACS Nano*, 2023, **17**(23), 23637–23648.
26 J. Liu, Y. Xu, R. Duan, M. Zhang, Y. Hu, M. Chen, B. Han, J. Dong, C. Lee and L. S. R. Kumara, Reaction-driven formation of anisotropic strains in FeTeSe nanosheets boosts low-concentration nitrate reduction to ammonia, *Nat. Commun.*, 2025, **16**(1), 3595.
27 C. Yang, Y. Tang, Q. Yang, B. Wang, X. Liu, Y. Li, W. Yang, K. Zhao, G. Wang and Z. Wang, Copper–nickel–MOF/nickel foam catalysts grown in situ for efficient electrochemical nitrate reduction to ammonia, *J. Hazard. Mater.*, 2024, **480**, 136036.
28 X.-X. Fu, H. Guo, D.-H. Si, H.-J. Zhu, Y.-Y. Lan, Y.-B. Huang and R. Cao, Hydrogen-bond mediated electrocatalytic nitrate reduction to ammonia over metal–organic frameworks with industrial current density, *Chem. Sci.*, 2025, **16**(29), 13503–13513.
29 Y. Li, Q. Lei, W. T. Hong, X. Liu, C. Xue and J. K. Kim, Transition Metal-Based Materials for Electrochemical and Photoelectrochemical Carbon-Free Nitrogen Cycling as H-Carrier. in Exploration, 2025, Wiley Online Library.
30 A. Yan, Y. Feng, X. Zhang, J. Sun, J. Wei, C. Shi, L. Xia, S. Wang, H. Zhang and Y. Guo, In Situ Reconstruction of Bimetallic MOFs to Form Copper-Cobalt Relay Catalysis for Efficient Nitrate Reduction to Ammonia, *Small*, 2025, **21**(39), e06256.
31 H. Sun, Z. Xia, Y. Qi, Q. Xu, J. Han, J. Wu, J.-S. Qin and H. Rao, Tandem Cu–Co Sites in MOF-818 for Efficient Ammonia Electrosynthesis from Nitrate in Neutral Media, *ACS Catal.*, 2025, **15**(19), 16581–16590.
32 Q.-N. Wang, A. S. Pittman, Y.-T. Xu and Y. Cao, The electrocatalytic reduction of nitrate to ammonia (NARR) using MOF-808 enhanced by the transition metal Fe(III) and $H_2PO_4^-$/HPO_4^{2-}, *Int. J. Hydrogen Energy*, 2024, **83**, 1143–1149.
33 X. Hu, M. Zhang, C. Lai, M. Cheng, F. Xu, D. Ma, L. Li, H. Yan, H. Sun and X. Fan, Metal-organic frameworks derived low-cost Cu-doped Co3O4 for efficient reduction of ultra-low nitrate concentrations to ammonia, *Chem. Eng. J.*, 2024, **493**, 152543.
34 J. Chen, T. Gong, Q. Hou, J. Li, L. Zhang, D. Zhao, Y. Luo, D. Zheng, T. Li and S. Sun, Co/N-doped carbon nanospheres derived from an adenine-based metal organic framework enabled high-efficiency electrocatalytic nitrate reduction to ammonia, *Chem. Commun.*, 2022, **58**(97), 13459–13462.
35 B. Han and A. Chakraborty, Ligand extension of aluminum fumarate metal-organic framework in transferring higher water for adsorption desalination, *Desalination*, 2024, **592**, 118135.
36 E. M. Johnson, S. Ilic and A. J. Morris, Design strategies for enhanced conductivity in metal–organic frameworks, *ACS Cent. Sci.*, 2021, **7**(3), 445–453.
37 B. Han and A. Chakraborty, Synergistic ionic liquid encapsulated MIL-101 (Cr) metal-organic frameworks for an innovative adsorption desalination system, *J. Cleaner Prod.*, 2024, **474**, 143565.
38 B. Han, L. Zhong, C. Chen, J. Ding, C. Lee, J. Liu, M. Chen, S. Tso, Y. Hu and C. Lv, Tuning Main Group Element-based Metal–Organic Framework to Boost Electrocatalytic Nitrogen Reduction Under Ambient Conditions, *Small*, 2024, **20**(9), 2307506.
39 X. Zhu, H. Huang, H. Zhang, Y. Zhang, P. Shi, K. Qu, S.-B. Cheng, A.-L. Wang and Q. Lu, Filling mesopores of conductive metal–organic frameworks with Cu clusters for selective nitrate reduction to ammonia, *ACS Appl. Mater. Interfaces*, 2022, **14**(28), 32176–32182.
40 Y. Zhu, G. Dong, F. Pan, T. Wang, L. Zhang, H. Wang, L. Ge and P. Zhang, Ir NCs Embedded Co-MOF Nanosheets for Boosting

Electrochemical Nitrate Reduction to Ammonia Performance, *ACS Appl. Mater. Interfaces*, 2025, **17**(19), 28084–28093.
41 F. Pan, J. Zhou, T. Wang, Y. Zhu, H. Ma, J. Niu and C. Wang, Revealing the activity origin of ultrathin nickel metal–organic framework nanosheet catalysts for selective electrochemical nitrate reduction to ammonia: Experimental and density functional theory investigations, *J. Colloid Interface Sci.*, 2023, **638**, 26–38.
42 J. Liu, T. A. Goetjen, Q. Wang, J. G. Knapp, M. C. Wasson, Y. Yang, Z. H. Syed, M. Delferro, J. M. Notestein and O. K. Farha, MOF-enabled confinement and related effects for chemical catalyst presentation and utilization, *Chem. Soc. Rev.*, 2022, **51**(3), 1045–1097.
43 P. Deria, W. Bury, I. Hod, C.-W. Kung, O. Karagiaridi, J. T. Hupp and O. K. Farha, MOF functionalization via solvent-assisted ligand incorporation: phosphonates vs carboxylates, *Inorg. Chem.*, 2015, **54**(5), 2185–2192.
44 F. Afshariazar, A. Morsali and P. Retailleau, Investigation of the influence of functionalization strategy on urea 2D MOF catalytic performance, *Inorg. Chem.*, 2023, **62**(8), 3498–3505.
45 Ö. Durak, H. Kulak, S. Kavak, H. M. Polat, S. Keskin and A. Uzun, Towards complete elucidation of structural factors controlling thermal stability of IL/MOF composites: Effects of ligand functionalization on MOFs, *J. Phys.: Condens. Matter*, 2020, **32**(48), 484001.
46 W. Zhang, W. Ji, L. Li, P. Qin, I. E. Khalil, Z. Gu, P. Wang, H. Li, Y. Fan and Z. Ren, Exploring the fundamental roles of functionalized ligands in platinum@ metal–organic framework catalysts, *ACS Appl. Mater. Interfaces*, 2020, **12**(47), 52660–52667.
47 L. Meng, Q. Cheng, C. Kim, W. Y. Gao, L. Wojtas, Y. S. Chen, M. J. Zaworotko, X. P. Zhang and S. Ma, Crystal engineering of a microporous, catalytically active fcu topology MOF using a custom-designed metalloporphyrin linker., *Angew. Chem., Int. Ed.*, 2012, **51**(40), 10082–10085.
48 W. Y. Gao, Y. Chen, Y. Niu, K. Williams, L. Cash, P. J. Perez, L. Wojtas, J. Cai, Y. S. Chen and S. Ma, Crystal engineering of an nbo topology metal–organic framework for chemical fixation of CO2 under ambient conditions, *Angew. Chem., Int. Ed.*, 2014, **53**(10), 2615–2619.
49 H. Jiang, W. Zhang, X. Kang, Z. Cao, X. Chen, Y. Liu and Y. Cui, Topology-based functionalization of robust chiral Zr-based metal–organic frameworks for catalytic enantioselective hydrogenation, *J. Am. Chem. Soc.*, 2020, **142**(21), 9642–9652.
50 B. Han and A. Chakraborty, Tailoring Zirconium-based metal organic frameworks for enhancing Hydrophilic/Hydrophobic Characteristics: Simulation and experimental investigation, *J. Mol. Liq.*, 2021, **341**, 117381.
51 M. D. Firouzjaei, F. A. Afkhami, M. R. Esfahani, C. H. Turner and S. Nejati, Experimental and molecular dynamics study on dye removal from water by a graphene oxide-copper-metal organic framework nanocomposite, *J. Water Process Eng.*, 2020, **34**, 101180.
52 F. Formalik, K. Shi, F. Joodaki, X. Wang and R. Q. Snurr, Exploring the Structural, Dynamic, and Functional Properties of Metal-Organic Frameworks through Molecular Modeling., *Adv. Funct. Mater.*, 2024, **34**(43), 2308130.
53 S. Majumdar, S. M. Moosavi, K. M. Jablonka, D. Ongari and B. Smit, Diversifying databases of metal organic frameworks for high-throughput computational screening, *ACS Appl. Mater. Interfaces*, 2021, **13**(51), 61004–61014.
54 Z. Shi, W. Yang, X. Deng, C. Cai, Y. Yan, H. Liang, Z. Liu and Z. Qiao, Machine-learning-assisted high-throughput computational screening of high performance metal–organic frameworks, *Mol. Syst. Des. Eng.*, 2020, **5**(4), 725–742.
55 A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein and R. Q. Snurr, High-throughput predictions of metal–organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration, *npj Comput. Mater.*, 2022, **8**(1), 1–10.
56 Y. Xie, C. Zhang, H. Deng, B. Zheng, J.-W. Su, K. Shutt and J. Lin, Accelerate synthesis of metal–organic frameworks by a robotic platform and bayesian optimization, *ACS Appl. Mater. Interfaces*, 2021, **13**(45), 53485–53491.
57 T. Jiang, J. L. Gradus and A. J. Rosellini, Supervised machine learning: a brief primer, *Behav. Ther.*, 2020, **51**(5), 675–687.
58 C. Campbell and Y. Ying, *Learning with Support Vector Machines*, Springer Nature, 2022.
59 P. Palimkar, R. N. Shaw and A. Ghosh, Machine learning technique to prognosis diabetes disease: Random forest classifier approach, in Advanced computing and intelligent technologies: proceedings of ICACIT 2021, 2022, Springer.
60 B. Charbuty and A. Abdulazeez, Classification based on decision tree algorithm for machine learning, *J. Appl. Sci. Technol. Trends*, 2021, **2**(01), 20–28.
61 A. V. Konstantinov and L. V. Utkin, Interpretable machine learning with an ensemble of gradient boosting machines, *Knowl.-Based Syst.*, 2021, **222**, 106993.
62 A. Kaveh, Applications of artificial neural networks and machine learning in civil engineering, *Stud. Comput. Intell.*, 2024, **1168**, 472.
63 M. Bansal, A. Goyal and A. Choudhary, A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning, *Decis. Anal.*, 2022, **3**, 100071.
64 B. Mihaljević, C. Bielza and P. Larrañaga, Bayesian networks for interpretable machine learning and optimization, *Neurocomputing*, 2021, **456**, 648–665.
65 S. Naeem, A. Ali, S. Anam and M. M. Ahmed, An unsupervised machine learning algorithms: Comprehensive review, *Int. J. Comput. Digital Syst.*, 2023, **13**(1), 911–921.
66 K. P. Sinaga and M.-S. Yang, Unsupervised K-means clustering algorithm, *IEEE Access*, 2020, **8**, 80716–80727.
67 P. Garikapati, K. Balamurugan, T. Latchoumi and R. Malkapuram, A cluster-profile comparative study on machining AlSi7/63% of SiC hybrid composite using agglomerative hierarchical clustering and K-means, *Silicon*, 2021, **13**(4), 961–972.
68 P. An, Z. Wang and C. Zhang, Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection, *Inf. Process. Manage.*, 2022, **59**(2), 102844.
69 A. A. Bushra and G. Yi, Comparative analysis review of pioneering DBSCAN and successive density-based clustering algorithms, *IEEE Access*, 2021, **9**, 87918–87935.
70 T. Kurita, Principal component analysis (PCA), *Computer vision: a reference guide*, Springer, 2021, pp. 1013–1016.
71 M. Balamurali, T-Distributed stochastic neighbor embedding, *Encyclopedia of mathematical geosciences*, Springer, 2023, pp. 1527–1535.
72 W. H. L. Pinaya, S. Vieira, R. Garcia-Dias and A. Mechelli, Auto-encoders, *Machine learning*, Elsevier, 2020, pp. 193–208.
73 J. E. Van Engelen and H. H. Hoos, A survey on semi-supervised learning, *Mach. Learn.*, 2020, **109**(2), 373–440.
74 L. Zhang, L. Yang, T. Ma, F. Shen, Y. Cai and C. Zhou, A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data, *Geoderma*, 2021, **384**, 114809.
75 E. Grolman, D. Cohen, T. Frenklach, A. Shabtai and R. Puzis, How and when to stop the co-training process, *Expert Syst. Appl.*, 2022, **187**, 115841.
76 K. Weiss, T. M. Khoshgoftaar and D. Wang, A survey of transfer learning, *J. Big data*, 2016, **3**, 1–40.
77 Y. Zhang and Q. Yang, An overview of multi-task learning, *Nat. Sci. Rev.*, 2018, **5**(1), 30–43.
78 A. G. Barto, Reinforcement Learning: An Introduction. By Richard's Sutton, *SIAM Rev.*, 2021, **6**(2), 423.
79 E. Osaro and Y. J. Colón, Optimizing the prediction of adsorption in metal–organic frameworks leveraging Q-learning., *AIChE J.*, 2024, **70**(12), e18611.
80 A. Hafiz, A survey of deep q-networks used for reinforcement learning: state of the art. Intelligent Communication Technologies and Virtual Mobile Networks: Proceedings of ICICV 2022, 2022, 393–402.
81 C. Daskalakis, D. J. Foster and N. Golowich, Independent policy gradient methods for competitive reinforcement learning, *Adv. Neural Inf. Process. Syst.*, 2020, **33**, 5527–5540.
82 H. Kumar, A. Koppel and A. Ribeiro, On the sample complexity of actor-critic method for reinforcement learning with function approximation, *Mach. Learn.*, 2023, **112**(7), 2433–2467.
83 S. Mo, X. Pei and C. Wu, Safe reinforcement learning for autonomous vehicle using monte carlo tree search, *IEEE Trans. Intell. Transp. Syst*, 2021, **23**(7), 6766–6773.
84 J. Qi, J. Du, S. M. Siniscalchi, X. Ma and C.-H. Lee, On mean absolute error for deep neural network based vector-to-vector regression, *IEEE Signal Process. Lett.*, 2020, **27**, 1485–1489.

85 D. Chicco, M. J. Warrens and G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Comput. Sci.*, 2021, **7**, e623.
86 T. O. Hodson, Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not, *Geosci. Model Dev. Discuss.*, 2022, **2022**, 1–10.
87 D. Jung and Y. Choi, Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation, *Minerals*, 2021, **11**(2), 148.
88 N. Kosaraju, S. R. Sankepally and K. Mallikharjuna Rao, Categorical data: Need, encoding, selection of encoding method and its emergence in machine learning models—a practical review study on heart disease prediction dataset using pearson correlation, in *Proceedings of International Conference on Data Science and Applications*: ICDSA 2022, Volume 1, 2023, Springer.
89 Y. J. Park, S. Yoon and S. E. Jerng, Machine learning of metal-organic framework design for carbon dioxide capture and utilization, *J. CO2 Utili.*, 2024, **89**, 102941.
90 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, Large-scale screening of hypothetical metal–organic frameworks, *Nat. Chem.*, 2012, **4**(2), 83–89.
91 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling and J. S. Camp, Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019, *J. Chem. Eng. Data*, 2019, **64**(12), 5985–5998.
92 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, Development of a Cambridge Structural Database subset: a collection of metal–organic frameworks for past, present, and future, *Chem. Mater.*, 2017, **29**(7), 2618–2625.
93 P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi and M. M. Maroto-Valer, Data-driven design of metal–organic frameworks for wet flue gas CO2 capture, *Nature*, 2019, **576**(7786), 253–256.
94 A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein and R. Q. Snurr, Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery, *Matter*, 2021, **4**(5), 1578–1597.
95 A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford and D. S. Sholl, *The Open DAC 2023 dataset and challenges for sorbent discovery in direct air capture*, ACS Publications, 2024.
96 J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod and V. Ogden, ARC–MOF: a diverse database of metal-organic frameworks with DFT-derived partial atomic charges and descriptors for machine learning, *Chem. Mater.*, 2023, **35**(3), 900–916.
97 L. T. Glasby, K. Gubsch, R. Bence, R. Oktavian, K. Isoko, S. M. Moosavi, J. L. Cordiner, J. C. Cole and P. Z. Moghadam, DigiMOF: a database of metal–organic framework synthesis information generated via text mining, *Chem. Mater.*, 2023, **35**(11), 4510–4524.
98 B. Han and A. Chakraborty, Experimental investigation for water adsorption characteristics on functionalized MIL-125 (Ti) MOFs: enhanced water transfer and kinetics for heat transformation systems, *Int. J. Heat Mass Transfer*, 2022, **186**, 122473.
99 K. Neikha and A. Puzari, Metal–Organic Frameworks through the Lens of Artificial Intelligence: A Comprehensive Review, *Langmuir*, 2024, **40**(42), 21957–21975.
100 B. Han and A. Chakraborty, Highly efficient adsorption desalination employing protonated-amino-functionalized MOFs, *Desalination*, 2022, **541**, 116045.
101 K. Mukherjee and Y. J. Colón, Machine learning and descriptor selection for the computational discovery of metal-organic frameworks, *Mol. Simul.*, 2021, **47**(10–11), 857–877.
102 I.-T. Sung, Y.-H. Cheng, C.-M. Hsieh and L.-C. Lin, Machine Learning for Gas Adsorption in Metal–Organic Frameworks: A Review on Predictive Descriptors, *Ind. Eng. Chem. Res.*, 2025, **64**(4), 1859–1875.
103 J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang and H. Liu, Feature selection: A data perspective, *ACM Comput. Surv*, 2017, **50**(6), 1–45.
104 Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu and M.-H. YangUniversal style transfer via feature transforms, *Advances in Neural Information Processing Systems 30*, Proc. 31st Conf. Neural Information Processing Systems (NeurIPS 2017), 2017, pp. 1–9.
105 S. Choi, T. Kim, H. Ji, H. J. Lee and M. Oh, Isotropic and anisotropic growth of metal–organic framework (MOF) on MOF: logical inference on MOF structure based on growth behavior and morphological feature, *J. Am. Chem. Soc.*, 2016, **138**(43), 14434–14440.
106 L. Friedman and O. V. Komogortsev, Assessment of the effectiveness of seven biometric feature normalization techniques, *IEEE Trans. Inf. Forensics Secur.*, 2019, **14**(10), 2528–2536.
107 R. Li, W. Zhong and L. Zhu, Feature screening via distance correlation learning, *J. Am. Stat. Assoc.*, 2012, **107**(499), 1129–1139.
108 B. F. Darst, K. C. Malecki and C. D. Engelman, Using recursive feature elimination in random forest to account for correlated variables in high dimensional data, *BMC Genet.*, 2018, **19**(1), 65.
109 X. Lu, Z. Xie, X. Wu, M. Li and W. Cai, Hydrogen storage metal-organic framework classification models based on crystal graph convolutional neural networks, *Chem. Eng. Sci.*, 2022, **259**, 117813.
110 N. Pouyanfar, M. Ahmadi, S. M. Ayyoubzadeh and F. Ghorbani-Bidkorpeh, Drug delivery system tailoring via metal-organic framework property prediction using machine learning: A disregarded approach, *Mater. Today Commun.*, 2024, **38**, 107938.
111 B. Han and A. Chakraborty, Machine learning optimized adsorption heat transformations from adsorbents screening to enhanced performances, *Energy Convers. Manage.*, 2025, **344**, 120272.
112 B. Han and A. Chakraborty, Evaluation on the performances of adsorption desalination employing functionalized metal-organic frameworks (MOFs), *Appl. Therm. Eng.*, 2023, **218**, 119365.
113 H. Demir, H. Daglar, H. C. Gulbalkan, G. O. Aksu and S. Keskin, Recent advances in computational modeling of MOFs: From molecular simulations to machine learning, *Coord. Chem. Rev.*, 2023, **484**, 215112.
114 Y. Yang, Z. Yu and D. S. Sholl, Machine Learning Models for Predicting Molecular Diffusion in Metal–Organic Frameworks Accounting for the Impact of Framework Flexibility, *Chem. Mater.*, 2023, **35**(23), 10156–10168.
115 J. Lee, I. Lee, J. Park, H. Kim, M. Kim, K. Min and S. Lee, Optimal Surrogate Models for Predicting the Elastic Moduli of Metal–Organic Frameworks via Multiscale Features, *Chem. Mater.*, 2023, **35**(24), 10457–10475.
116 B. Han and A. Chakraborty, Evaluation of energy flow, dissipation and performances for advanced adsorption assisted heat transformation systems: Temperature-entropy frameworks, *Energy Convers. Manage.*, 2021, **240**, 114264.
117 P. Ying, T. Liang, K. Xu, J. Zhang, J. Xu, Z. Zhong and Z. Fan, Submicrometer phonon mean free paths in metal–organic frameworks revealed by machine learning molecular dynamics simulations, *ACS Appl. Mater. Interfaces*, 2023, **15**(30), 36412–36422.
118 J. Park, H. Kim, Y. Kang, Y. Lim and J. Kim, From data to discovery: recent trends of machine learning in metal–organic frameworks, *JACS Au*, 2024, **4**(10), 3727–3743.
119 B. Han, Y. Sim, Q. Yan, N. Mathews and J.-C. P. Gabriel, From electronic wastes to efficient and specific filtration membranes: A photovoltaic upcycling case enabling silver urban mining, *J. Cleaner Prod.*, 2025, **505**, 145528.
120 B. Han, Z. Liu, D. Xia, G. Zante, Q. Yan and J.-C. P. Gabriel, Enhanced silver recovery from electronic wastes using ionic liquid-integrated nanocomposite membrane, *Sep. Purif. Technol.*, 2025, **366**, 132689.
121 B. Han and J.-C. P. Gabriel, Thin-film nanocomposite (TFN) membrane technologies for the removal of emerging contaminants from wastewater, *J. Cleaner Prod.*, 2024, **480**, 144043.
122 B. Han, S. M. Chevrier, Q. Yan and J.-C. P. Gabriel, Tailorable metal–organic framework based thin film nanocomposite membrane for lithium recovery from wasted batteries, *Sep. Purif. Technol.*, 2024, **334**, 125943.
123 H. Mashhadimoslem, M. A. Abdol, P. Karimi, K. Zanganeh, A. Shafeen, A. Elkamel and M. Kamkar, Computational and Machine Learning Methods for CO2 Capture Using Metal–Organic Frameworks, *ACS Nano*, 2024, **18**(35), 23842–23875.
124 B. Han and A. Chakraborty, Functionalization on metal–organic frameworks to enhance water adsorption uptakes and kinetics for

cooling applications. Advancements in Non-Conventional Cooling and Thermal Storage Strategies, 2024, 65–103.
125 B. Han and A. Chakraborty, Synthesis and characteristics of ionic liquid-implanted HKUST-1 metal–organic frameworks for transforming heat into extraordinary water transfer, *ACS Sustainable Chem. Eng.*, 2024, **12**(21), 8115–8127.
126 B. Han, Synthesisation, characterisation and water adsorption on novel functional metal organic frameworks for heat transformation process, 2020.
127 X. Wei, D. Peng, L. Shen, Y. Ai and Z. Lu, Analyzing of metal organic frameworks performance in CH4 adsorption using machine learning techniques: A GBRT model based on small training dataset, *J. Environ. Chem. Eng.*, 2023, **11**(3), 110086.
128 A. Ahmed and D. J. Siegel, Predicting hydrogen storage in MOFs via machine learning, *Patterns*, 2021, **2**(7), 100291.
129 B. Han, M. S. Ng and A. Chakraborty, Optimizing metal-organic frameworks for adsorption-based atmospheric water harvesting: A systematic evaluation of pristine and modified MOFs for enhanced performance in diverse climates, *Chem. Eng. J.*, 2025, 166607.
130 B. Han, A. Chakraborty and B. B. Saha, Enhancing Cooling and Atmospheric Water Harvesting via Zeolite-MOF Composites: Experimental and Thermodynamic Evaluation of MIL-160 (Al) and AFI-Type Zeolite Hybrid Adsorbents, *ACS Sustainable Chem. Eng.*, 2025.
131 H. Daglar and S. Keskin, Combining machine learning and molecular simulations to unlock gas separation potentials of MOF membranes and MOF/polymer MMMs, *ACS Appl. Mater. Interfaces*, 2022, **14**(28), 32134–32148.
132 R. Syah, A. Al-Khowarizmi, M. Elveny and A. Khan, Machine learning based simulation of water treatment using LDH/MOF nanocomposites, *Environ. Technol. Innovation*, 2021, **23**, 101805.
133 X. Lu, P. Liu, K. Bisetty, Y. Cai, X. Duan, Y. Wen, Y. Zhu, L. Rao, Q. Xu and J. Xu, An emerging machine learning strategy for electrochemical sensor and supercapacitor using carbonized metal–organic framework, *J. Electroanal. Chem.*, 2022, **920**, 116634.
134 B. Han, T. H. Rupam, A. Chakraborty and B. B. Saha, A comprehensive review on VOCs sensing using different functional materials: Mechanisms, modifications, challenges and opportunities, *Renewable Sustainable Energy Rev.*, 2024, **196**, 114365.
135 Z. Shi, X. Yuan, Y. Yan, Y. Tang, J. Li, H. Liang, L. Tong and Z. Qiao, Techno-economic analysis of metal–organic frameworks for adsorption heat pumps/chillers: From directional computational screening, machine learning to experiment, *J. Mater. Chem. A*, 2021, **9**(12), 7656–7666.
136 B. Han and A. Chakraborty, Recent advances in metal-organic frameworks for adsorption heat transformations, *Renewable Sustainable Energy Rev.*, 2024, **198**, 114411.
137 P. Wang, J. Xing, X. Jiang and J. Zhao, Transition-metal interlink neural network: Machine learning of 2D metal–organic frameworks with high magnetic anisotropy, *ACS Appl. Mater. Interfaces*, 2022, **14**(29), 33726–33733.
138 Y. Wang, L. He, M. Wang, J. Yuan, S. Wu, X. Li, T. Lin, Z. Huang, A. Li and Y. Yang, The drug loading capacity prediction and cytotoxicity analysis of metal–organic frameworks using stacking algorithms of machine learning, *Int. J. Pharm.*, 2024, **656**, 124128.
139 V. C. Anadebe, V. I. Chukwuike, M. A. Chidiebere and R. C. Barik, Synthesis, Characterization, and evaluation of Co-MOF based ZIF-67 for CO2 corrosion inhibition of X65 steel: insights from electrochemical studies and a machine learning algorithm, *J. Phys. Chem. C*, 2023, **127**(20), 9871–9886.
140 L. Zhao, F. Xiao, X. Zeng, Z. Huang, W. Fang, X. Du, X. He, W. Li, D. Wang and H. Chen, Synergistic Effects of MHD Dynamics and Oxygen Vacancies on Electrode Polarization in Photoelectrocatalysis CO_2 Reduction Systems in Exploration, 2025, Wiley Online Library.
141 B. Han, J. Liu, C. Lee, C. Lv and Q. Yan, Recent advances in metal-organic framework-based nanomaterials for electrocatalytic nitrogen reduction., *Small Methods*, 2023, **7**(9), 2300277.
142 K. E. McCullough, D. S. King, S. P. Chheda, M. S. Ferrandon, T. A. Goetjen, Z. H. Syed, T. R. Graham, N. M. Washton, O. K. Farha and L. Gagliardi, High-throughput experimentation, theoretical modeling, and human intuition: lessons learned in metal–organic-framework-supported catalyst design, *ACS Cent. Sci.*, 2023, **9**(2), 266–276.
143 X. Bai, Y. Li, Y. Xie, Q. Chen, X. Zhang and J.-R. Li, High-throughput screening of CO_2 cycloaddition MOF catalyst with an explainable machine learning model, *Green Energy Environ.*, 2025, **10**(1), 132–138.
144 S. Kwon and Y. Lee, Explainability-based mix-up approach for text data augmentation, *ACM Trans. Knowl. Discov. Data*, 2023, **17**(1), 1–14.
145 Y. Song, J. Li, D. Chi, Z. Xu, J. Liu, M. Chen and Z. Wang, AI-driven advances in metal–organic frameworks: from data to design and applications, *Chem. Commun.*, 2025, **61**(82), 15972–16001.
146 J. Zhou, A. Mroz and K. E. Jelfs, Deep generative design of porous organic cages via a variational autoencoder, *Digital Discov.*, 2023, **2**(6), 1925–1936.
147 H. Park, X. Yan, R. Zhu, E. A. Huerta, S. Chaudhuri, D. Cooper, I. Foster and E. Tajkhorshid, A generative artificial intelligence framework based on a molecular diffusion model for the design of metal-organic frameworks for carbon capture, *Commun. Chem.*, 2024, **7**(1), 21.
148 B. Han and A. Chakraborty, Synthesis and characterization of various MOFs and MOFs-zeolite composites: Water adsorption and rapid cooling production, *Rapid Refrigeration and Water Protection: Next Generation Adsorbents*, 2022, Springer, pp. 29–85.
149 M. Phuong and C. Lampert. Towards understanding knowledge distillation. in International conference on machine learning. 2019. PMLR.
150 J. Tang, R. Shivanna, Z. Zhao, D. Lin, A. Singh, E. H. Chi and S. Jain, Understanding and improving knowledge distillation, *arXiv*, preprint, arXiv:2002.03532, 2020.
151 H. W. B. Teo, A. Chakraborty and B. Han, Water adsorption on CHA and AFI types zeolites: Modelling and investigation of adsorption chiller under static and dynamic conditions, *Appl. Therm. Eng.*, 2017, **127**, 35–45.
152 B. Han, E. Hu, B.-E. Jia, Z. Liu, S. Tso, A. Sumboja, I. T. Anggraningrum, A. Z. Syahrial and Q. Yan, Metal-organic frameworks (MOFs) in aqueous batteries (ABs): unlocking potential through innovative materials design, *Sci. China Chem.*, 2025, 1–35.
153 B. Han and A. Chakraborty, Functionalization, protonation and ligand extension on MIL-53 (Al) MOFs to boost water adsorption and thermal energy storage for heat transformations, *Chem. Eng. J.*, 2023, **472**, 145137.