scientific reports

Check for updates

OPEN

# Accelerating discovery of MOFs for hydrogen storage via machine learning in energy related applications

Saeid Khairandesh[1], Marzieh Lotfi[2]✉, Afsanehsadat Larimi[3]✉, Ali Akbar Asgharinezhad[4]✉ & Cyrus Ghotbi[1]

**Hydrogen is a promising clean energy carrier, but its low energy density necessitates advanced storage solutions. Metal–Organic Frameworks (MOFs) offer high tunability and porosity for efficient hydrogen adsorption. This work combines Grand Canonical Monte Carlo (GCMC) simulations with machine learning, employing Feed-Forward (FNN) and Pattern Recognition (PRNN) neural networks optimized via Equilibrium Optimizer and Genetic Algorithm. The integrated approach predicts gravimetric and volumetric hydrogen storage capacities across 98,695 metal–organic frameworks under temperature–pressure swing conditions. Pore volume and void fraction emerged as dominant structural descriptors. The models identified 12 top-performing MOFs exceeding MOF-5 in both gravimetric (8.27 wt.%) and volumetric (51.94 g-$H_2$/L) capacities, demonstrating the power of ML-accelerated screening for next-generation hydrogen storage materials.**

**Keywords** Hydrogen storage, Metal–organic-framework (MOFs), Artificial neural networks (ANNs), Grand canonical Monte Carlo (GCMC), Temperature–pressure swing conditions

The environmental impact of fossil fuels—particularly their role in greenhouse gas emissions and climate change—has intensified the search for clean energy alternatives. Hydrogen is a promising candidate because it produces zero emissions when used in fuel cells, releasing only water vapor and heat as byproducts[1]. This clean profile aligns with global efforts to reduce carbon footprints and transition toward a sustainable energy future[2].

Despite its promise, hydrogen faces a major obstacle: efficient storage. Under ambient conditions, hydrogen has very low energy density, requiring large volumes to store meaningful amounts of energy. This poses significant challenges for onboard vehicle storage and large-scale applications[3]. Consequently, the development of safe, efficient, and cost-effective hydrogen storage systems remains a critical priority[4].

Several storage strategies are under investigation. Compressed hydrogen gas (CHG) stores $H_2$ at high pressure (typically 350–700 bar) in carbon-fiber-reinforced tanks and is the most common method in current fuel cell vehicles (FCVs)[5]. However, it requires bulky and expensive containment systems. Cryo-compressed hydrogen, which stores $H_2$ as a liquid at cryogenic temperatures (~ 30–80 K) and, pressure (250–350 bar), offers higher volumetric density but demands energy-intensive cooling and specialized infrastructure[6]. Solid-state hydrogen storage—using materials such as metal hydrides, complex hydrides, and metal–organic frameworks (MOFs)—provides a promising alternative by adsorbing hydrogen into porous structures, potentially achieving high storage capacities[7].

MOFs are highly porous, crystalline materials formed by linking metal ions or clusters with organic ligands[8]. This architecture creates a three-dimensional network with extensive internal void space, making MOFs exceptionally suitable for gas storage. Notably, some MOFs exhibit surface areas exceeding 2500 $m^2\ g^{-1}$, as measured by the Brunauer–Emmett–Teller[9] method. By tuning the metal nodes and organic linkers, researchers can engineer MOFs with tailored pore sizes to optimize hydrogen adsorption[2,10]. In fact, structural properties—such as void fraction and pore volume—often influence hydrogen uptake more than chemical composition.

[1]Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran, Iran. [2]Chemical Engineering Department, Jundi-Shapur University of Technology, Dezful, Iran. [3]Department of Chemical Engineering, School of Engineering and Applied Sciences, Swansea University, Wales, UK. [4]Chemistry and Process Research Department, Niroo Research Institute, Tehran, Iran. ✉email: marzyeh.lotfi@gmail.com; a.larimi@swansea.ac.uk; aasgharinezhad@nri.ac.ir

Importantly, MOFs can adsorb and release hydrogen reversibly and rapidly, enabling repeated charge–discharge cycles[11].

Despite this potential, only a small fraction of known MOFs has been synthesized and tested. While over 100,000 MOF structures are cataloged in the Cambridge Structural Database (CSD), fewer than 10% are porous and suitable for gas storage[12]. Synthesis challenges thus limit experimental validation. To address this, computational screening has emerged as a powerful tool to explore the vast space of hypothetical MOFs and identify high-performing candidates without exhaustive synthesis[13].

Grand Canonical Monte Carlo (GCMC) simulations are widely used to predict hydrogen uptake in MOFs[14,15] by statistically sampling adsorption configurations under prescribed thermodynamic conditions. In large-scale screening studies, GCMC provides high-fidelity gravimetric and volumetric uptake data that enable the identification and ranking of high-capacity frameworks based on their structural characteristics[13,16]. These insights support the development of MOFs with optimized pore geometry and functionality for specific storage requirements[17]. However, the computational cost of GCMC becomes prohibitive when evaluating hundreds of thousands of candidate structures, motivating the integration of data-driven surrogate models.

Machine learning (ML) has emerged as a transformative solution. By learning from existing GCMC or experimental data, ML models can predict hydrogen uptake across vast MOF databases in seconds—bypassing costly simulations for each candidate[18]. This accelerates the identification of top-performing materials and guides experimental efforts toward the most viable targets[3].

While diverse machine learning approaches—including Random Forest (RF)[18], XGBoost, and Graph Neural Networks (GNNs)[19]—have been applied to MOF property prediction, they face limitations in this context. Tree-based models (e.g., RF, XGBoost) struggle to generalize across continuous, high-dimensional descriptor spaces[20], while GNNs require atomistic structural data that are often inconsistent or unavailable across large, heterogeneous MOF databases[12,19]. In contrast, multilayer perceptron (MLPs)—specifically Feed-Forward (FNN) and Pattern Recognition (PRNN) variants—leverage standardized, physically interpretable crystallographic descriptors (e.g., pore volume, void fraction) and excel at modeling smooth, differentiable structure–property relationships at scale[11,18]. Their simplicity, robustness, and compatibility with high-throughput screening make MLPs particularly well-suited for predicting deliverable hydrogen storage capacities across tens of thousands of MOFs[3,15,21].

In this study, computational screening and deep learning are integrated to predict both gravimetric and volumetric hydrogen storage capacities under temperature–pressure swing[22] conditions. Figure 1 outlines the overall workflow of this study, encompassing database curation, crystallographic feature selection, neural network implementation, capacity prediction, and model evaluation. The study leverages a dataset of 98,695 MOFs and deploys two complementary neural architectures—Feed-Forward (FNN) and Pattern Recognition (PRNN) networks—optimized via the Equilibrium Optimizer (EO). Our models are calibrated against GCMC

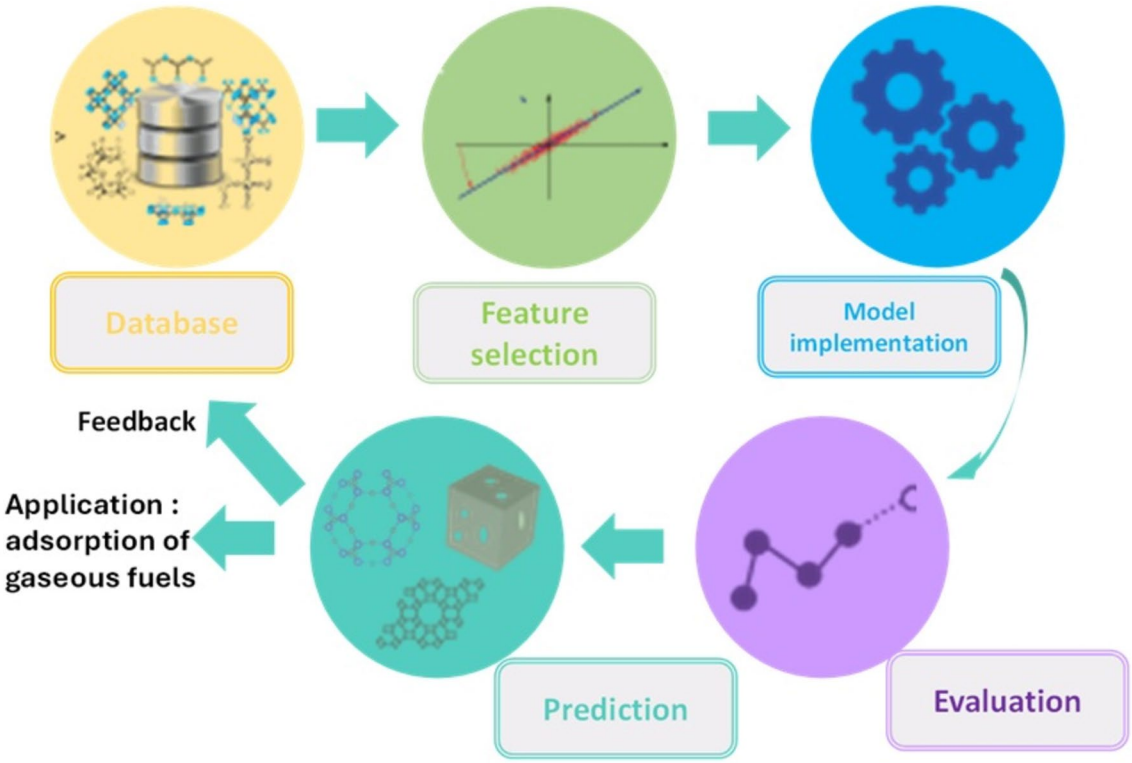


**Fig. 1**. General schematic of this study.

simulations to ensure physical fidelity. Ultimately, this framework enables rapid, accurate, and interpretable screening of MOFs, accelerating the discovery of viable materials for real-world hydrogen storage applications.

## Method

### Database and dataset preparation

This study leverages the Hydrogen Materials Advanced Research Consortium (HyMARC)[23] database encompassing 98,695 MOFs initially compiled by Ahmed et al.[18]. Each MOF entry includes structural characteristics such as gravimetric and volumetric surface areas (GSA, VSA) $m^2/g$, $m^2/cm^3$ respectively, pore volume (PV) $cm^3/g$, density (D) $g/cm^3$, void fraction (VF), largest cavity diameter (LCD) Å, and pore-limiting diameter (PLD) Å. The target properties predicted by our machine learning models are the usable gravimetric hydrogen capacity (UG, wt.%) and usable volumetric hydrogen capacity (UV, g-$H_2$/L) under temperature–pressure swing conditions[22] (loading at 77 K and 100 bar; delivery at 160 K and 5 bar). Usable capacity is defined as the difference in adsorbed hydrogen between the loading and delivery states, reflecting deliverable storage relevant to real-world applications[5,14]. These detailed crystallographic properties were calculated using the zeo++ code with a probe radius of 1.86 Å which corresponds to the kinetic diameter of $H_2$ and is the conventional choice for hydrogen-related porosity characterization in computation-ready MOF databases[24]. This standardized probe radius ensures consistency with high-throughput screening protocols and aligns theoretical porosity metrics with experimental adsorption behavior under cryogenic conditions[25–27]. The crystallographic properties of MOFs are well established to govern hydrogen physisorption behavior[18]. To ensure physical interpretability, computational efficiency, and compatibility with high-throughput screening, the input features were restricted to a minimal yet representative set of seven intrinsic structural descriptors: density (D), pore volume (PV), gravimetric surface area (GSA), volumetric surface area (VSA), void fraction (VF), largest cavity diameter (LCD), and pore-limiting diameter (PLD). These properties are directly computable from a MOF's CIF file in seconds using open-source tools like zeo++, and are universally recognized as first-order descriptors that encode critical information about pore geometry, surface accessibility, and packing density—key factors controlling hydrogen uptake[1,7,10,11,20,28–32].

For the development of the Artificial Neural Network (ANN) model, the dataset comprising 98,695 MOF structures was randomly shuffled prior to partitioning to eliminate any ordering bias inherent in the original database. It was then split into training (70%), validation (15%), and test (15%) subsets using MATLAB's dividerand function with a fixed random seed to ensure reproducibility. approximately 69,086 MOFs were allocated to the training set. The remaining data were evenly divided between the validation and test sets, with each comprising approximately 14,804 MOFs. Since this study addresses a regression task—predicting continuous gravimetric and volumetric hydrogen capacities—stratified sampling was not applied, as it is primarily designed for classification problems with discrete labels. Each MOF corresponds to a unique, independent entry; verification of sample indices confirmed no overlap between subsets, thereby preventing data leakage. Furthermore, all input features were normalized using mean and standard deviation computed only from the training set, and these same normalization parameters were applied to the validation and test sets to preserve data integrity.

The dataset incorporates 61,250 MOFs sourced from the University of Ottawa, 578 entries from the University of Michigan, 20,156 structures from the Northwestern University Library, and 5047 MOFs from the CoRE database (Fig. 2). Hydrogen storage capacities—both gravimetric and volumetric—were computed using Grand Canonical Monte Carlo (GCMC) simulations conducted via the RASPA software package[33].These simulations were performed under varying thermodynamic conditions, with temperatures ranging from 77 to 160 K and pressures between 5 and 100 bar. Framework atoms were described using the Universal Force Field (UFF)[34], while hydrogen molecules were modeled using a Lennard–Jones potential with Feynman–Hibbs quantum corrections to account for nuclear quantum effects at cryogenic temperatures[35,36]. MOF–$H_2$ and $H_2$–$H_2$ interactions employed Lorentz–Berthelot mixing rules[37] with a cutoff radius of 12.8 Å, and long-range corrections were applied to account for truncated interactions. To avoid finite-size effects, unit cells with lattice parameters smaller than 24 Å were replicated in all directions following established protocols. Each simulation consisted of 20,000 Monte Carlo cycles, with the first 10,000 cycles used for equilibration and the remaining cycles for adsorption averaging. Translation, insertion, and deletion moves were attempted with equal probability. Structural and chemical descriptors for each MOF were derived based on prior estimations provided by Ahmed et al.[18], serving as the basis for their characterization.

### Neural network model architecture

Two neural network models were employed to predict hydrogen storage capacities in MOFs: a Feed-Forward Neural Network (FNN) and a Pattern Recognition Neural Network (PRNN).

The FNN serves as a general-purpose regressor. It consists of an input layer (seven crystallographic descriptors), three fully connected hidden layers (13–25–30 neurons), and an output layer with two neurons, predicting usable gravimetric (UG) and volumetric (UV) hydrogen capacities jointly. The FNN uses hyperbolic tangent sigmoid activation functions in hidden layers and a linear activation in the output layer. It incorporates Layer Normalization (LayerNorm) to stabilize training and capture complex, hierarchical input–output relationships. This deeper architecture is well suited for modeling the near-linear dependence of $U_G$ on descriptors like pore volume and void fraction.

The PRNN, in contrast, adopts a shallower but wider topology with two hidden layers (29–26 neurons) and the same input/output structure. While it uses identical activation functions and training protocols as the FNN, its reduced depth and increased per-layer width are designed to emphasize global pattern recognition. This makes the PRNN particularly effective at capturing the non-monotonic, saturating trends observed in volumetric capacity ($U_V$) under TPS conditions.

Both models are trained using backpropagation with the mean squared error (MSE) loss function (Eq. 1):

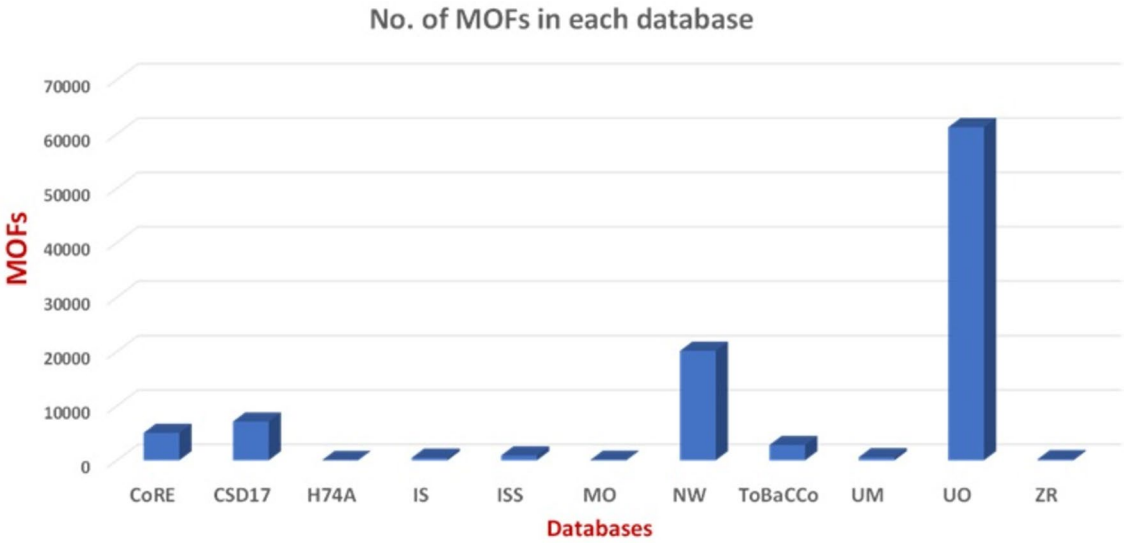


**Fig. 2**. Number of MOFs in each database used for the study. The largest contributions come from the University of Ottawa (UO) and Northwestern (NW) databases, which collectively provide the majority of MOFs, highlighting their significance as key sources for high-performing MOFs. Smaller contributions are observed from databases such as CSD17 and CoRE, which represent real MOF datasets, as well as ToBaCCo, UM, and ZR, reflecting the diversity of the datasets employed in the analysis.

$$\boldsymbol{MSE} = \frac{1}{\boldsymbol{N}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} (\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y}_{\boldsymbol{i}})^2 \tag{1}$$

where $\hat{y}_i$ denotes the actual (GCMC-calculated) hydrogen capacity, $y_i$ is the model prediction, and $N$ is the number of samples. Further architectural and training details are provided in Sect. 2 of the Supplementary Information. Figure 3 shows the simple MLP model.

## Optimization algorithm

Meta-heuristic optimization algorithm, Equilibrium Optimizer (EO) enhance the neural networks' performance by fine-tuning network parameters to reduce prediction error and improve convergence.

*Equilibrium optimizer (EO)*

EO is a physics-based optimization technique inspired by control volume mass balance models. It optimizes solutions by simulating dynamic equilibrium processes, where particles, representing parameter configurations, iteratively update their positions to converge towards an equilibrium or optimal state[38]. EO incorporates a generation rate enabling it to adjust step sizes dynamically for particles to avoid local optima and maintain diverse solutions. The balancing factor modulates particle movement, ensuring convergence towards the optimal solution while preserving exploration capacity. The EO continues iterating until the average position change of particles falls below a threshold, ensuring robust convergence with minimal error. For more detail refer to the supplemental information. Figure 4 illustrates the main objective of optimizer in constitution of artificial neural network.

## Neural network training and optimization procedure

The overall process for predicting hydrogen storage capacities in MOFs using neural networks followed a structured pipeline (Fig. 5). Initially, a search space was defined, encompassing possible neural network configurations with 1–3 hidden layers, 1–30 neurons per layer, and three different transfer functions (Elliotsig, Sigmoid, Tanh). Subsequently, an architecture optimization phase was conducted using the Equilibrium Optimizer (EO), a meta-heuristic algorithm. In this step, EO generated candidate architectures, evaluated their prediction performance on validation data, and selected the most promising configuration based on minimum mean squared error (MSE).

Once the optimal architecture was identified, the neural network (either FNN or PRNN) was trained using Stochastic Gradient Descent (SGD) algorithms. During this stage, the network's weights and biases were updated iteratively to minimize the loss function (Eq. 1), defined as the average squared difference between predicted and actual hydrogen capacities.

To enhance prediction accuracy, the EO algorithm was applied again, this time to refine the neural architecture and hyperparameters, iteratively improving the model's structure and convergence until minimal performance improvement was observed. A final training phase followed, in which the best architecture underwent fine-tuning of weights and biases to reduce residual error and ensure optimal generalization.

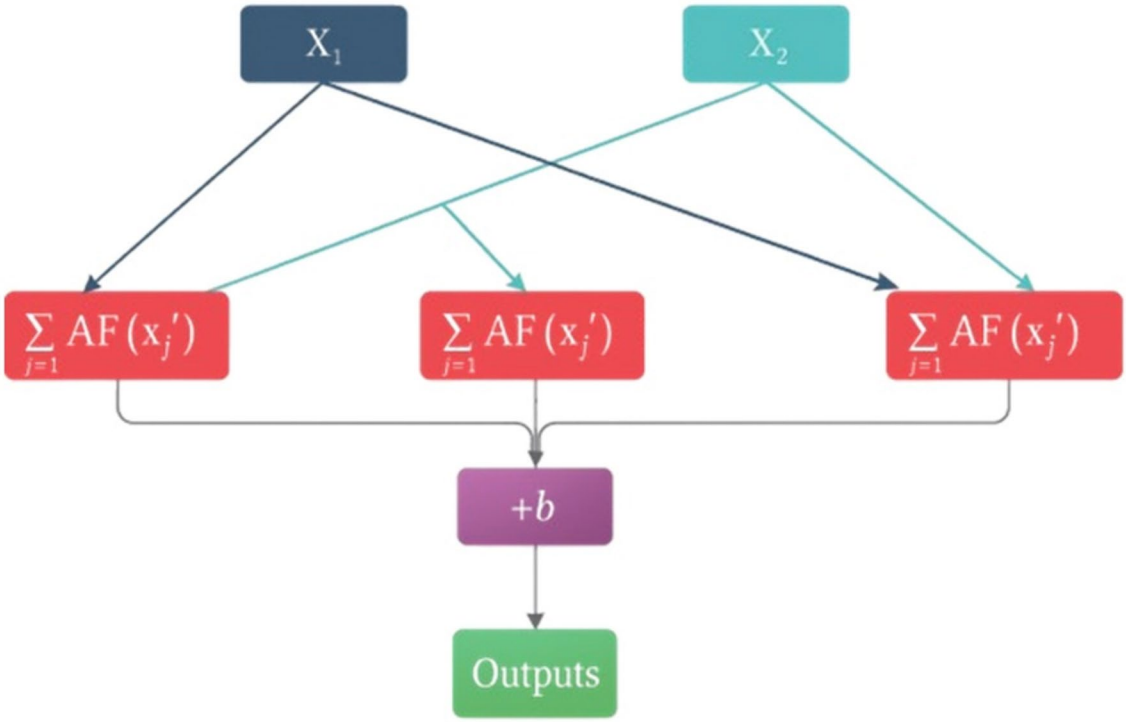


**Fig. 3**. A simple MLP model with one hidden layer.

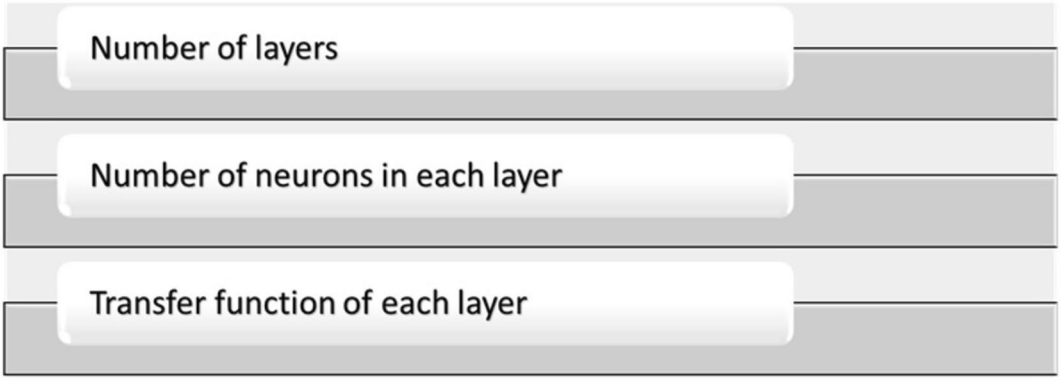


**Fig. 4**. Role of optimizer in the ANN architecture.

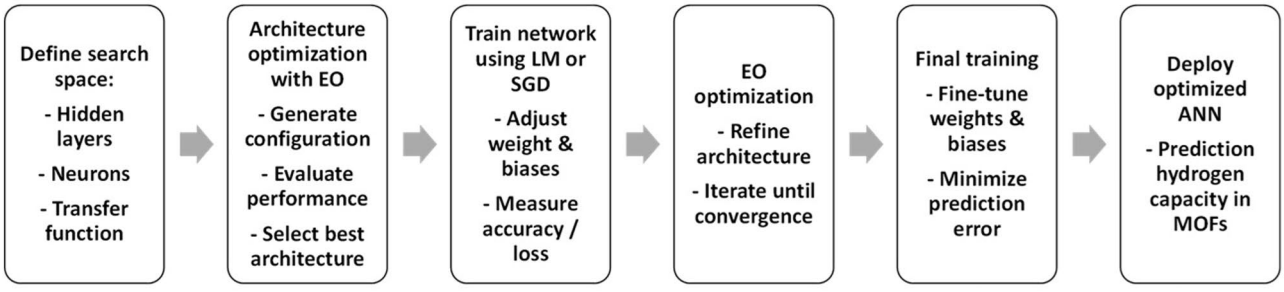


**Fig. 5**. Workflow of the EO-based artificial neural network (ANN) development for predicting hydrogen storage capacities in MOFs.

The optimized ANN model was then deployed to predict two hydrogen storage capacities under TPS conditions (UG and UV), based solely on seven crystallographic descriptors. Both FNN and PRNN models incorporated activation functions to capture non-linearities and employed Layer Normalization for training stability.

## Results & discussion

### Evaluating ML algorithm

This study utilizes two neural network models Feed-Forward Neural Network (FNN) and Pattern Recognition Neural Network (PRNN) to predict hydrogen storage in MOFs based on gravimetric (UG) and volumetric (UV) capacities. These models were evaluated using key metrics, Root-Mean-Square Error (RMSE) and Coefficient of Determination ($R^2$), to compare their predictive performance and reliability. It provides a measure of how well the predicted values generated by a model match the observed or actual calculated values, such as those derived from Grand Canonical Monte Carlo (GCMC) simulations. Table 1 summarizes the performance of the ANN algorithm in further detail. Coefficient of Determination ($R^2$), ranging from 0 to 1, represents the predictive power or "goodness of fit" of the model[3,10]. Root Mean Square Error (RMSE) provides a concrete physical error margin in the same units as the property being measured (e.g., wt.% or g-$H_2$/L)[18]. More details have prepared in the supplemental information Sect. S4 accuracy metric.

### Comparative analysis of FNN and PRNN

As shown in Table 1, both models achieve nearly identical $R^2$ values for UV prediction ($R^2 = 0.911$), but PRNN yields a marginally lower RMSE (3.144 g-$H_2$/L) compared to FNN (3.203 g-$H_2$/L), indicating its slight advantage in capturing volumetric storage trends. The high $R^2$ values obtained for both models indicate that the dominant relationships between MOF structural descriptors and hydrogen storage capacities are well captured. For gravimetric uptake (UG), $R^2$ values close to 0.99 and low RMSE (< 0.4) demonstrate strong predictive reliability, indicating that UG can be accurately modeled using the selected descriptors[11,29]. In contrast, volumetric uptake (UV) shows higher RMSE values (~ 3.2) despite maintaining relatively high $R^2$ (~ 0.91), indicating that while overall trends are well reproduced, local prediction errors remain more pronounced. This reflects the greater complexity of volumetric storage, which depends on multiple interacting structural factors within MOFs[15,39–41].

A comparison between models reveals that while FNN performs marginally better for UG, PRNN achieves a slightly lower RMSE for UV. Although the FNN and PRNN use identical descriptors and training protocols, their architectural differences introduce distinct inductive biases. Gravimetric hydrogen capacity is primarily governed by strongly correlated, monotonic features such as pore volume, void fraction, and gravimetric surface area, which are effectively captured by the deeper hierarchical structure of the FNN. In contrast, volumetric capacity depends on nonlinear trade-offs among density, volumetric surface area, and porosity, leading to non-monotonic optimal regimes. The shallower but wider PRNN better captures these global, coupled feature interactions, resulting in slightly improved volumetric predictions. This highlights that architectural bias, rather than input selection alone, governs model performance across different hydrogen storage metrics[28,29,42–47].

### Univariate feature importance

To assess the individual contribution of crystallographic descriptors to hydrogen storage capacity, a univariate feature importance analysis was performed using Pearson's correlation coefficient (r)[48,49] as an interpretive tool rather than for feature selection[1]. Figure 6 examines the capacity-property trend, conducted across approximately 98,695 MOFs evaluated via GCMC simulations, provides physical insight into structure–property relationships and validates the relevance of the selected input features. As summarized in Table 2, pore volume (PV) $cm^3/g$, gravimetric surface area (GSA) $m^2/g$, and void fraction (VF) exhibit strong positive correlations with both usable gravimetric (UG) and volumetric (UV) hydrogen capacities, indicating that increases in accessible porosity and some ranges of gravimetric surface area (4500–5000 $m^2/g$) generally enhance storage performance. In contrast, single-crystal density (D) exhibits a non-monotonic relationship with usable hydrogen capacities: storage is maximized at a density 'sweet spot' of approximately 0.6 g $cm^{-3}$, and further reductions in density lead to diminishing returns (Fig. 6)[3,37]. This optimal density reflects a balance between two competing effects governing volumetric hydrogen storage. At high densities, limited accessible pore volume restricts adsorption, whereas at very low densities the framework contains insufficient adsorption sites per unit volume despite high porosity. Consequently, volumetric capacity is maximized at an intermediate density where pore accessibility and framework packing efficiency are optimally balanced, consistent with prior theoretical and computational studies[28,50]. Similarly, volumetric surface area (VSA) correlates positively with volumetric capacity only within an intermediate range (~ 1600–2200 $m^2$ $cm^{-3}$), highlighting the complexity of volumetric packing effects. Notably, PV and VF consistently emerge as the most influential descriptors across operating conditions and

| | FNN | | PRNN | |
|---|---|---|---|---|
| $H_2$ capacity type | $R^2$ | RMSE | $R^2$ | RMSE |
| UG at TPS | 0.994 | 0.353 | 0.991 | 0.364 |
| UV at TPS | 0.911 | 3.203 | 0.911 | 3.144 |

**Table 1**. Performance of ANN algorithms in predicting UG and UV $H_2$ capacities of MOFs under TPS condition. $R^2$, RMSE represent coefficient of determination, and root-mean-square error. UG: usable gravimetric hydrogen capacity (wt.%); UV: usable volumetric hydrogen capacity (g-$H_2$/L).

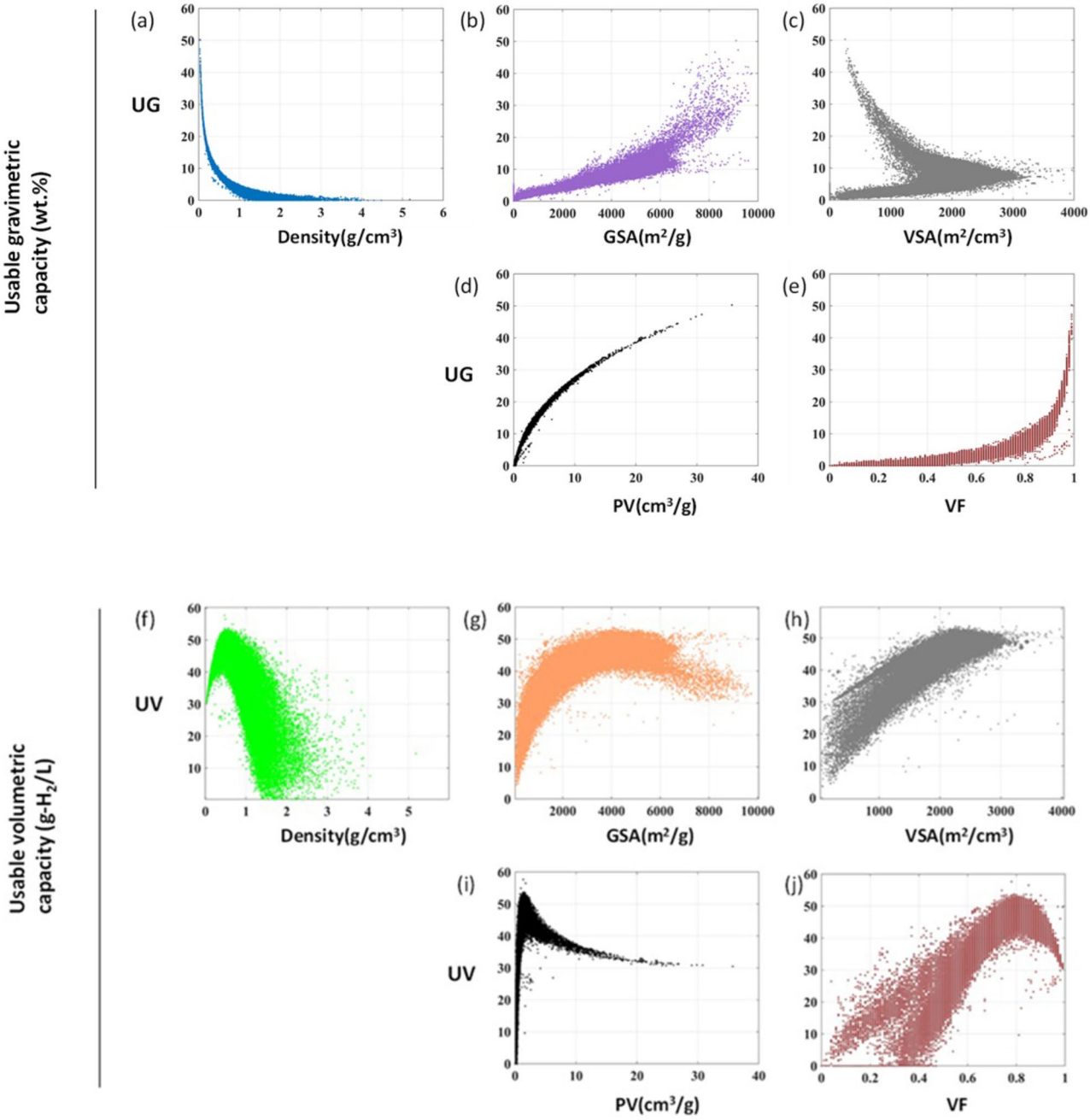


**Fig. 6**. The usable capacities of 98,695 MOFs are presented as a function of five crystallographic properties, under the assumption of TPS condition between 100 bar/77 K and 5 bar at 160 K. Panels (**a**, **b**, **c**, **d**, **e**) illustrate gravimetric capacities, while panels (**f**, **g**, **h**, **i**, **j**) show volumetric capacities. (GSA: gravimetric surface area; VSA: volumetric surface area; PV: pore volume; VF: void fraction); UG: gravimetric capacity; UV: volumetric capacity.

capacity metrics (Table 2, Fig. 6), in agreement with prior studies by Ahmed et al.[18] and the empirical Chahine rule[29,51], which establishes a linear relationship between pore volume and gravimetric hydrogen uptake[37,52]. Importantly, all seven descriptors were retained in subsequent modeling to preserve physical interpretability and ensure compatibility with high-throughput screening workflows, with the univariate analysis serving as a sanity check against known adsorption behavior rather than a basis for model pruning.

## Equilibrium optimizer performance

The EO's primary function was to systematically tune the neural network's architecture, including the number of hidden layers, the number of neurons in each layer, and the selection of transfer functions, to minimize the Root Mean Square Error (RMSE). To rigorously demonstrate the added value of this optimization, a direct

| Pearson correlation coefficient (r) | Usable grav. capacity (wt.%) | Usable vol. capacity (g-$H_2$/L) |
|---|---|---|
| Density (g/cm$^3$) | **−0.812** | **−0.818** |
| Grav. surface area (m$^2$/g) | **0.908** | **0.776** |
| Vol. surface area (m$^2$/cm$^3$) | **0.465** | **0.890** |
| Pore volume (cm$^3$/g) | **0.909** | **0.330** |
| Void fraction | **0.832** | **0.878** |

**Table 2**. The relative significance of five features in predicting hydrogen storage performance in MOFs.; LCD and PLD are omitted due to their minimal standalone predictive contribution (see Fig. S2). An r value of + 1 indicates a perfect positive linear correlation. An r value of −1 signifies a perfect negative linear correlation. An r value of 0 suggests no linear correlation. LCD: largest cavity diameter (Å); PLD: pore-limiting diameter (Å).

| Model configuration | Target | RMSE | $R^2$ | RMSE reduction (%) |
|---|---|---|---|---|
| Non-optimized (baseline) | UG | 0.397 | 0.907 | – |
| | UV | 3.669 | 0.856 | – |
| EO-optimized (proposed) | UG | 0.364 | 0.991 | 8.46% |
| | UV | 3.144 | 0.911 | 14.33% |

**Table 3**. Comparative performance of optimized vs. non-optimized models. UG: usable gravimetric hydrogen capacity (wt.%); UV: usable volumetric hydrogen capacity (g-$H_2$/L).

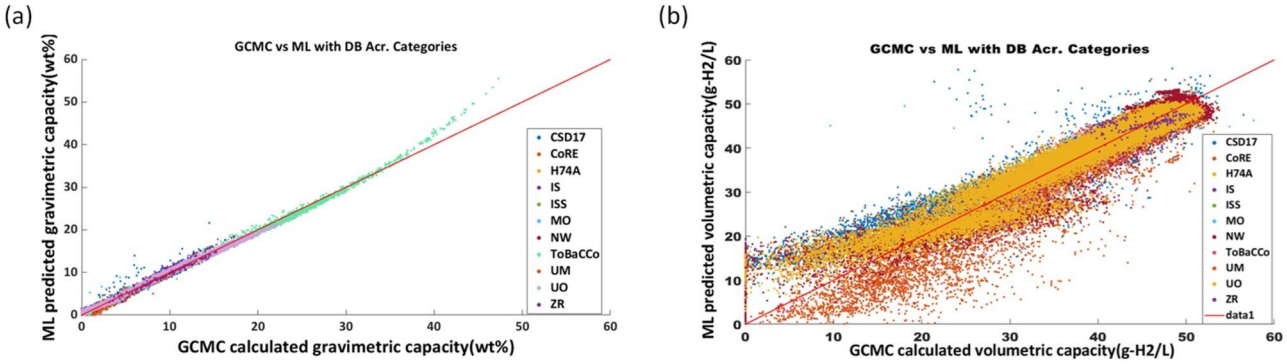


**Fig. 7**. Presents PRNN predictions with GCMC calculated, which are shown as representative since FNN and PRNN exhibit nearly identical volumetric prediction dispersion and comparable $R^2$ values, with only marginal differences in RMSE (Table 1). (**a**) Usable gravimetric hydrogen capacity (**b**) Usable volumetric hydrogen capacity. Colors indicate different MOF databases (e.g., CoRE, CSD17), highlighting model performance across diverse MOF datasets. DB Acr. Refer to the Database Acronyms.

comparison was performed against a non-optimized baseline model. The baseline model was configured with a typical, yet arbitrary, architecture (e.g., two hidden layers with 15 neurons each and standard transfer functions) that falls within the EO's search boundaries but lacks the benefit of metaheuristic tuning. The results of this comparative analysis, presented in Table 3, unequivocally confirm the necessity of the EO algorithm. As shown, the EO-optimized model achieved a significant reduction in error, with the RMSE for the UV prediction decreasing by 14.33% and the $R^2$ value increasing to 0.991. This improvement is a direct consequence of the EO's ability to precisely locate the optimal hyperparameter combination (e.g., 2 hidden layers with 37 and 11 neurons, respectively, and a specific mix of transfer functions) that maximizes model stability and predictive accuracy. The use of the Equilibrium Optimizer thus moves the model from a state of acceptable performance to one of high-fidelity prediction, validating its selection for this critical hydrogen storage modeling task.

## Evaluating model accuracy using GCMC comparison

To validate the predictive fidelity of the ML models, their outputs were compared against independent Grand Canonical Monte Carlo (GCMC) simulations across all 98,695 MOFs (Fig. 7). GCMC simulations serve as validated surrogates for experimental adsorption data in high-throughput studies, as they reproduce measured $H_2$ isotherms for benchmark MOFs with high accuracy under TPS conditions[3,18]. TPS condition.

here refers to pressure Temperature swing; loading with hydrogen at a cryogenic temperature of 77 K and a high pressure of 100 bar, during the delivery cycle, hydrogen is released until the system reaches a discharge

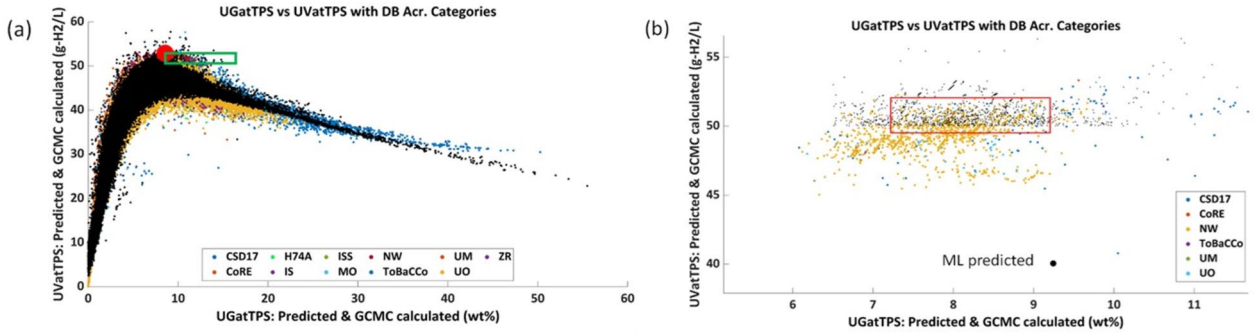


**Fig. 8**. Comparative analysis of the UGatTPS and UVatTPS predicted and GCMC calculated data, as visualized in Figure (**a**), shows that the model's predictions (indicated by black points) closely align with the original calculated results from various databases (color-coded). Bigger red circle is presents MOF-5 in the figure. (**b**) The Northwestern (NW) database, represented by yellow points, demonstrated the highest consistency with the ML-predicted MOFs, indicating the reliability and accuracy of predictions in this dataset compared to others highlights the top 1289 metal–organic frameworks (MOFs) that were identified to surpass the defined performance thresholds of 6.5 wt.% for gravimetric capacity and 50 g-$H_2$/L for volumetric capacity. This subset of MOFs confirms the potential of the Northwestern database's contributions, among others, in housing high-capacity materials. The distribution underlines a significant clustering within these high-performance categories, with predictive modeling matching the original values and showcasing the strong predictive capabilities of the employed machine learning algorithms. UGatTPS: usable gravimetric capacity (wt.%); UVatTPS: usable volumetric capacity (g-$H_2$/L) both at TPS (pressure Temperature swing) condition.

| Name | Source | Density (g/cm$^3$) | Grav. surf. area (m$^2$/g) | Vol. surf. area (m$^2$/cm$^3$) | Void fraction | Pore volume (cm$^3$/g) | Largest cavity diameter (Å) | Pore limiting diameter (Å) | Usable grav. capacity (wt.%) GCMC ML | Usable vol. capacity (g-$H_2$/L) GCMC ML |
|---|---|---|---|---|---|---|---|---|---|---|
| TPS condition | | | | | | | | | | |
| SUKYON | CoRE | 0.53 | 5130 | 2701 | 0.85 | 1.47 | 10.8 | 7.3 | 8.85 8.96 | 51.96 51.81 |
| hMOF_cat_1 | NW | 0.57 | 5206 | 2979 | 0.80 | 1.39 | 7.0 | 8.1 | 8.27 8.37 | 51.94 51.69 |
| NASZOW | CSD17 | 0.58 | 5264 | 3073 | 0.78 | 1.33 | 5.6 | 4.8 | 8.15 8.06 | 52.25 51.27 |
| XUFFUA | CSD17 | 0.49 | 7724 | 3808 | 0.78 | 1.58 | 6.0 | 5.4 | 9.47 10.7 | 52.08 54.10 |
| QATDAQ | CSD17 | 0.50 | 5266 | 2612 | 0.76 | 1.54 | 7.2 | 6.2 | 9.35 10.7 | 52.77 56.30 |

**Table 4**. The highest capacity MOFs, as identified by ML and verified by GCMC, under temperature + pressure swing[22] condition. NW refers to the Northwestern database[13]. Grav.; gravimetric, Vol.; volumetric capacities.

pressure of 5 bar while the temperature is simultaneously raised to 160 K[7,15,37]. More explanation is provided in the supplemental information Sect. 5.1.1.

The results show near-perfect agreement for gravimetric capacity (UG), with a parity slope of 0.998 and $R^2 = 0.994$, confirming that the model accurately captures the dominant linear trends driven by pore volume, void fraction, and gravimetric surface area (Fig. 7a). For volumetric capacity (UV), agreement remains strong (slope = 0.982, $R^2 = 0.911$), though with greater scatter—consistent with UV's heightened sensitivity to structural trade-offs (Fig. 7b).

The largest discrepancies between ML and GCMC predictions occur primarily in real MOF datasets (e.g., CoRE and CSD; red and blue points in Fig. 7b). These deviations could stem from several sources for instance model overfitting [37], inaccuracies in GCMC simulations[53,54], solvent removal [27,55], limitations of volumetric capacity prediction[56], areas where future work could refine both simulation and prediction.

## Top-performing MOFs

For TPS conditions, MOF-5 (7.8 wt.%, 51.9 g-$H_2$/L)[18,28,37]serves as the primary performance benchmark. In this study, a "top-performing" MOF is defined as one that simultaneously exceeds both the U.S. Department of Energy's 2050 targets (6.5 wt.%, 50 g-$H_2$/L)[7] and the capacity of MOF-5.

Figure 8a illustrates the strong agreement between ML-predicted (black points) and GCMC-calculated (colored points) capacities across all databases. From the full dataset of 98,695 MOFs, the model identified 1289 candidates meeting or exceeding the 6.5 wt.% and 50 g-$H_2$/L thresholds, highlighted in the green rectangle in Fig. 8a among these, 12 MOFs were confirmed via GCMC simulation to surpass MOF-5's performance—validating the model's ability to prioritize high-capacity materials[57]. The top five are summarized in Table 4.

Notably, most of these top performers originate from the Northwestern University (NW) database (yellow points in Fig. 3b), suggesting its prominence in hosting high-capacity frameworks[13].

In Fig. 8b, the high-capacity MOFs within the red rectangle—those showing strongest alignment with ML predictions—exhibit consistent structural features: gravimetric surface areas of 2000–4000 $m^2/g$, void fractions of 0.5–0.85, pore volumes of 0.5–1.2 $cm^3/g$, and an average density of ~ 0.6 $g/cm^3$. This suggests.

that while high surface area enhances gravimetric uptake (beneficial for pressure-swing applications[18]), it often compromises volumetric capacity under TPS conditions. Optimal UV performance arises at intermediate surface areas, where porosity and packing efficiency are balanced—a key insight for rational MOF design[50].

Based on the dominance of the Northwestern (NW) database[58] in the top 1289 candidates (Fig. 9), the study identified hypothetical.

MOF_5026659_i_0_j_28_k_1_m_0_cat_1 as the highest-capacity predicted MOF under TPS conditions, with GCMC-calculated capacities of 8.27 wt.% and 51.94 g-$H_2$/L—exceeding MOF-5 and positioning it as a leading candidate for experimental synthesis.

As no Crystallographic Information File (CIF) exists for this exact structure, Fig. 10 presents a closely related NW MOF with comparable properties (void fraction, surface area, density), the values for which are shown in Table 5. As no CIF (Crystallographic Information File) exists for this exact structure, a closely related NW MOF with comparable properties (void fraction, surface area, density) is presented in. Its structure features highly symmetric Zn-based nodes and carboxylate linkers, promoting uniform porosity and enhanced $H_2$ uptake—hallmarks of high-performing MOFs in this class[59–62].

## Limitation and feature consideration

While our study successfully identified high-capacity MOFs through machine learning predictions, certain limitations must be acknowledged. A key challenge lies in the synthetic feasibility of these materials, especially those derived from hypothetical datasets. In some cases, real MOFs—despite appearing theoretically promising—may suffer from structural instabilities such as framework collapse during activation, which can significantly reduce their hydrogen storage performance. These issues may be even more pronounced in the case of selected hypothetical MOFs[15,18,37]. Advances in synthesis techniques may mitigate these challenges, making previously impractical structures achievable in the future. Our ML models, while powerful, do not differentiate between MOFs with non-defective crystal structures and those that may have imperfections or unrealistically modeled features. Virtual solvent removal or inaccuracies in reported structures can result in partial occupancies or symmetry disorders, leading to erroneous predictions for some candidates[29]. This emphasizes the importance of post-prediction validation through GCMC simulations and detailed structural inspection to confirm the viability of these promising MOFs.

## Conclusion

This study presents a machine learning framework to accelerate the discovery of high-capacity MOFs for hydrogen storage under TPS conditions. Leveraging a curated dataset of 98,695 MOFs from diverse sources, this work integrates crystallographic descriptors with deep learning models—specifically Feed-forward Neural Networks (FNN) and Pattern Recognition Neural Networks (PRNN)—to predict usable hydrogen storage capacities. The models were optimized using the Equilibrium Optimizer (EO) meta-heuristic algorithm to enhance prediction

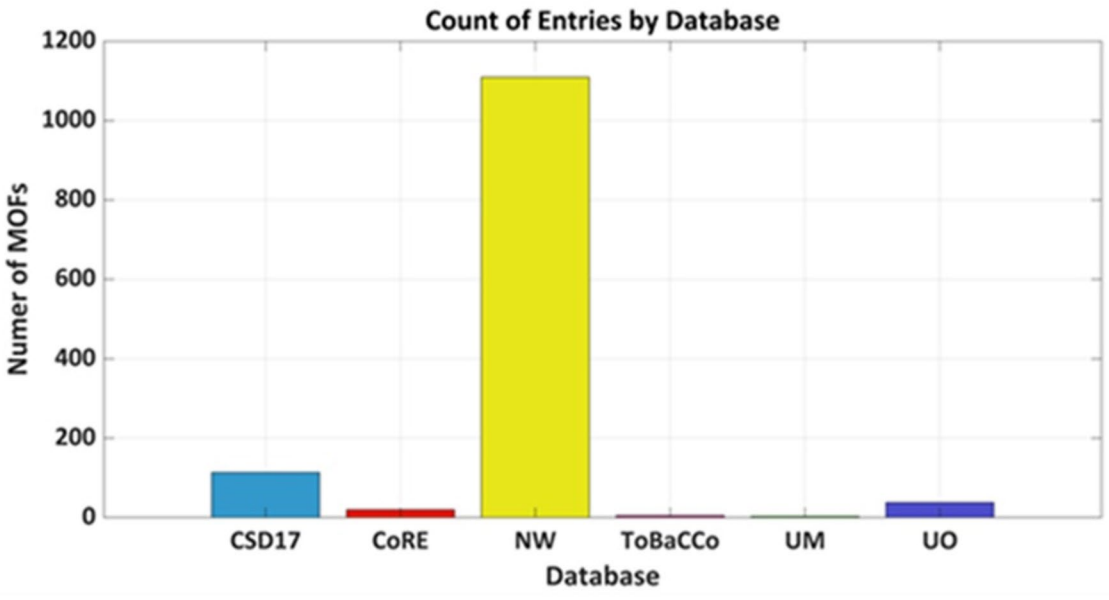


**Fig. 9**. Distribution of predicted MOFs across databases, showing a total of 1289 entries. The majority are sourced from the Northwestern (NW) database, indicating its reliability and prominence as a key source of high-performing MOFs. This suggests that MOFs with superior hydrogen storage capacities are more likely to be found in the NW database compared to others in the figure.

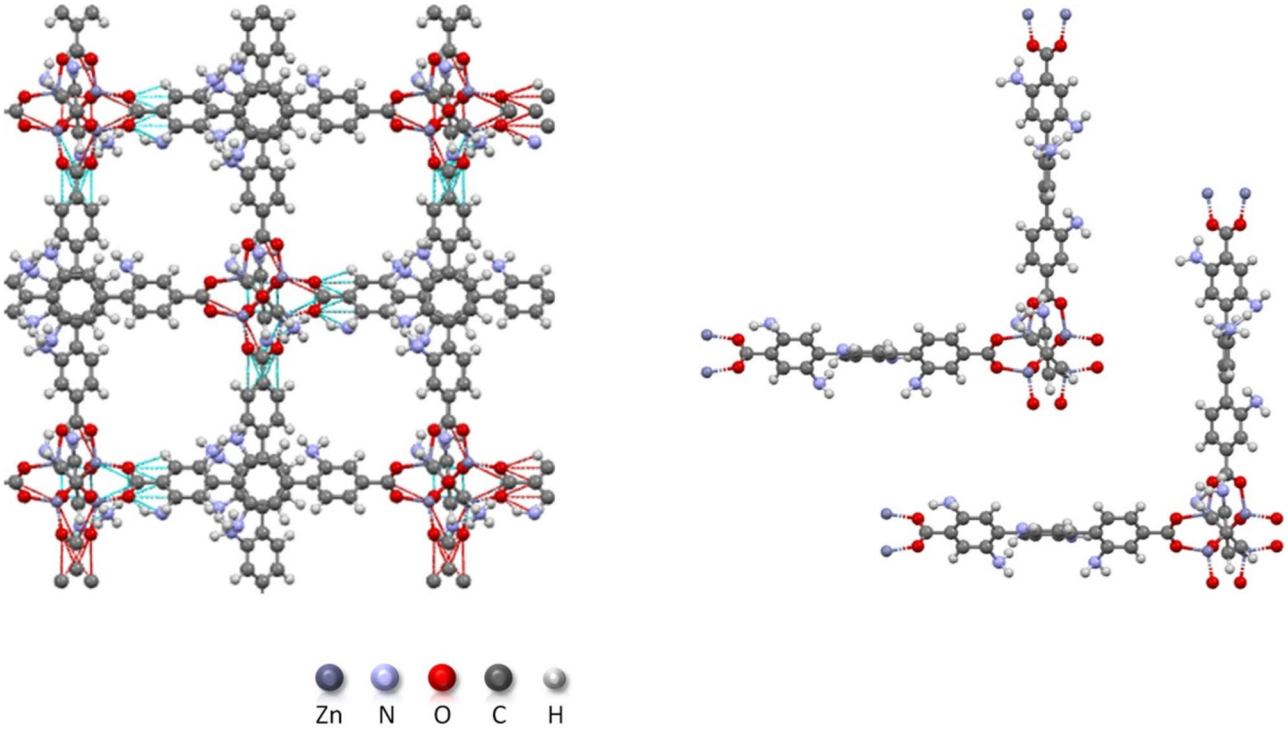


**Fig. 10**. Crystal structures of hMOF_500 Highest-capacity MOFs under TPS conditions. This MOFs originate from the Northwestern University (NW) databases[13] extracted from Free Mercury 4.2.0 that is released by CCDC.

| Name | Source | Density | GSA | VSA | VF | PV | LCD | PLD | UG | | UV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | ML | GCMC | ML | GCMC |
| hmof-500 | NW | 0.48 | 5445.9 | 2625.1 | 0.80 | 1.58 | 10.75 | 6.25 | 8.27 | 8.37 | 51.94 | 51.69 |

**Table 5**. Property of selected MOF from NW database hMOF-500. GSA: gravimetric surface area ($m^2/g$); VSA: volumetric surface area ($m^2/cm^3$); PV: pore volume ($cm^3/g$; VF: void fraction); UG: gravimetric capacity (wt.%); UV: volumetric capacity (g-$H_2$/L); LCD: largest cavity diameter (Å); PLD: pore-limiting diameter (Å).

accuracy and training stability. Trained solely on seven structural features, the neural networks achieved high predictive performance, with the FNN model yielding an $R^2$ of 0.994 for gravimetric capacity and 0.911 for volumetric capacity.

These outcomes underscore the strength of using crystallographic descriptors in combination with neural architectures to approximate Grand Canonical Monte Carlo (GCMC) simulation results. Feature importance analysis confirmed pore volume, void fraction, and surface area as dominant predictors. The models successfully identified 12 MOFs outperforming MOF-5, demonstrating their effectiveness in screening large databases. While challenges remain regarding the synthesis and stability of hypothetical MOFs, this approach offers a scalable path toward efficient materials discovery in energy applications. This work advances the field by integrating physically interpretable ML, metaheuristic optimization, and TPS-based deliverable capacity prediction to prioritize MOFs that are not only high-performing but also synthetically plausible—bridging the gap between computational screening and real-world application.

## Data and code availability

The dataset used in this study is taken from: https://datahub.hymarc.org/dataset/computational-prediction-of-hydrogen-storage-capacities-in-mofs/resource/968256f6-85af-4519-b0bb-c1af7c567592 and the MATLAB codes developed in this study are openly available at: https://github.com/samfkh/MOF-H2-ML-ANN-EO.

## Material availability

This study did not generate new reagent.

Received: 4 November 2025; Accepted: 11 March 2026
Published online: 18 March 2026

## References


1. Schweitzer, B. *Hydrogen Storage in Porous Crystalline Materials: Insights on the Role of Interaction Strength from Simulation and Machine Learning* (Colorado School of Mines, Golden, 2018).
2. Zelenak, V. & Saldan, I. Factors affecting hydrogen adsorption in metal-organic frameworks: A short review. *Nanomaterials (Basel)* **11**, 1638. https://doi.org/10.3390/nano11071638 (2021).
3. Shekhar, S. & Chowdhury, C. Prediction of hydrogen storage in metal-organic frameworks: A neural network based approach. *Results Surf. Interfaces* **14**, 100166. https://doi.org/10.1016/j.rsurfi.2023.100166 (2024).
4. Segakweng, T. *Hydrogen Physisorptive Storage in Metal-Organic Frameworks (MOFs)* (University of Pretoria (South Africa), Pretoria, 2016).
5. Katsumata, S. et al. Experimental and analytical study of a high-pressure hydrogen storage tank made of CFRP with dome–cylinder split molding structure for fuel cell vehicles. *Int. J. Hydrogen Energy* **101**, 269–279 (2025).
6. Moreno-Blanco, J. et al. The storage performance of automotive cryo-compressed hydrogen vessels. *Int. J. Hydrogen Energy* **44**, 16841–16851 (2019).
7. García-Holley, P. et al. Benchmark study of hydrogen storage in metal–organic frameworks under temperature and pressure swing conditions. *ACS Energy Lett.* **3**, 748–754. https://doi.org/10.1021/acsenergylett.8b00154 (2018).
8. Mani-Biswas, M. & Cagin, T. Insights from theoretical calculations on structure, dynamics, phase behavior and hydrogen sorption in nanoporous metal organic frameworks. *Comput. Theor. Chem.* **987**, 42–56. https://doi.org/10.1016/j.comptc.2012.02.023 (2012).
9. Zanobetti, F. et al. Onboard carbon capture and storage (OCCS) for fossil fuel-based shipping: A sustainability assessment. *J. Clean. Prod.* **470**, 143343. https://doi.org/10.1016/j.jclepro.2024.143343 (2024).
10. Sarikas, A. P., Fanourgakis, G. S., Gkagkas, K. & Froudakis, G. E. Comparison of machine learning approaches for the identification of top-performing materials for hydrogen storage. *Sustain. Chem. Environ.* **5**, 100056. https://doi.org/10.1016/j.scenv.2023.100056 (2024).
11. Gómez-Gualdrón, D. A. et al. Understanding volumetric and gravimetric hydrogen adsorption trade-off in metal–organic frameworks. *ACS Appl. Mater. Interfaces* **9**, 33419–33428. https://doi.org/10.1021/acsami.7b01190 (2017).
12. Chong, S., Lee, S., Kim, B. & Kim, J. Applications of machine learning in metal-organic frameworks. *Coord. Chem. Rev.* **423**, 213487. https://doi.org/10.1016/j.ccr.2020.213487 (2020).
13. Wilmer, C. E. et al. Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **4**, 83–89. https://doi.org/10.1038/nchem.1192 (2012).
14. Cabria, I. Grand canonical Monte Carlo simulations of the hydrogen and methane storage capacities of novel but MOFs at room temperature. *Int. J. Hydrogen Energy* **50**, 160–177. https://doi.org/10.1016/j.ijhydene.2023.06.298 (2024).
15. Purewal, J. et al. Estimation of system-level hydrogen storage for metal-organic frameworks with high volumetric storage density. *Int. J. Hydrogen Energy* **44**, 15135–15145. https://doi.org/10.1016/j.ijhydene.2019.04.082 (2019).
16. Altintas, C. & Keskin, S. On the shoulders of high-throughput computational screening and machine learning: Design and discovery of MOFs for $H_2$ storage and purification. *Mater. Today Energy* https://doi.org/10.1016/j.mtener.2023.101426 (2023).
17. Dureckova, H. *Robust Machine Learning QSPR Models for Recognizing High Performing MOFs for Pre-combustion Carbon Capture and Using Molecular Simulation to Study Adsorption of Water and Gases in Novel MOFs* (Université d'Ottawa/University of Ottawa, Ottawa, 2018).
18. Ahmed, A. & Siegel, D. J. Predicting hydrogen storage in MOFs via machine learning. *Patterns* https://doi.org/10.1016/j.patter.2021.100291 (2021).
19. Lu, X., Xie, Z., Wu, X., Li, M. & Cai, W. Hydrogen storage metal-organic framework classification models based on crystal graph convolutional neural networks. *Chem. Eng. Sci.* **259**, 117813 (2022).
20. Mukherjee, K. & Colón, Y. J. Machine learning and descriptor selection for the computational discovery of metal-organic frameworks. *Mol. Simul.* **47**, 857–877 (2021).
21. Wu, X., Wang, Y., Cai, Z., Zhao, D. & Cai, W. Revealing enhancement mechanism of volumetric hydrogen storage capacity of nanoporous frameworks by molecular simulation. *Chem. Eng. Sci.* **226**, 115837 (2020).
22. Fuel OoEEREFFaHa. Cells (Department of Energy accessed 28 Oct hegea-f-f-a-a-f-c. Office of Energy Efficiency & Renewable Energy. 5 Fast Facts about Hydrogen and Fuel Cells). https://energy.gov/eere/articles/5-fast-facts-abouthydrogen-and-fuel-cells
23. Ahmed AD. MOF_Crystallographic_Properties. https://datahub.hymarc.org/dataset/c753c9ea-3bc6-44e1-8880-cfa6e41ce418/resource/cb1f9049-c3fa-4acb-bc6e-bd375ea3c961/download/mof_crystallographic_properties.xlsx (2019)
24. Willems, T. F., Rycroft, C. H., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **149**, 134–141 (2012).
25. Martin, R. L., Smit, B. & Haranczyk, M. Addressing challenges of identifying geometrically diverse sets of crystalline porous materials. *J. Chem. Inf. Model.* **52**, 308–318. https://doi.org/10.1021/ci200386x (2012).
26. Moghadam, P. Z. et al. Development of a Cambridge structural database subset: A collection of metal–organic frameworks for past, present, and future. *Chem. Mater.* **29**, 2618–2625 (2017).
27. Yaghi, O. M., Kalmutzki, M. J. & Diercks, C. S. *Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks* (Wiley, New York, 2019).
28. Ahmed, A. et al. Balancing gravimetric and volumetric hydrogen density in MOFs. *Energy Environ. Sci.* **10**, 2459–2471. https://doi.org/10.1039/C7EE02477K (2017).
29. Goldsmith, J., Wong-Foy, A. G., Cafarella, M. J. & Siegel, D. J. Theoretical limits of hydrogen storage in metal–organic frameworks: Opportunities and trade-offs. *Chem. Mater.* **25**, 3373–3382. https://doi.org/10.1021/cm401978e (2013).
30. Richard, M.-A., Bénard, P. & Chahine, R. Gas adsorption process in activated carbon over a wide temperature range above the critical point. Part 1: Modified Dubinin–Astakhov model. *Adsorption* **15**, 43–51. https://doi.org/10.1007/s10450-009-9149-x (2009).
31. Düren, T., Bae, Y.-S. & Snurr, R. Q. Using molecular simulation to characterise metal–organic frameworks for adsorption applications. *Chem. Soc. Rev.* **38**, 1237–1247. https://doi.org/10.1039/B803498M (2009).
32. Allendorf, M. D., Bauer, C. A., Bhakta, R. & Houk, R. Luminescent metal–organic frameworks. *Chem. Soc. Rev.* **38**, 1330–1352. https://doi.org/10.1039/B802352M (2009).
33. Dubbeldam, D., Calero, S., Ellis, D. E. & Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **42**, 81–101 (2016).
34. Rappé, A. K., Casewit, C. J., Colwell, K., Goddard, W. A. III. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
35. Fischer, M., Hoffmann, F. & Fröba, M. Preferred hydrogen adsorption sites in various MOFs—a comparative computational study. *ChemPhysChem* **10**, 2647–2657 (2009).
36. Colón, Y. J., Fairen-Jimenez, D., Wilmer, C. E. & Snurr, R. Q. High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature. *J. Phys. Chem. C* **118**, 5383–5389 (2014).
37. Ahmed, A. et al. Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. *Nat. Commun.* **10**, 1568. https://doi.org/10.1038/s41467-019-09365-w (2019).
38. Faramarzi, A., Heidarinejad, M., Stephens, B. & Mirjalili, S. Equilibrium optimizer: A novel optimization algorithm. *Knowl. Based Syst.* **191**, 105190. https://doi.org/10.1016/j.knosys.2019.105190 (2020).
39. Pareek, P. & Prema, K. Efficient design of feedforward network for pattern classification. *IOSR J. Comput. Eng. IOSR-JCE* **10**, 78–82 (2013).

40. Benny, D. & Soumya, K.R. Feed-forward neural network processing speed analysis and an experimental evaluation of neural network frameworks. In *2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)* 1–5 (IEEE, 2015). https://doi.org/10.1109/ISCO.2015.7282337
41. Shrivastava, S. & Singh, M. P. Performance evaluation of feed-forward neural network with soft computing techniques for hand written English alphabets. *Appl. Soft Comput.* **11**, 1156–1182. https://doi.org/10.1016/j.asoc.2010.02.015 (2011).
42. Iqbal, A. & Aftab, S. A feed-forward and pattern recognition ANN model for network intrusion detection. *Int. J. Comput. Netw. Inf. Secur.* **14**, 19 (2019).
43. Zhang, J., Li, Y., Li, C., Mei, X. & Zhou, J. Application of soft computing represented by regression machine learning model and artificial lemming algorithm in predictions for hydrogen storage in metal-organic frameworks. *Materials* **18**, 3122 (2025).
44. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).
45. Raghu, M., Poole, B., Kleinberg, J., Ganguli, S. & Sohl-Dickstein, J. On the expressive power of deep neural networks. In *International Conference on Machine Learning: PMLR* 2847–54 (2017).
46. Bengio, Y., Courville, A. & Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).
47. Montúfar, G., Pascanu, R., Cho, K. & Bengio, Y. On the number of linear regions of deep neural networks. In *Advances in Neural Information Processing Systems* 27 (2014).
48. Kokoska, S., Zwillinger, D., Probability, C. S. & Tables, S. *Formulae* (CRC Press, London, 2000).
49. Livas, C. G., Trikalitis, P. N. & Froudakis, G. E. MOFSynth: A computational tool toward synthetic likelihood predictions of MOFs. *J. Chem. Inf. Model.* **64**, 8193 (2024).
50. Allendorf, M. D. et al. An assessment of strategies for the development of solid-state adsorbents for vehicular hydrogen storage. *Energy Environ. Sci.* **11**, 2784–2812. https://doi.org/10.1039/C8EE01085D (2018).
51. Broom, D. P. *Hydrogen Storage Materials: The Characterisation of Their Storage Properties* (Springer, Cham, 2011).
52. Panella, B., Hirscher, M. & Roth, S. Hydrogen adsorption in different carbon nanostructures. *Carbon* **43**, 2209–2214. https://doi.org/10.1016/j.carbon.2005.03.037 (2005).
53. Altintas, C. et al. An extensive comparative analysis of two MOF databases: High-throughput screening of computation-ready MOFs for $CH_4$ and $H_2$ adsorption. *J. Mater. Chem. A* **7**, 9593–9608. https://doi.org/10.1039/C9TA01378D (2019).
54. Chen, T. & Manz, T. A. Identifying misbonded atoms in the 2019 CoRE metal–organic framework database. *RSC Adv.* **10**, 26944–26951. https://doi.org/10.1039/D0RA02498H (2020).
55. Wuttke, S. Introduction to reticular chemistry metal-organic frameworks and covalent organic frameworks Von Omar M. Yaghi, Markus J. Kalmutzki und Christian S. Diercks. *Angew. Chem.* **131**, 6230 (2019).
56. Sturluson, A. et al. The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation*. *Mol. Simul.* **45**, 1082–1121. https://doi.org/10.1080/08927022.2019.1648809 (2019).
57. Makridis, S. Hydrogen storage and compression. arXiv preprint arXiv:170206015. https://doi.org/10.1049/PBPO101E_ch1 (2017).
58. Snurr* NSBSJBCT-ALSHMISWSQ. MOFX-DB: An online database of computational adsorption data for nanoporous materials. https://mof.tech.northwestern.edu/ (2023)
59. Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, 1230444. https://doi.org/10.1126/science.1230444 (2013).
60. Kreno, L. E. et al. Metal–organic framework materials as chemical sensors. *Chem. Rev.* **112**, 1105–1125. https://doi.org/10.1021/cr200324t (2012).
61. Murray, L. J., Dincă, M. & Long, J. R. Hydrogen storage in metal–organic frameworks. *Chem. Soc. Rev.* **38**, 1294–1314. https://doi.org/10.1039/B802256A (2009).
62. Rowsell, J. L. & Yaghi, O. M. Metal–organic frameworks: A new class of porous materials. *Microporous Mesoporous Mater.* **73**, 3–14. https://doi.org/10.1016/j.micromeso.2004.03.034 (2004).


## Acknowledgements

S.Kh. acknowledges Dr. Abtahi, Head of the R&D Department, Energy Transition and Alternative Fuel Sector, for providing financial support to access the HyMARC database, offering guidance, and facilitating research progress through workplace flexibility. The authors also thank Mr. Jamal Miyahi, former Technical and Ship Management Director, for his instrumental role in shaping the research direction.

## Author contributions

All authors contributed to the study conception and design. Data collection and analysis were performed by all authors. The first draft of the manuscript was written by Saeid Khairandesh and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Funding

Not applicable.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-44340-8.

**Correspondence** and requests for materials should be addressed to M.L., A.L. or A.A.A.