Results in Surfaces and Interfaces 14 (2024) 100166

Contents lists available at ScienceDirect

Results in Surfaces and Interfaces

journal homepage: www.elsevier.com/locate/rsurfi

Full length article

Prediction of hydrogen storage in metal-organic frameworks: A neural network based approach

Shivanshu Shekhar [a], Chandra Chowdhury [b,*]

[a] Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[b] Institute of Catalysis Research and Technology (IKFT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany

ARTICLE INFO

Dataset link: https://datahub.hymarc.org/dataset

Keywords:
MOFs
Hydrogen storage
Deep learning
Neural network
GCMC

ABSTRACT

Gas capture, sensing, and storage systems are all within the capabilities of metal–organic frameworks (MOFs). It is common practise to choose the MOF with the best adsorption property from a large database before running an adsorption calculation. High-throughput computational research is sometimes hampered by the expense of computing thermodynamic values, slowing the progress of MOFs for separations and storage applications. When trying to predict material properties, machine learning has recently emerged as a possible alternative to more conventional methods like experiments and simulations. The H_2 capacities of 918,734 MOFs drawn from 19 databases were recently predicted using ML by Ahmed and Siegel (2021). Several ML methods were utilized, and the extremely randomised tree (ERT) model emerged as the most accurate predictor of hydrogen delivery capacity in terms of both gravimetric and volumetric quantities. Interestingly we have used deep learning model (Feed-forward neural network) as well as ERT model for the prediction of H_2 deliverable capacities of a huge number of MOFs developed from the previous studies and got till date best results for predictions. To verify our model's efficacy, we also performed Grand Canonical Monte Carlo (GCMC) simulations. We show our method by forecasting the hydrogen storage capacity of MOFs during a temperature and pressure swing from 100bar/77K to 5bar/160K.

1. Introduction

The depletion of fossil fuels and the resulting climate disruption from human-caused carbon dioxide emissions triggered the need to develop alternative fuel sources. Hydrogen (H_2), because to its high specific energy and capacity to be produced and utilized without CO_2 emissions, is a possible future car fuel. H_2 has the potential to replace fossil fuels in the energy industry and has numerous benefits over existing energy sources, such as its abundance, low environmental effect during combustion, and high specific energy, hence it is commonly regarded as the clean, sustainable fuel of the future (Satyapal et al., 2007; Greene and Duleep, 2013). Despite of its many potential uses, its low volumetric energy density limits its use in portable applications like fuel cell vehicles (Allendorf et al., 2018; Thornton et al., 2017). Researchers have been concentrating on developing inexpensive hydrogen storage technologies in order to overcome this restriction.

Storage via adsorption in porous hosts is a possible substitute for high-pressure compression. A new class of functional porous crystalline solids, metal–organic frameworks (MOFs) can be synthesized in a modular fashion from metal centers and organic ligands, resulting in a wide range of chemical and structural forms with properties that can be finely tuned (Öhrström, 2015; Batten et al., 2012). The presence of the metal centers and organic linkers provide access to a plethora of building blocks and can be altered by adding functional groups, switching out metals, or combining metals and linkers. The use of metal–organic frameworks (MOFs) as adsorbents has shown much promise (Farha et al., 2011; Sculley et al., 2011). Due to their high gravimetric densities, quick kinetics, reversibility, and the flexibility to adjust their chemistry, metal–organic frameworks (MOFs) have emerged as attractive hydrogen sorbents (Ahmed et al., 2019; Ahmed and Siegel, 2021).

Despite the diversity of possible MOFs, however, only a small percentage of them have actually been synthesized (Groom et al., 2016). Computational modeling can be difficult to apply to many of the synthesized MOFs due to their disorder, missing atoms, or low porosity. Computational design can be used to avoid such complexities. Nearly a million "hypothetical" MOFs have been reported, and it is reasonable to anticipate that many more materials may be proposed in the future (Wilmer et al., 2012; Bao et al., 2015; Gomez-Gualdron et al., 2014).

High-throughput screening with Grand Canonical Monte Carlo (GCMC) (Dubbeldam et al., 2013) has been successful in uncovering

* Corresponding author.
E-mail address: pc.chandra12@gmail.com (C. Chowdhury).

https://doi.org/10.1016/j.rsurfi.2023.100166
Received 18 November 2023; Accepted 21 November 2023
Available online 30 November 2023

interesting candidates with improved gas storage capabilities. There are approximately a million hypothetical MOFs have been reported to date. However, it might be difficult to undertake a systematic search of all of these resources, and discrepancies in execution can make it hard to compare screening studies. As a result, it would be helpful to have a more reliable method of screening MOFs for their gas storage properties prediction.

For this reason, machine learning (ML) is a fascinating discipline to investigate (Butler et al., 2018; Borboudakis et al., 2017). Faster than molecular simulations, ML systems can screen massive MOF datasets and anticipate their properties. Predicting MOF hydrogen storage with machine learning has multiple advantages as like accelerating discovery, identifying optimum MOFs, minimal input data for prediction, discovering high-capacity MOFs from huge database, etc.

ML could be useful, but only if sufficient training data is provided. Data on H_2 storage in MOFs obtained through experimentation is scant and very condition- and sample-specific. As a result, it could be preferable to use a dataset built on reliable computational forecasts. Using the pseudo-Feynman-Hibbs potential in conjunction with generic interatomic potentials, previous research has shown that H_2 isotherms in MOFs may be accurately predicted (Ahmed et al., 2019; Allendorf et al., 2018). Several MOFs, including IRMOF-20, SNU-70, UMCM-9, and PCN-610/NU-100, have been found as having a desirable H_2 density through these screening experiments. These prior investigations have established a database of MOF features that can be used to train machine learning algorithms to predict H_2 uptake across increasingly larger MOF datasets.

Predictions of H_2 storage capacities in MOFs have been made using a wide range of ML techniques, such as neural networks (NN), ridge linear regression (RR), support vector machines (SVM), and multilinear regression (MLR) (Lu et al., 2022; Wu et al., 2020; Purewal et al., 2019; Thornton et al., 2017; Anderson et al., 2018; Ma et al., 2020). Various features and training methods have been used in these studies to provide predictions about gravimetric and volumetric capacity.

Inspired by the above studies, in this paper, Deep Learning (DL) is used to sift through 918,734 items from various MOF repositories, a database of known and planned MOFs. Several NN models with various mathematical structures have been proposed thus far. The feed-forward NN (FNN) model has been shown to be especially robust and has seen extensive use in the fields of function fitting and data processing. The straightforward nature of FNN architectures makes them useful in a wide variety of chemistry and biology applications (Burns and Whitesides, 1993; Blount et al., 2017; Aziz et al., 2021; Cirovic, 1997). With such a large dataset, FNN method can be used to accurately evaluate H_2 uptake in MOFs. As an additional comparison, we used the ERT model and discovered that while the FNN model is superior for forecasting H_2 storage capacity gravimetrically, the ERT model is superior for predicting H_2 storage capacity in terms of volume. We have used GCMC simulations to determine the H_2 storage capabilities of MOFs, which allows us to make predictions about the accuracy of our ML model. We hope that our manuscript will have a great impact in current gas storage adsorption predictions where standard molecular simulations would take huge time whereas our model would take few minutes to compute the similar property with a reasonable accuracy compared to molecular simulation.

2. Computational methodology

2.1. Dataset

A total of 98 695 MOFs reported by Ahmed et al. (2019), are represented in the MOF database collected from the Hydrogen Materials Advanced Research Consortium (HyMARC) website. Included are 61 250 MOFs from the University of Ottawa database, 578 data from the University of Michigan database, 20 156 MOFs from the Northwestern University Library, and 5047 CoRE MOFs. Hydrogen storage capacity (both gravimetric and volumetric) was calculated using GCMC simulation as implemented in the RASPA code, and crystallographic characteristics such as gravimetric and volumetric surface areas (GSA and VSA), pore volume (PV), density (D), void fraction (VF), largest cavity diameter (LCD), and pore limiting diameter (PLD) were calculated using the Zeo++ code (Willems et al., 2012), as discussed in the paper. Measurements of H_2 storage capacity were made at temperatures and pressures ranging from 100 bar/77 K to 5 bar/160 K. Structure and chemical properties of MOFs were estimated by Ahmed et al. and used to characterize each structure.

2.2. Feedforward neural network (FNN)

A basic artificial neural network for pattern recognition is a feedforward neural network (Svozil et al., 1997). It has input, hidden, and output layers. See Fig. 1. The network learns the relationship between the variables it receives and the variables it returns by changing neuron link weights. The network is trained with known samples, spectra, and class labels. Adjusting the network's weights reduces the difference between predicted and true outputs. After training, the network can use spectra to classify unknown samples.

In a forward neural network, data flows from input to output. Loops or connections to the input layer are absent. All neurons in a layer are connected. These connections weight input signals. An activation function adds inputs in a non-linear way and sends the result to the next layer.

Common activation functions include sigmoid, tanh, and ReLU. These functions let the network learn complex data patterns and connections. The output layer estimates based on input. Backpropagation uses labeled training data and gradient descent to repeatedly change the network's weights.

Pattern recognition, classification, regression, and time series analysis use forward neural networks. They can be used with other neural network designs like convolution neural networks (CNNs) for image processing or recurrent neural networks (RNNs) for sequential data analysis. We used root mean squared error (RMSE) as our loss metric and which is determined by the equation given Eq. (1) where y_i, y_i^* and m denote the true values, predicted values and the number of samples in the dataset, respectively.

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - y_i^*)^2} \tag{1}$$

2.3. Extremely Randomised Tree (ERT)

Random Forests and Decision Trees are the building blocks of ERT, an ensemble learning method. It is a tweak on the Random Forest algorithm that promises better accuracy and less variety in the predictions. Because of its ability to deal with big datasets, noisy data, and high-dimensional feature spaces, the Extra Trees model is a popular choice. It is useful for solving classification and regression issues alike. The model can be trained quickly and accurately. There are several previous studies where ERT model is successfully utilized as a property predictor (Acosta et al., 2022; Ma and Liu, 2020; Chen et al., 2022).

2.4. Grand Canonical Monte Carlo simulation (GCMC)

We have also utilized GCMC simulation, which is commonly used to predict gas adsorption behavior in porous materials, to make predictions about the amount of H_2 that can be delivered and compared these to our ML findings to ensure the accuracy of our ML model. We carried out GCMC simulations using the RASPA software package (Dubbeldam et al., 2016). All of the porous materials studied had their framework atoms described using the universal force field (UFF) (Rappé et al., 1992). Lennard-Jones (LJ) potential was used to depict the interactions

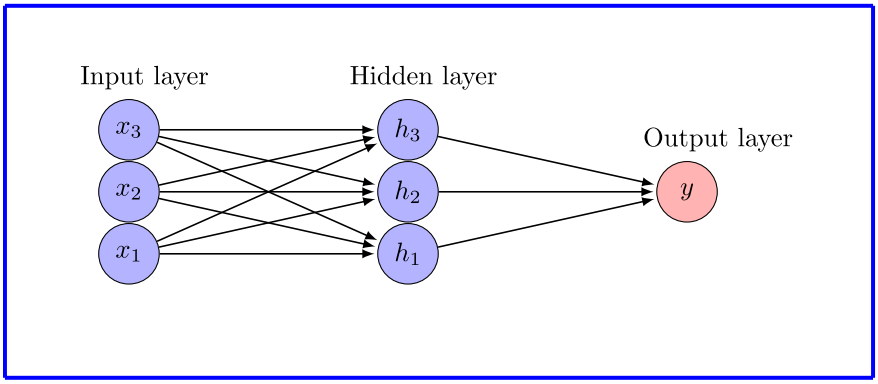

Fig. 1. Forward neural network architecture with 3 neurons in the input layer, 3 neurons in the hidden layer, and 1 neuron in the output layer.

between the framework atoms and H_2 while taking into account the Coulomb interaction. Cross LJ parameters were calculated using the Lorentz–Berthelot combining methods. Lennard-Jones (LJ) potential was adjusted to account for the quantum effects of hydrogen molecules at cryogenic temperatures using the Feynman-Hibbs adjustments (Fischer et al., 2009; Colón et al., 2014; Ahmed et al., 2019). Both H_2-H_2 and H_2-MOF interactions were calculated down to a cutoff radius of 12 Å. Following the procedure applied in the previous study (Ahmed et al., 2019), the MOF unit cell was replicated if the cell length was less than 24 Å in any direction. This means a replication of the unit cell twice in all the directions if the cell length was less than 24 Å. To account for the fact that interactions outside of the cutoff radius were ignored, long range corrections were applied. Table S1 and S2 in the Supporting Information file detailed all of the LJ parameters that were ultimately chosen for H_2 molecule as well as for MOFs. The GCMC simulations were conducted, employing a total of 20,000 cycles for the MOFs under consideration (Dubbeldam et al., 2016; Ahmed et al., 2019). In these particular cases, the final 10,000 cycles were dedicated to the computation of H2 adsorption. Every GCMC cycle consisted of a number of motions that was equivalent to the initial number of molecules in the system. The likelihood of performing translation, insertion, and deletion moves were equal.

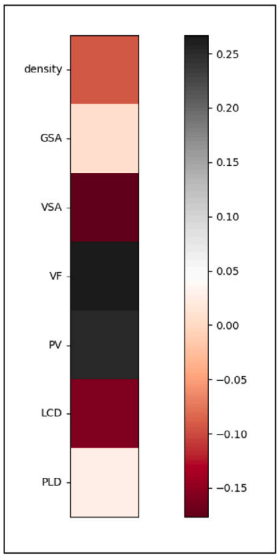

Fig. 2. Feature importance analysis from FNN model.

3. Results and discussion

First of all, to predict the H_2 deliverable capacities in both gravimetric and volumetric ways we have chosen seven feature vectors such as GSA, VSA, VF, D, PLD, LCD and PV as our input. In the following subsections we will describe how we predict using different ML methods.

3.1. Detailed descriptions of our FNN model

In our implementation, we utilize a specific type of feed-forward neural network known as an "input transformative" network. The primary purpose of this network is to take an input feature vector with a length of 7 (in our case as we have chosen 7 structural feature vectors as input) and apply a transformation that results in a vector of the same shape. This transformed vector is then passed through three subsequent hidden layers, ultimately generating the final output of the network.

To achieve the transformation step, we employ a fully connected layer, where each neuron in the layer is connected to every element in the input vector. The activation function used in this layer is the hyperbolic tangent (tanh) function. After conducting thorough experiments, we determined that tanh yielded the most favorable outcomes for the first layer of our network. Tanh operates by mapping the input values to a range between −1 and 1, providing non-linear transformations that facilitate better learning and representation capabilities.

It is crucial to note that our model does not work directly with the raw input data. Before feeding the inputs into the network, we preprocess them using a technique called zero-centering and unit variance scaling. This process involves adjusting the values of each feature so that they are centered around zero (subtracting the mean) and have a standard deviation of 1 (dividing by the standard deviation). This normalization procedure allows the model to handle inputs with varying scales and distributions more effectively.

Through extensive experimentation, we discovered that this preprocessing step significantly enhances the overall performance of our model. By preparing the inputs in this manner, we provide the network with standardized and consistent data, which aids in faster convergence, better gradient flow, and improved generalization.

In summary, our input transformative network utilizes a fully connected layer with a tanh activation function to transform the input feature vector. We preprocess the inputs by centering them around zero and scaling them to have unit variance. These design choices have been carefully selected and validated through experiments to optimize the network's performance.

In addition to the previous details, we incorporate Layer Normalization (LayerNorm) into each layer of our network. LayerNorm is a technique that normalizes the activations within each layer, ensuring that they follow a Gaussian distribution. This normalization aids in stabilizing the learning process and can improve the overall performance of the network.

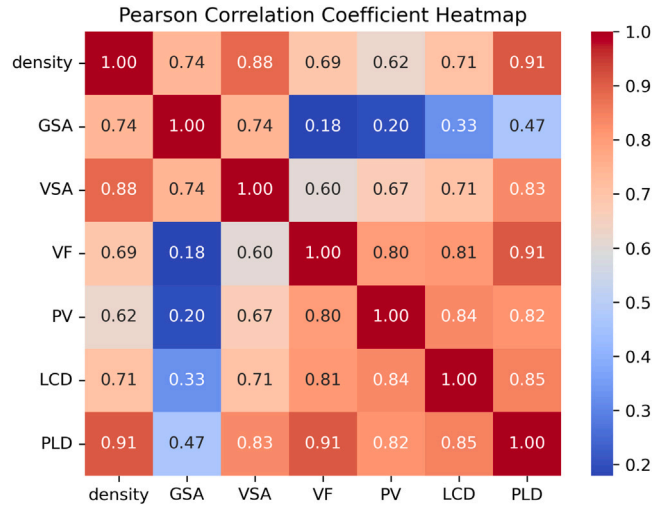

Fig. 3. Heatmap of the Pearson correlation coefficient matrix of 7 features.

Table 1
Feature importance analysis using polynomial method.

Polynomial degree	Number of feature vectors	Validation loss	
		Average	Minimum
1	8	1.95	1.90
2	36	1.88	1.83
3	120	1.87	1.83

During our extensive experimentation, we found that the Mish activation function yielded superior results for the hidden layers compared to other activation functions. Mish is a non-linear activation function that operates on a function of x rather than directly on x itself, similar to the tanh function. It provides smooth and continuous gradients, facilitating better optimization during the training process.

Furthermore, we leveraged our input data to obtain the outputs specifically from the first layer of the network. This allowed us to visualize the impact and importance of each feature in the model's decision-making process. By analyzing the outputs, we could identify if a particular feature was being scaled down to zero, indicating that it had little to no significance in the network's overall computations. This feature importance analysis aided us in understanding the relevance of each input feature and potentially making informed decisions regarding feature selection or engineering. From Fig. 2 shows the feature importance analysis we got from our FNN model for predicting gravimetric H_2 deliverable capacity.

Our FNN class constructor takes the following arguments:

- The size or number of neurons in the hidden layers of the model: 128
- The total number of hidden layers in the model: 3
- The organization of the hidden layers: Constant.
- The normalization used per-Layer: LayerNorm
- Learning rate: 0.001
- The Activation function used for the hidden layers: Mish

A hyperbolic tangent activation function and layer normalization are added after the input layer. Among 98 695 data we use 25% of the data 24 000 for validation, and the remaining 75% 74 000 are used for training. For the training, 300 epochs are used.

3.2. Detailed description of our ERT model

We have employed the polynomial strategy to perform feature engineering and applied it to our model (ERT) to ensure about the model's viability, even if the reduced-feature dataset yields machine learning models that are generally accurate. We can expand the feature space by generating polynomial features from the existing features in the dataset. This approach enabled us to capture intricate relationships and non-linear dependencies between the features, allowing our machine learning models to learn more complex patterns and improve their predictive performance. We perform 100 experiments. From Fig. 3, we can see that the Pearson correlation coefficient confirms our observations that several of the initial seven features are highly connected. Strong associations with outcomes are shown by a high correlation coefficient.

Using our polynomial feature importance analysis, a series of operators for expanding the feature space are implemented recursively and as a result of this total number of features increase from 7 to 120. The Pearson correlation coefficient between the target quantity and these meta-features (obtained through mathematical operations) is used to evaluate their usefulness. Table 1 shows the number of features taken for first, second and third degree polynomials and the corresponding validation loss values. Third-degree polynomial application yields the lowest validation loss compared to the other degrees tested.

3.3. Prediction of gravimetric H_2 deliverable capacity

To begin, we use the standard 7 feature vectors to train the FNN model to predict the hydrogen deliverable capacity of MOFs in the unit namely, UG at TPS (usable gravimetric hydrogen capacity for the temperature+pressure swing between 100 bar/77 K and 5 bar/160 K in units of weight percent). The neural network is trained using a cross-validation procedure with five iterations. Our model maintains a constant behavior across all five stages. From Fig. 4 it is seen that the 7 feature vectors used in the FNN model resulted in a validation loss of around 0.217 which is better than the previously reported ERT value (Ahmed and Siegel, 2021).

3.4. Prediction of volumetric H_2 deliverable capacity

Similarly for the prdiction of UV at TPS (usable volumetric hydrogen capacity for the temperature+pressure swing between 100 bar/77 K and 5 bar/160 K in units of g H_2 L^{-1} we have tried FNN model using the same 7 features). In this case we are getting validation loss of around 1.93 which is close to the ERT model using 7 features reported previously. Now, for the sake of tweaking the ERT model, we have used the features predicted by our polynomial method to train the model. In this case we observe the RMSE of around 1.83 which is better than the reported value. Also, one point to be noted, previously authors showed an increase in RMSE value to 2.10 for selecting univariate to multivariate features. But in our predicted model for ERT we show a better performance in terms of RMSE value while considering combination of features originated from polynomial feature importance analysis. In Fig. 5, through a parity plot, we demonstrate how the GCMC values and the ML projected values for volumetric H_2 deliverable capacity compare with one another.

3.5. Prediction of best performing MOFs through ML

After validating our ML models for predicting H_2 deliverable capacity (both UG and UV) in TPS condition, we estimate the H_2 storage capacity of a total of 820k MOFs for whom the H_2 deliverable capacity was previously unknown, therefore completing the validation process. A total of 820k MOFs were extracted from the previously disclosed database.

Considering MOF-5 as the benchmark for comparing the H_2 storage capacity in TPS which is 7.8 wt% and 51.9 g-H_2 L^{-1} (Ahmed et al., 2017), we found a total of 20 MOFs exceeding this condition. Based on the results of our ML model, the 20 MOFs in terms of their ability to predict H_2 storage capacity have been compiled and are presented in Table S3.

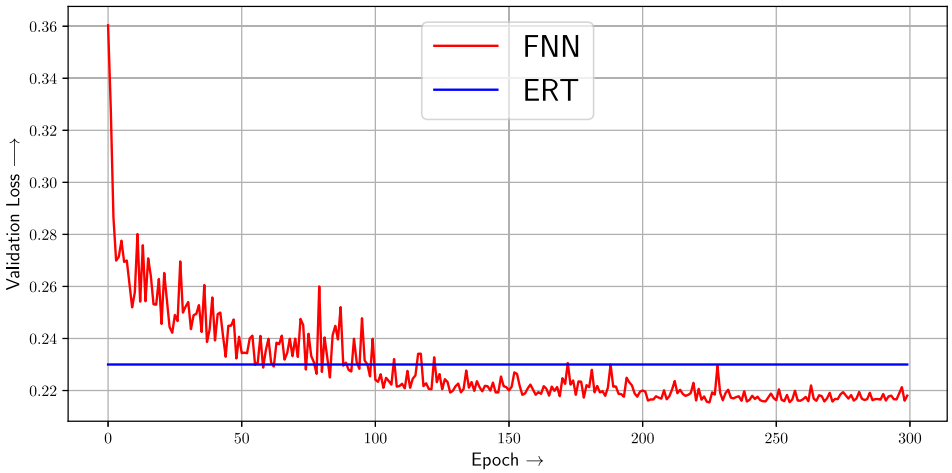

Fig. 4. Comparison of model performances of FNN with that of ERT in predicting H_2 deliverable capacity at UG at TPS condition in terms of validation loss.

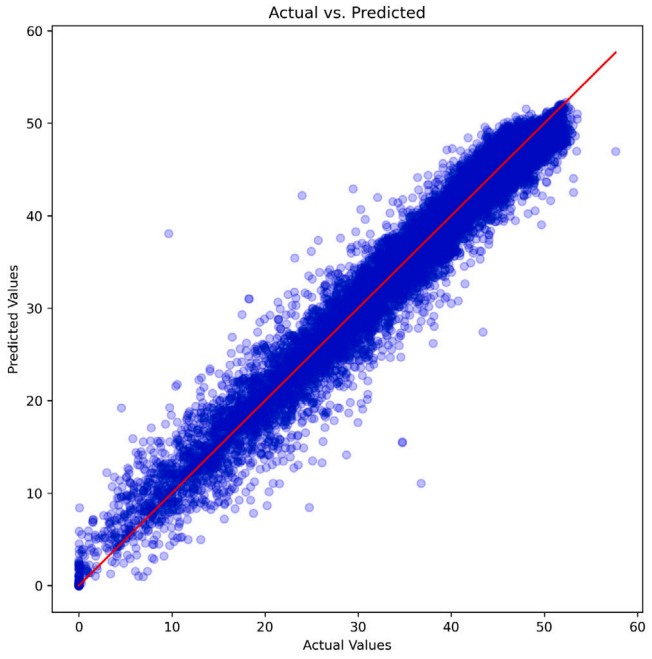

Fig. 5. Parity plot comparing the performance of ERT model with that of GCMC predicted values on predicting volumetric H_2 deliverable capacity.

Interestingly among these MOFs few are real MOFs which are already synthesized and two are hypothetical MOFs reported from theoretical calculations. Both of the hypothetical MOFs originated from R-WLLFHS database (Chung et al., 2016). Below Fig. 6 represents the best two hypothetical MOFs found from our ML study and it would be interesting for future research to study further for the purpose of H_2 storage capacity. A lower void proportion, as predicted by feature importance analysis, is observed in all MOFs that meet the requirements. Other than void fraction, it is seen that the highest capacity MOFs in TPS condition typically have the lower surface areas as well as higher densities. The hydrogen storage capacity of Metal-Organic Frameworks (MOFs) is greatly influenced by their density, as it has a direct effect on important factors such as surface area, pore volume, and molecular interaction. A positive correlation is frequently observed between higher density and an augmented internal surface area, hence facilitating the availability of additional sites for hydrogen molecules to adhere to and be stored within the structure of the MOFs. Furthermore, it has been observed that MOFs with higher density exhibit a higher pore volume, which allows for a greater amount of hydrogen to be stored, resulting in an increased storage capacity. The enhanced density facilitates the opportunity to optimize the dimensions, configuration, and intermolecular associations of pores, which are pivotal determinants influencing the efficacy of hydrogen adsorption and desorption processes occurring inside the framework of the MOFs. Our study shows a good agreement with the previous research (Ahmed et al., 2019; Ahmed and Siegel, 2021).

3.6. Comparison with GCMC results

We calculated the gravimetric and volumetric hydrogen delivery capacities of the two MOFs with the best ML predictions and ran GCMC simulations to compare our ML predicted results. Using GCMC simulations, we determined the hydrogen storage capacities in each MOF for the aforementioned temperature and pressure swing adsorption conditions (77 K/100 bar for charge, and 160 K/5 bar for discharge). The MOF's ability to store hydrogen was assessed using its deliverable capacity (N_{DC}), which is determined by the following equation:

$$N_{DC} = N_{77\ K,100\ bar} - N_{160\ K,5\ bar} \quad (2)$$

The outcomes of our GCMC simulations are in agreement with those of ML, within a tiny margin of error. The comparative results from GCMC and ML are shown in Table 2.

4. Conclusion

To reliably forecast the performance of MOFs in the context of hydrogen storage applications, we present a deep learning-based technique in this study. For predicting gravimetric hydrogen deliverable capacity in temperature and pressure swing situation with temperatures and pressures ranging from 100 bar/77 K to 5 bar/160 K, our FNN model outperforms the current result. Intriguingly, we built our FNN model to do feature importance analysis, which is critical for justifying the model's feature choices in light of the desired output. We also created an ML model to estimate the volumetric hydrogen storage capacity under the similar condition (i.e. temperature and pressure swing condition within a range of 100 bar/77 K to 5 bar/160 K), and the results showed that the ERT model was superior to the FNN model. We have included polynomial feature significance analysis, which involves generating polynomial features from the current features in the dataset in order to enlarge the feature space. Using this method, we were able to find a better outcome than what had previously been found.

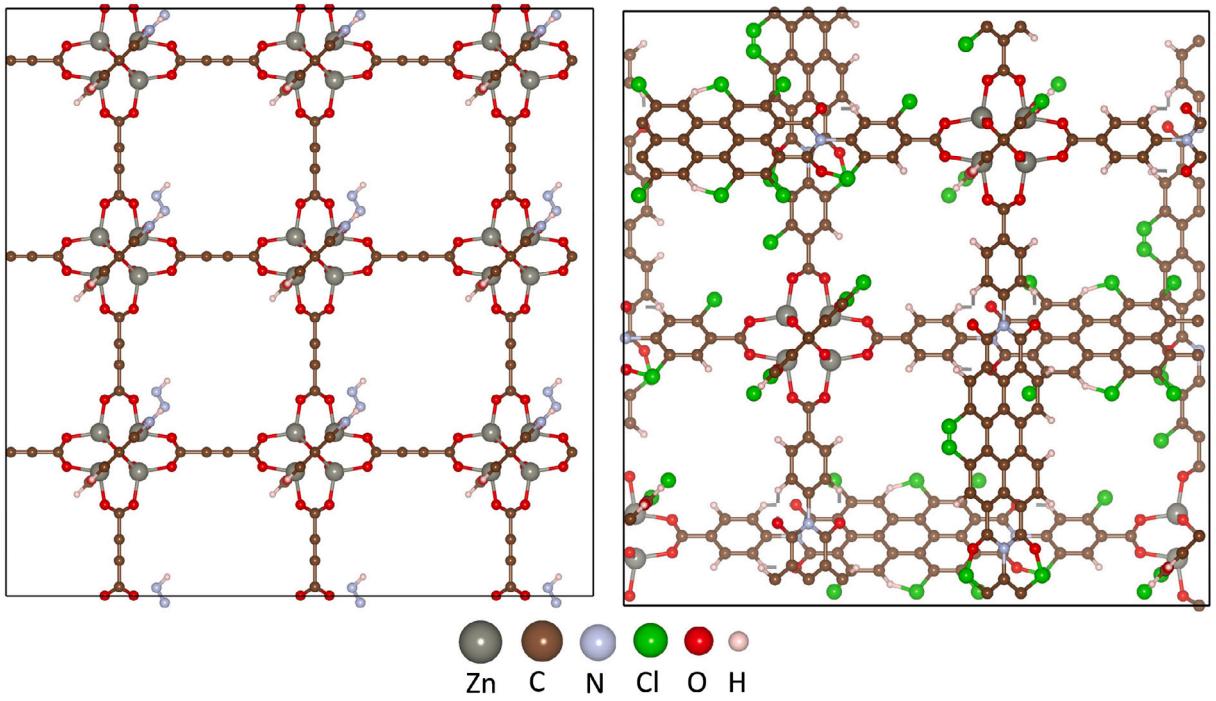

Fig. 6. Crystal structures of the hypothetical MOFs (hypotheticalMOF_5027108_1_0_0_28_2_8, hypotheticalMOF_50322703_2_0_29_29_2) showing both good gravimetric and volumetric H_2 storage capacities found from our ML study.

Table 2
Comparison of prediction performances of the highest capacity MOFs according to ML and GCMC under temperature + pressure swing conditions.

MOF	Density	GSA	VSA	VF	PV	LCD	PLD	Prediction (UG)		Prediction (UV)	
								ML	GCMC	ML	GCMC
hypotheticalMOF_5027108_1_0_0_28_2_8	0.55	4473.24	2466.39	0.81	1.47	7.45	9.46	8.23	7.58	52.25	50.04
hypotheticalMOF_5032270_3_2_0_29_29_2	0.52	4276.5	2231.77	0.81	1.55	10.84	12.76	8.36	7.39	52.30	50.21

CRediT authorship contribution statement

Shivanshu Shekhar: Data curation, Formal analysis, Investigation, Methodology. **Chandra Chowdhury:** Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing 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.

Data availability

Data used in this study can be taken from HyMARC database: https://datahub.hymarc.org/dataset.

Acknowledgments

C.C. gratefully acknowledges her sincere gratitude to Prof. Dr. Felix Studt of Karlsruhe Institute of Technology (KIT), Germany for the generous support. We would like to express our gratitude to Professor Randall Snurr of Northwestern University for his helpful discussions regarding the GCMC calculations using the RASPA code.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rsurfi.2023.100166. Force field parameters used for H_2 molecule, MOF atoms. List of MOFs predicted through ML.

References

Acosta, Carlos Mera, Ogoshi, Elton, Souza, Jose Antonio, Dalpian, Gustavo M, 2022. Machine learning study of the magnetic ordering in 2D materials. ACS Appl. Mater. Interfaces 14 (7), 9418–9432.

Ahmed, Alauddin, Liu, Yiyang, Purewal, Justin, Tran, Ly D, Wong-Foy, Antek G, Veenstra, Mike, Matzger, Adam J, Siegel, Donald J, 2017. Balancing gravimetric and volumetric hydrogen density in MOFs. Energy Environ. Sci. 10 (11), 2459–2471.

Ahmed, Alauddin, Seth, Saona, Purewal, Justin, Wong-Foy, Antek G, Veenstra, Mike, Matzger, Adam J, Siegel, Donald J, 2019. Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. Nat. Commun. 10 (1), 1568.

Ahmed, Alauddin, Siegel, Donald J., 2021. Predicting hydrogen storage in MOFs via machine learning. Patterns 2 (7), 100291.

Allendorf, Mark D, Hulvey, Zeric, Gennett, Thomas, Ahmed, Alauddin, Autrey, Tom, Camp, Jeffrey, Cho, Eun Seon, Furukawa, Hiroyasu, Haranczyk, Maciej, Head-Gordon, Martin, et al., 2018. An assessment of strategies for the development of solid-state adsorbents for vehicular hydrogen storage. Energy Environ. Sci. 11 (10), 2784–2812.

Anderson, Grace, Schweitzer, Benjamin, Anderson, Ryther, Gómez-Gualdrón, Diego A, 2018. Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: Molecular simulation and machine learning. J. Phys. Chem. C 123 (1), 120–130.

Aziz, N., Wijaya, A., Fathona, R.S., et al., 2021. Application of feed forward neural network model to predict the limiting current of tin magneto electrodeposition. In: IOP Conference Series: Materials Science and Engineering, Vol. 1011. IOP Publishing, 012005.

Bao, Yi, Martin, Richard L, Haranczyk, Maciej, Deem, Michael W, 2015. In silico prediction of MOFs with high deliverable capacity or Internal Surface Area. Phys. Chem. Chem. Phys. 17 (18), 11962–11973.

Batten, Stuart R, Champness, Neil R, Chen, Xiao-Ming, Garcia-Martinez, Javier, Kitagawa, Susumu, Öhrström, Lars, O'Keeffe, Michael, Suh, Myunghyun Paik, Reedijk, Jan, 2012. Coordination polymers, metal–organic frameworks and the need for terminology guidelines. CrystEngComm 14 (9), 3001–3004.

Blount, Drew, Banda, Peter, Teuscher, Christof, Stefanovic, Darko, 2017. Feedforward chemical neural network: An in silico chemical system that learns XOR. Artif. Life 23 (3), 295–317.

Borboudakis, Giorgos, Stergiannakos, Taxiarchis, Frysali, Maria, Klontzas, Emmanuel, Tsamardinos, Ioannis, Froudakis, George E, 2017. Chemically intuited, large-scale screening of MOFs by machine learning techniques. Npj Comput. Mater. 3 (1), 40.

Burns, John A., Whitesides, George M., 1993. Feed-forward neural networks in chemistry: Mathematical systems for classification and pattern recognition. Chem. Rev. 93 (8), 2583–2601.

Butler, Keith T, Davies, Daniel W, Cartwright, Hugh, Isayev, Olexandr, Walsh, Aron, 2018. Machine learning for molecular and materials science. Nature 559 (7715), 547–555.

Chen, Liang, Huang, Huan, Thangavelu, Lakshmi, Abdelbasset, Walid Kamal, Bokov, Dmitry Olegovich, Algarni, Mohammed, Ghazali, Sami, Alashwal, May, 2022. Optimization and comparison of machine learning methods in estimation of carbon dioxide loading in chemical solvents for environmental applications. J. Mol. Liq. 349, 118513.

Chung, Yongchul G, Gómez-Gualdrón, Diego A, Li, Peng, Leperi, Karson T, Deria, Pravas, Zhang, Hongda, Vermeulen, Nicolaas A, Stoddart, J Fraser, You, Fengqi, Hupp, Joseph T, et al., 2016. In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm. Sci. Adv. 2 (10), e1600909.

Cirovic, Dragan A., 1997. Feed-forward artificial neural networks: Applications to spectroscopy. TRAC Trends Anal. Chem. 16 (3), 148–155.

Colón, Yamil J, Fairen-Jimenez, David, Wilmer, Christopher E, Snurr, Randall Q, 2014. High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature. J. Phys. Chem. C 118 (10), 5383–5389.

Dubbeldam, David, Calero, Sofía, Ellis, Donald E, Snurr, Randall Q, 2016. RASPA: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. Mol. Simul. 42 (2), 81–101.

Dubbeldam, David, Torres-Knoop, Ariana, Walton, Krista S., 2013. On the inner workings of Monte Carlo codes. Mol. Simul. 39 (14–15), 1253–1292.

Farha, Omar K, Shultz, Abraham M, Sarjeant, Amy A, Nguyen, SonBinh T, Hupp, Joseph T, 2011. Active-site-accessible, porphyrinic metal- organic framework materials. J. Am. Chem. Soc. 133 (15), 5652–5655.

Fischer, Michael, Hoffmann, Frank, Fröba, Michael, 2009. Preferred hydrogen adsorption sites in various MOFs—a comparative computational study. ChemPhysChem 10 (15), 2647–2657.

Gomez-Gualdron, Diego A, Gutov, Oleksii V, Krungleviciute, Vaiva, Borah, Bhaskarjyoti, Mondloch, Joseph E, Hupp, Joseph T, Yildirim, Taner, Farha, Omar K, Snurr, Randall Q, 2014. Computational design of metal–organic frameworks based on stable zirconium building units for storage and delivery of methane. Chem. Mater. 26 (19), 5632–5639.

Greene, David L., Duleep, G., 2013. Worldwide Status of Hydrogen Fuel Cell Vehicle Technology and Prospects for Commercialization. US Department of Energy.

Groom, Colin R, Bruno, Ian J, Lightfoot, Matthew P, Ward, Suzanna C, 2016. The cambridge structural database. Acta Crystallogr. B 72 (2), 171–179.

Lu, Xiuyang, Xie, Zhizhong, Wu, Xuanjun, Li, Mengmeng, Cai, Weiquan, 2022. Hydrogen storage metal-organic framework classification models based on crystal graph convolutional neural networks. Chem. Eng. Sci. 259, 117813.

Ma, Ruimin, Colon, Yamil J., Luo, Tengfei, 2020. Transfer learning study of gas adsorption in metal–organic frameworks. ACS Appl. Mater. Interfaces 12 (30), 34041–34048.

Ma, Sicong, Liu, Zhi-Pan, 2020. Machine learning for atomic simulation and activity prediction in heterogeneous catalysis: Current status and future. ACS Catal. 10 (22), 13213–13226.

Öhrström, Lars, 2015. Let's talk about MOFs—Topology and terminology of metal-organic frameworks and why we need them. Crystals 5 (1), 154–162.

Purewal, Justin, Veenstra, Mike, Tamburello, David, Ahmed, Alauddin, Matzger, Adam J, Wong-Foy, Antek G, Seth, Saona, Liu, Yiyang, Siegel, Donald J, 2019. Estimation of system-level hydrogen storage for metal-organic frameworks with high volumetric storage density. Int. J. Hydrogen Energy 44 (29), 15135–15145.

Rappé, Anthony K., Casewit, Carla J., Colwell, K.S., Goddard III, William A, Skiff, W Mason, 1992. UFF, A full periodic table force field for molecular mechanics and molecular dynamics simulations. J. Am. Chem. Soc. 114 (25), 10024–10035.

Satyapal, Sunita, Petrovic, John, Read, Carole, Thomas, George, Ordaz, Grace, 2007. The US department of energy's National hydrogen storage project: Progress towards meeting hydrogen-powered vehicle requirements. Catal. Today 120 (3–4), 246–256.

Sculley, Julian, Yuan, Daqiang, Zhou, Hong-Cai, 2011. The current status of hydrogen storage in metal–organic frameworks—Updated. Energy Environ. Sci. 4 (8), 2721–2735.

Svozil, Daniel, Kvasnicka, Vladimir, Pospichal, Jiri, 1997. Introduction to multi-layer feed-forward neural networks. Chemometr. Intell. Lab. Syst. 39 (1), 43–62.

Thornton, Aaron W, Simon, Cory M, Kim, Jihan, Kwon, Ohmin, Deeg, Kathryn S, Konstas, Kristina, Pas, Steven J, Hill, Matthew R, Winkler, David A, Haranczyk, Maciej, et al., 2017. Materials genome in action: Identifying the performance limits of physical hydrogen storage. Chem. Mater. 29 (7), 2844–2854.

Willems, Thomas F, Rycroft, Chris H, Kazi, Michaeel, Meza, Juan C, Haranczyk, Maciej, 2012. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. Microporous Mesop. Mater. 149 (1), 134–141.

Wilmer, Christopher E, Leaf, Michael, Lee, Chang Yeon, Farha, Omar K, Hauser, Brad G, Hupp, Joseph T, Snurr, Randall Q, 2012. Large-scale screening of hypothetical metal–organic frameworks. Nat. Chem. 4 (2), 83–89.

Wu, Xuanjun, Wang, Yuguo, Cai, Zhijun, Zhao, Dongmei, Cai, Weiquan, 2020. Revealing enhancement mechanism of volumetric hydrogen storage capacity of nano-porous frameworks by molecular simulation. Chem. Eng. Sci. 226, 115837.