Scientific Reports

https://doi.org/10.1038/s41598-026-45915-1

Article in Press

AI-driven optimization of hydrogen storage in porous carbon adsorbents

Received: 20 January 2026

Accepted: 23 March 2026

Published online: 26 March 2026

Cite this article as: Rocha H.R.O., Romanos J., Abou Dargham S. *et al.* AI-driven optimization of hydrogen storage in porous carbon adsorbents. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-45915-1

Helder R. O. Rocha, Jimmy Romanos, Sara Abou Dargham, Roy Roukos, Jair A. L. Silva & Heinrich Joh. Wörtche

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

AI-Driven Optimization of Hydrogen Storage in Porous Carbon Adsorbents

Helder R. O. Rocha[1], Jimmy Romanos[2], Sara Abou Dargham[3,*], Roy Roukos[2], Jair A. L. Silva[1,3], and Heinrich Wörtche[3,4]

[1]Department of Electrical Engineering, Federal University of Espírito Santo, Vitória, Espírito Santo, Brazil
[2]Department of Natural Science, Lebanese American University (LAU), Byblos, P.O.Box. 36, Lebanon
[3]Sensors and Smart Systems, Institute of Engineering, Hanze University of Applied Sciences, Groningen, 9747 AS, The Netherlands
[4]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, 5612 AZ, The Netherlands
[*]s.abou.dargham@pl.hanze.nl

ABSTRACT

Efficient hydrogen (H_2) storage remains a major challenge for clean energy applications. This study presents an AI-driven methodology to optimize H_2 storage in porous carbon adsorbents. A comprehensive dataset of 917 literature-derived entries was used to develop two machine learning models: Random Forest (RF) and Convolutional Neural Network (CNN). Both models accurately predicted hydrogen uptake based on material properties and experimental conditions. ~~, with the CNN achieving superior performance (R^2 = 0.9353; RMSE = 0.0406).~~ Within the range of the experimental dataset, the CNN demonstrated strong interpolation performance, accurately predicting hydrogen uptake with a high coefficient of determination (R^2 = 0.9353) and a Root Mean Squared Error (RMSE) of 0.0406. The CNN was integrated into a multi-objective optimization framework to maximize hydrogen uptake while minimizing average pore diameter (AVD). Through extrapolative optimization beyond the training data range, the AI-driven technique and optimization method (AiDO) identified theoretical Pareto-optimal solutions extending beyond the experimental dataset, predicting H_2 uptake of up to 16.66 wt% at an AVD of 0.08 nm. While these extrapolated solutions are not directly validated by experiments, constrained optimization scenarios (e.g., realistic pore-size limits) provide physically meaningful design targets. ~~The AI-driven technique and optimization method (AiDO) identified material parameters beyond the experimental dataset, predicting H_2 uptake of 16.66 wt% at an AVD of 0.08 nm.~~ Sensitivity analysis confirmed the robustness of the methodology to different normalization techniques. This approach demonstrates the potential of combining predictive ML with optimization to accelerate the design of high-performance hydrogen adsorbents, reducing experimental costs and supporting sustainable energy systems.

ARTICLE IN PRESS

Introduction

The urgent need to decarbonize the transportation sector has intensified the search for viable clean energy storage solutions. Hydrogen (H_2), with its clean combustion properties, emerges as a promising alternative to fossil fuels. However, effective hydrogen storage remains a critical challenge. Conventional methods, such as compression and liquefaction, are limited by safety, efficiency, and cost. Material-based storage, based on chemisorption or physisorption, offers a safer and potentially more efficient approach, particularly for vehicular applications.

Chemisorption involves the absorption of hydrogen in metal hydrides through ionic, covalent, or metallic-type bonds, resulting in liquids or solids that release hydrogen upon heating or catalytic exposure[1,2]. However, this method is typically irreversible, with metal hydrides requiring high temperatures and chemical hydrides necessitating recycling for each charging cycle. Physisorption, in contrast, attracts hydrogen molecules to sorbents like activated carbon (AC), zeolites, and metal-organic frameworks (MOF) through weak van der Waals forces, allowing for reversible storage[3–8].

Despite advances, near-ambient hydrogen uptake remains low due to weak interactions, making optimization of adsorbent properties, such as specific surface area (SSA), pore volume, and average pore diameter (AVD), essential for achieving high gravimetric and volumetric capacities[9]. To advance hydrogen-powered systems, the United States Department of Energy (DOE) has established stringent targets for hydrogen storage systems[10]. Among the leading materials for reversible hydrogen storage, ACs have been particularly effective due to their high porosity and surface area. Computational modeling has played a crucial role in optimizing the pore structure and surface chemistry of ACs for enhanced hydrogen physisorption[11–13]. Improvements focus on tailoring pore structures through controlled activation and modifying surface chemistry via elemental doping—such as boron, aluminum, lithium, and calcium[14–17]. These efforts aim to meet the U.S. DOE's targets for hydrogen storage by reducing pressure requirements, improving both gravimetric and volumetric capacity, and lowering costs. However, low uptake

ARTICLE IN PRESS

at ambient conditions remains a key challenge due to weak hydrogen-adsorbent interactions. Achieving efficient storage at near-ambient temperatures and safe pressures depends on optimizing parameters like specific surface area (SSA), pore volume, and pore size distribution (PSD)[15]. Common strategies include precursor selection, carbonization, and activation (physical or chemical), which enhance porosity and surface area[18–20]. Additionally, incorporating metal nanoparticles, such as nickel, has been shown to improve hydrogen uptake from 0.15 wt% to 0.53 wt%[20]. Storage performance is typically evaluated at 77 K or 298 K under fixed pressures[21,22]. Therefore a multidisciplinary approach, combining experimental and computational, is essential for accelerating progress. While experimental data provides the foundation for understanding hydrogen-material interactions, computational models offer insights into the underlying mechanisms at the atomic level. Furthermore, leveraging the predictive power of artificial intelligence (AI), specifically machine learning (ML), can enable rapid prediction of hydrogen uptake, accelerating the discovery of promising materials and optimization of storage conditions.

ML has emerged as a powerful tool for predicting gas adsorption in porous materials. Several studies have demonstrated its application to methane (CH_4)[23–25] carbon dioxide (CO_2)[26–28], and hydrogen (H_2) systems. Applications span a range of materials, MOFs, COFs, zeolites, hydrides, and ACs. For instance, explainable ML models have been successfully used to predict hydrogen uptake in MOFs, providing insights into structure–performance relationships and guiding material design[29]. More broadly, representative ML studies in gas storage are summarized in Table 1, illustrating the range of models and target gases. This comparison underscores the growing role of machine learning in materials design for energy storage and illustrates the research gap that motivates the present work, namely the application of advanced models to hydrogen storage in activated carbons.

For hydrogen storage, Rahnama et al. predicted material class and hydrogen content in metal hydrides using DOE data[30,31], while Wang et al. modeled storage efficiency in metal hydride chambers[37]. Artificial neural networks (ANN) was used to predict uptake in 28 zeolites from 349 data points at 77 K[32]. Anderson et al. combined molecular simulation and ML to explore hydrogen storage in porous crystals[38]. ML models also assessed H_2 uptake in MOFs, identifying key features like surface area, pore volume, and operating conditions[29,33,34]. Additionally, in materials-oriented research, ML was applied to evaluate magnesium-containing (Mg) intermetallic compounds for hydrogen storage and identified a promising candidate containing beryllium (Be) for achieving high gravimetric hydrogen density[39]. For porous carbons, Kusdhany et al. predicted uptake at 77 K using textural and chemical data[35], and Davoodi et al. compared four ML models for prediction performance[36].

Although substantial experimental and computational efforts have improved hydrogen physisorption in activated carbons by tailoring surface area, pore volume, pore size distribution, and surface chemistry, progress is still constrained by the large combinatorial design space of precursors/activation strategies and by the cost and time associated with trial-and-error experimentation[13,36,40]. In addition, reported adsorption data are often fragmented across different temperatures and pressures, which complicates cross-study comparisons and limits direct transferability of empirical design rules. More recently, machine learning has been introduced as a surrogate to predict hydrogen uptake from material descriptors and operating conditions; however, most ML studies remain primarily predictive and are not systematically coupled to an optimization step that can propose new combinations of material parameters and conditions and quantify trade-offs among competing design objectives relevant to adsorbent design[36,41]. ~~While prior studies have used ML primarily for predictive modeling, this work introduces a novel approach that combines ML with multi-objective optimization. Our method not only predicts hydrogen uptake but also identifies optimal material parameters without relying on pre-existing experimental datasets.~~ While prior work has largely focused on prediction, some studies in related fields demonstrate the potential of combining ML with optimization. For example, in chemical engineering, Alatefi et al. integrated ML with evolutionary and metaheuristic optimization to identify optimal material and process parameters[42].

Similarly, to address these limitations in hydrogen storage, this study proposes an AI-driven technique and optimization framework (AiDO) for hydrogen storage in porous carbon adsorbents. The objectives are to: compile and curate a comprehensive literature-derived dataset linking key textural/material descriptors and experimental conditions to hydrogen uptake; develop and validate two predictive models (Random Forest and Convolutional Neural Network) to accurately estimate hydrogen uptake; and integrate the best-performing model into a bi-objective optimization strategy to maximize hydrogen uptake while minimizing average pore diameter, thereby generating Pareto-optimal design targets that can guide experimental synthesis. In summary, this paper introduces a novel methodology to enhance hydrogen storage in porous carbon adsorbents by combining AI-driven techniques and optimization methods (AiDO). The main novelty of AiDO is the tight coupling of a data-driven predictive model with multi-objective optimization to move beyond prediction-only workflows and provide actionable, quantitatively optimized parameter sets, supported by sensitivity analysis to ensure robustness against data preprocessing choices. While prior studies have primarily focused on predictive modeling, our approach advances the field by integrating multi-objective optimization to identify optimal parameters for maximizing hydrogen uptake. This dual strategy enables both accurate prediction and data-driven material design, independent of existing experimental datasets. Key contributions include the creation of a comprehensive dataset of 917 literature-derived data points, the development and validation of two ML models on this dataset, the integration of these models into a multi-objective optimization framework, and the identification of optimal adsorbent

ARTICLE IN PRESS

Table 1. Overview of machine learning models applied for gas storage materials.Abbreviations are defined in the Supplementary Material.

Study	Application	Dataset	ML Model(s)	Best Performance
Pardakhti et al.[23]	Predict methane uptake in hypothetical MOFs	130,398 MOFs	DT, PR, SVM and RF (8% of data trained)	RF ($R^2 = 0.98$; MAPE = 7%)
Kim et al.[24]	Predict methane isotherms at various temperatures	4,951 CoRE-MOFs with 8 features	Linear and non-linear SVM, DT and RF	RF ($R^2 = 0.947$; MAPE = 3.903%)
Meng et al.[25]	Methane adsorption on shale	Data from 20 cases	ANN, RF, SVM and XGBoost	XGBoost ($R^2 = 0.9781$; RMSE = 0.0053)
Abdi et al.[26]	Predict CO_2 uptake in MOFs	20 MOFs-based with 1,191 data points	CatBoost, LightGBM, XGBoost, and RF	XGBoost ($R^2 = 0.9955$; RMSE = 0.5682)
Yuan et al.[27]	CO_2 adsorption in biomass carbons	527 data points; 11 features	GBDT, XGB, and LGB	GBDT ($R^2 = 0.84$; RMSE = 0.66)
Maheri et al.[28]	Improve ML for CO_2 adsorption	Yuan et al.[27] dataset with 13 features	MGBR, MLP, LSTM, and CNN	CNN ($R^2 = 0.8573$; MSE = 1.36)
Alatefi et al.[29]	Predict H_2 storage capacity in MOFs	1729 dataset	BRANN, LSSVM, and ET	ET ($R^2 = 0.995$; RMSE = 0.1445).
Rahnama et al.[30,31]	Predict metal hydride class & hydrogen weight percent	DOE metal hydrides database	BDTR, BLR, NNR, and LR	BDTR ($R^2 = 0.83$; RMSE = 0.0117862)
Alizadeh et al.[32]	Simulate and predict hydrogen uptake ability of zeolite	28 zeolites; 349 experimental data points	MLPNN, CFFNN, GRNN, and RNN	CFFNN ($R^2 = 0.99429$; RMSE = – 0.058)
Salehi et al.[33]	Predict H_2 storage capacity in MOFs	293 dataset	MLP, SVM, RF, CatBoost, LightGBM, XGBoost, and CMIS	CMIS ($R^2 = 0.982$; RMSE = 0.088).
Meduri et al.[34]	Enhance H_2 uptake prediction in MOFs	141 MOFs	LR, RR, DT, SVR, KNN, ET, RF, AB, GB, and GRNN	ET ($R^2 = 0.9917$; RMSE = 0.1827).
Meduri et al.[34]	Enhance H_2 uptake prediction in MOFs	141 MOFs	LR, RR, DT, SVR, KNN, ET, RF, AB, GB, and GRNN	ET ($R^2 = 0.9917$; RMSE = 0.1827).
Kusdhany et al.[35]	Hydrogen uptake in porous carbon materials	68 samples with 1745 data points	LR, SVR(L), SVR (RBF), XGBT, RF	RF ($R^2 = 0.910$; RMSE = 0.542).
Davoodi et al.[36]	Hydrogen uptake in porous carbon materials	2072 data records	GRNN, LSSVM, ANFIS, ELM	LSSVM ($R^2 = 0.9910$; RMSE = 0.1861).
This study	**Predict and optimize H_2 storage in porous carbon adsorbents**	**917 literature-derived data points**	**RF and CNN**	**CNN ($R^2 = 0.9353$; RMSE = 0.0406)**

ARTICLE IN PRESS

ARTICLE IN PRESS

parameters, potentially reducing both experimental costs and development time.

Methods

Datasets description and processing

The data collection process utilized activated carbons (AC) for hydrogen storage in porous materials, incorporating 566 patterns derived from the work of Rahimi et al.[41], as well as an additional 351 patterns extracted from previously published research[19,21,40,43–72]. All collected datasets are provided in the Supplementary Material.

Table 2 presents a statistical analysis of various variables related to adsorption and the characteristics of the adsorbent materials. A numerical value 0, 1 and 2 was assigned to each type of adsorbent of activated carbon, biochar, and metal-doped carbon, respectively. Same was done for the chemical activation agent that can be alkali, acidic, and metal chloride types. The value 0 was assigned when no additional chemical agent was used. the numbers 1, 2, 3, 4, 5 and 6 corresponded to the use of phosphoric acid (H_3PO_4), calcium oxide (CaO), zinc chloride ($ZnCl_2$), carbon dioxide (CO_2), sodium hydroxide (NaOH) and potassium hydroxide (KOH), respectively. These numerical labels are purely categorical and serve only as identifiers without any inherent mathematical relationship or physical significance. During data preprocessing, categorical variables were encoded using these numerical labels. For the Random Forest model, this encoding is appropriate because tree-based algorithms partition features through hierarchical splits rather than relying on numerical ordering. For the CNN model, the encoded categorical variables are processed together with other inputs through nonlinear transformations within the network layers, allowing the model to learn interactions between categorical and continuous features. Consequently, statistical operations such as averaging or standard deviation calculations are not meaningful in this context. The activating agent ratio (Ratio) defined as the ratio of the mass of the activation agent to the mass of the carbon precursor, has a mean of 1.42 with a standard deviation (STD) of 1.63. The specific surface area (SSA) shows a mean of 2218.11 m^2/g and an STD of 959.63. Micropore volume (Vmic) ranges from 0 to 2.01 cm^3/g, with a mean of 0.87 and an STD of 0.51, while mesopore volume (Vmes) has a mean of 0.29 and an STD of 0.33. Pressure ranges from 0.70 to 298 bar, with a mean of 37.81 bar and an STD of 49.85. Temperature (Temp.) ranges from 1 to 303 K, with a mean value of 141.27 K and a standard deviation of 98.51. The amount of adsorbed hydrogen (H_2), reported as gravimetric Gibbs excess adsorption, ranges from 0 to 12.60 wt%, with a mean of 2.68 wt% and a standard deviation of 2.17. The average pore diameter (AVD) has a mean value of 2.05 nm and a standard deviation of 0.50.

Table 2. Statistical analysis of dataset variables.

	Mean	STD	Min	Max
Adsorbent			0.00	2.00
Agent			0.00	6.00
Ratio	1.42	1.63	0.00	5.00
SSA (m^2g^{-1})	2218.11	959.63	0.42	4310.00
Vmic (cm^3g^{-1})	0.87	0.51	0.00	2.01
Vmes (cm^3g^{-1})	0.29	0.33	0.00	1.80
Pressure (bar)	37.81	49.85	0.70	298.00
Temp. (K)	141.27	98.51	1.00	303.00
H_2 (%wt)	2.68	2.17	0.00	12.60
AVD (nm)	2.05	0.50	0.00	4.43

All variables were normalized prior to model training to ensure that features with different numerical ranges contribute comparably to the learning process and to improve numerical stability of the machine-learning models. The normalized distributions of the variables are illustrated in Figure 1, highlighting the multivariate relationships between hydrogen storage and the material descriptors and operating conditions included in the dataset. ~~Figure 1 shows the normalized distribution of the variables in the dataset revealing the complex and multivariate relationship between hydrogen storage and these variables.~~ Notably, the measurement conditions pressure and temperature, which affect the interaction between hydrogen molecules and the adsorbent, are not uniformly distributed between the maximum and minimum limits of these variables. Because hydrogen uptake is strongly dependent on operating conditions, a literature-curated dataset can exhibit uneven sampling of pressure and temperature. As shown in Figure 1, pressure and temperature are not uniformly distributed across their ranges, which implies that predictive performance will be most reliable near the densely sampled regions of (P, T), and material ranking can be biased if comparisons are made across varying measurement conditions. In this work, pressure and temperature are therefore treated as explicit conditional inputs to the model alongside material descriptors (Agent, Ratio, SSA, Vmic, Vmes). Accordingly,

ARTICLE IN PRESS

model outputs should be interpreted as the predicted uptake at the specified (P,T), and fair comparison of materials should be performed by fixing (P,T) to standardized conditions of interest. In contrast, the other variables show a uniform distribution of the data.

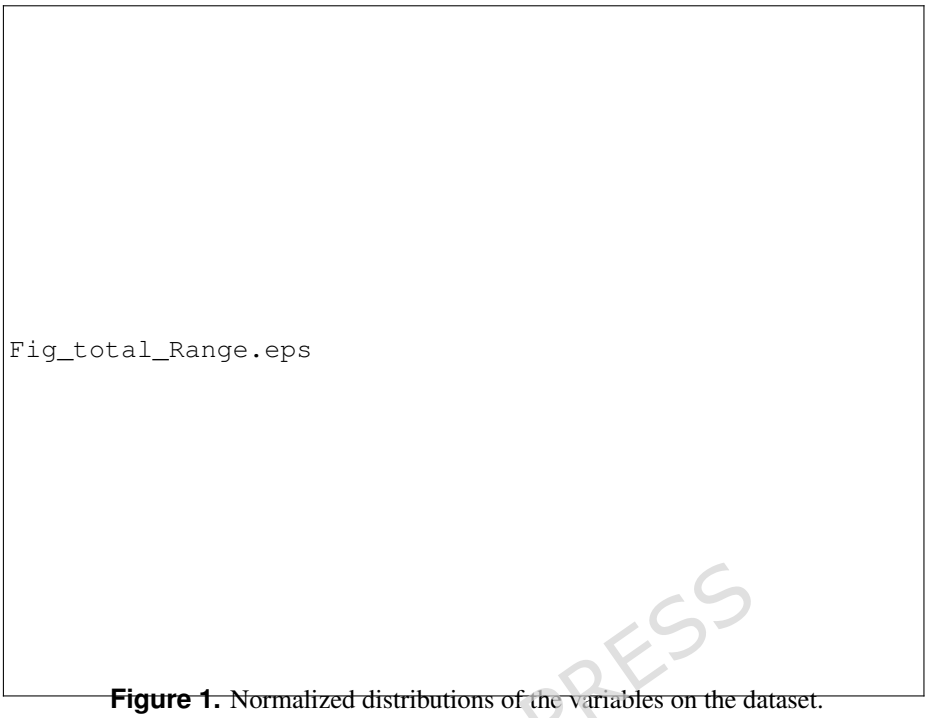

Figure 1. Normalized distributions of the variables on the dataset.

~~Figure 2 shows that high values of Vmic, along with SSA, result in significant hydrogen storage for the ACs adsorbent. This indicates that a higher SSA, combined with a lower proportion of micropore volume to total volume, favors hydrogen adsorption.~~ The relationship between micropore volume (Vmic), specific surface area (SSA), and hydrogen storage performance is illustrated in Figure 2. The trend suggests that adsorbents with larger SSA and a lower proportion of micropore volume to total volume, provide more adsorption sites, thereby enhancing hydrogen uptake. These results suggest that optimizing these parameters could be crucial for developing more efficient adsorbents for hydrogen storage.

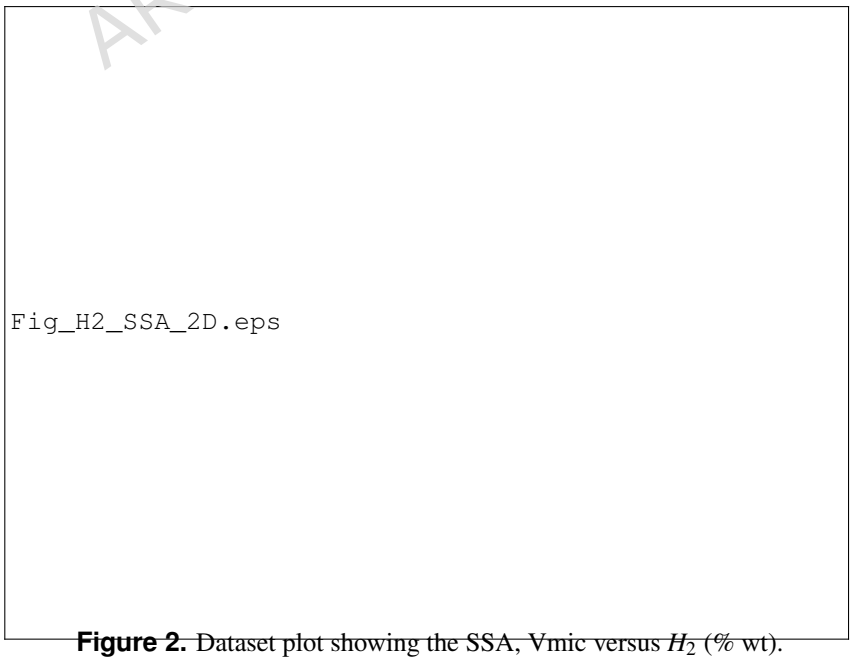

Figure 2. Dataset plot showing the SSA, Vmic versus H_2 (% wt).

ARTICLE IN PRESS

The AI-Driven methodology

Once the main dataset is assembled, the neural network training process begins with Max-Min Scaler normalization. Table 3 presents the maximum and minimum thresholds used for this normalization, which are set to aid in the extrapolation of results during the optimization process.

Table 3. Maximum and minimum limits for normalization.

Variable	Min	Max
Adsorbent	0.00	2.00
Agent	0.00	6.00
Ratio	0.00	10.00
SSA (m^2g^{-1})	0.00	5310.00
Vmic (cm^3g^{-1})	0.00	4.00
Vmes (cm^3g^{-1})	0.00	4.00
Pressure (bar)	0.00	298.00
Temp. (K)	0.00	303.00
H2 (%wt)	0.00	30.00
AVD (nm)	0.00	8.00

We developed a CNN architecture by tuning various hyperparameters with the Optuna framework. These hyperparameters include the number of convolutional layers, the number of neurons per convolutional layer, kernel size, the number of pooling operations, the kernel size of pooling layers, the number of dense layers, the number of neurons per dense layer, the activation function, as well as the BatchNormalization and Dropout layers. Although CNN is typically associated with image processing, 1D-CNN has proven effective for tabular data by capturing local feature dependencies. In our architecture, the input vector of material properties (SSA, Pore Volume, etc.) is treated as a 1D sequence. The convolutional layer applies learnable filters across this sequence, extracting non-linear interactions between adjacent features—effectively performing automated feature engineering—before passing the data to the fully connected layers. The final architecture of CNN, detailed in Table 4, comprises a feature extraction procedure and a regression procedure. The network receives an input of 8 dimensions. The first layer is a Conv1D layer with 64 filters, a kernel size of 2, and a ReLU activation function. It includes three fully connected layers (FC1, FC2 and FC3) with 32, 8, and 32 neurons, respectively, and each using a ReLU activation function. To stabilize and accelerate training, a BatchNormalization layer with a momentum of 0.98 and epsilon of 0.006 is applied after FC1. Before applying a Dropout layer with a rate of 0.25, a multiplication operation is performed between FC1 and FC3. The output layer consisted of 1 neuron with a linear activation function, suitable for regression tasks.

ARTICLE IN PRESS

Table 4. Chosen parameters for CNN and RF.

CNN	Parameters
Input	8
Conv1D	64, 2, ReLu
Flatten	–
FC1	32, ReLu
BatchNormalization	0.98, 0.006
FC2	8, ReLu
FC3	32, ReLu
Multiplication	FC1xFC3
Dropout	0.25
Output	1, Linear
RF	**Parameters**
Input	8
number of estimators	300
max depth	30
max features	sqrt
Output	1

The CNN was trained using the RMSprop optimizer with Mean Squared Error (MSE) as the loss function, a batch size

ARTICLE IN PRESS

of 16, and for 200 epochs. During training, a learning rate reduction strategy was employed, beginning with a learning rate of 0.004 and progressively decreasing it with a patience of 4 epochs, a reduction factor of 0.9, and a minimum learning rate of 0.0001. The ModelCheckpoint function was utilized to save the weights of the best model from the 6 folds, facilitating subsequent predictions on the test dataset. Performance was evaluated using the Root Mean Squared Error (RMSE) and the coefficient of determination (R^2), defined by the following equations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{1}$$

where y_i are the actual values and $\hat{y}_i$ are the predicted values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{2}$$

where $\bar{y}$ is the mean of the actual values.

Table 4 also presents the baseline (RF) architecture used for comparison with the CNN, along with its hyperparameters. The RF model uses 300 estimators, each with a maximum tree depth of 30. This setup aims to create a strong ensemble learning method by averaging the predictions from multiple decision trees to improve accuracy and control overfitting. The comparison between the RF baseline and the CNN architecture aims to evaluate the performance of traditional ensemble methods against deep learning techniques in modeling complex patterns within the data.

Mono and Bi-objective Optimization in Optuna

As established in the literature, the relationship between average pore diameter (AVD) and hydrogen adsorption is complex and influenced by factors such as pore shape, size distribution, temperature, pressure, and adsorbent properties. Generally, micropores exhibit higher hydrogen uptake due to stronger adsorption forces at their walls, while mesopores show lower uptake due to weaker forces. Therefore, after the models training, the CNN model was used by the Optuna optimizer to determine the maximum potential of the ACs for hydrogen storage based on the AVD. The optimization procedure consisted of two steps: (i) training the neural network to accurately predict hydrogen storage, and (ii) tuning input parameters to maximize H_2 uptake while minimizing AVD, according to the following objective function:

$$AVD = 4 \times \left(\frac{V_{mic} + V_{mes}}{SSA}\right) \times 10^3 \tag{3}$$

It should be noted that the average pore diameter (AVD) is calculated from the specific surface area (SSA) and pore volume, which introduces a degree of mathematical dependence between these structural descriptors. Therefore, the optimization objectives are not entirely independent. In this study, AVD is interpreted as a structural constraint representing the characteristic pore size of the adsorbent, while hydrogen uptake reflects adsorption performance. The multi-objective optimization is therefore intended to identify synthesis conditions that balance high surface area with controlled pore structure, which is a common practical consideration in porous material design.

The workflow adopted in this study integrates the dataset, AI model, and Optuna optimization framework, as illustrated in Figure 3. The process begins with configuring the ANN architecture and Optuna hyperparameters, followed by loading and preparing the dataset for model training. ~~Figure 3 presents a diagram that outlines the dataset, AI model, and Optuna optimization procedures, providing a comprehensive overview. The process begins with configuring the ANN and Optuna parameters, followed by loading the dataset.~~

The ANN model was trained using a dataset comprising seven input parameters: Agent, Ratio, SSA, Vmic, Vmes, Pressure, and Temperature. The model predicted H_2 and AVD. Hyperparameter optimization was conducted using Optuna to refine the ANN's architecture. A multi-objective optimization approach was employed to maximize H_2 production while minimizing AVD, involving 1000 trials across the input parameter space. The optimal solutions were visualized on a Pareto front. Figure 4 provides a graphical representation of this optimization process.

The entire process was implemented using Python, with the Keras and scikit-learn (sklearn) libraries. The platform used for this work is equipped with a 12th Gen Intel(R) Core(TM) i7-12700H processor, operating at 2.3 GHz and 16 gigabytes (GB) of RAM (random-access memory).

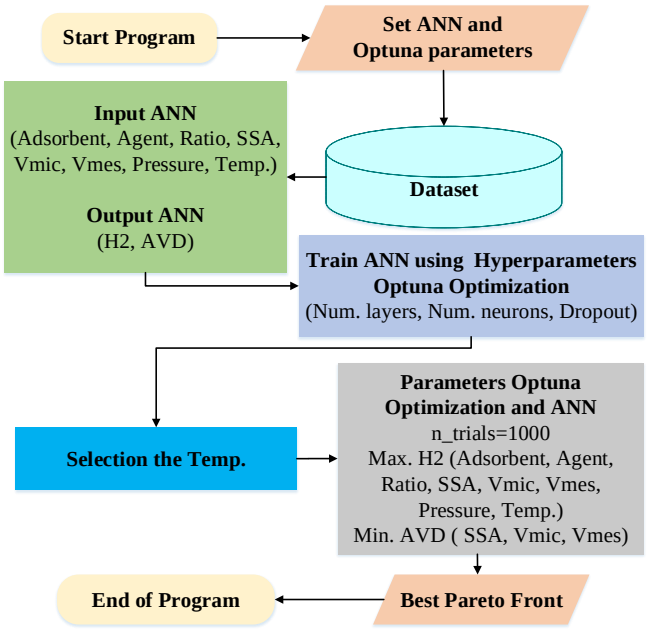

Figure 3. Diagram of AiDO methodology.

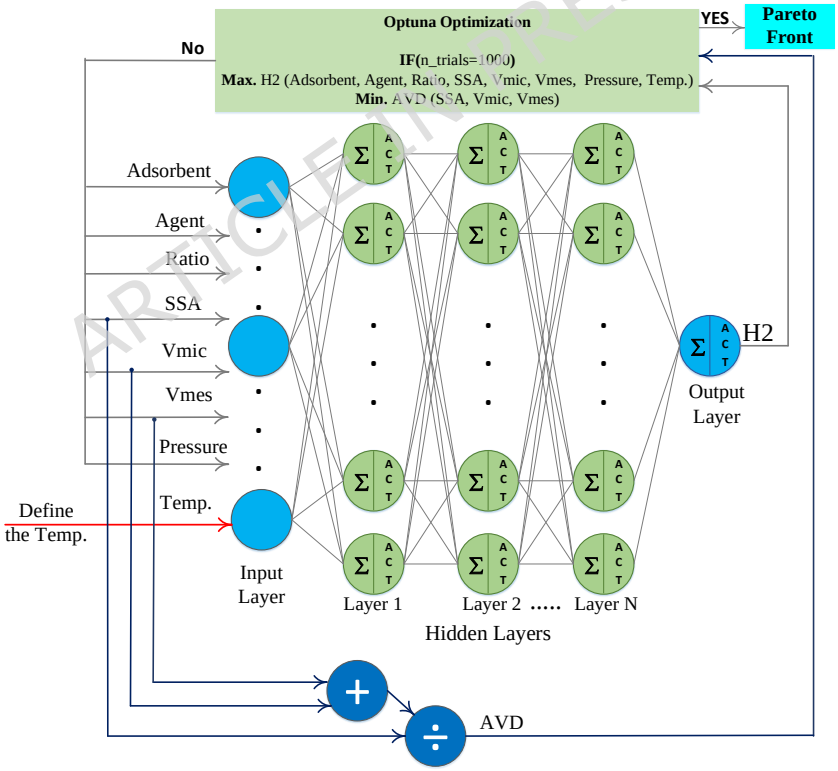

Figure 4. ANN model and Optuna optimization parameters.

Results and Discussion

The dataset, focused on hydrogen storage using activated carbons (AC), included 566 patterns from Rahimi et al.[41] and 351 from other studies. It covers variables related to adsorption and adsorbent properties, such as activation agent type, activating

ARTICLE IN PRESS

agent ratio (mass ratio of agent to carbon precursor), specific surface area (SSA), micropore (Vmic) and mesopore (Vmes) volumes, average pore diameter (AVD), pressure (0.70–298 bar), temperature (1–303 K), and hydrogen uptake (0–12.60 wt%).

Examining correlations among variables helps identify which features most strongly influence hydrogen uptake and can guide feature selection or interpretation of model outputs. It also highlights potential dependencies between derived and fundamental descriptors. The correlation matrix of the dataset, displayed in Figure 5, reveals key relationships between variables. SSA, Ratio, and Agent exhibit moderate positive correlations with H_2 uptake, proving that increases in these variables enhance hydrogen storage. Vmic also shows positive, though weaker, correlation. In contrast, temperature has a moderate negative correlation, confirming that lower temperatures are essential for effective hydrogen storage.Pressure and AVD exhibit weak negative correlations, implying minimal impact on storage. The negative correlation with AVD indicates that smaller pore diameters favor higher hydrogen uptake. However, since AVD is derived from Vmic and SSA, as shown in Equation (3), its effect reflects the combined influence of these parameters: maximizing the surface area while minimizing the average pore diameter increases hydrogen uptake. When considering the pore size distribution in more detail, specifically, maximizing Vmic (pores < 2 nm) and minimizing Vmes (pores between 2–5 nm) leads to improved hydrogen storage performance.

In order to evaluate the model's performance and insights, this section is divided into three subsections: Machine Learning Results, Optimization Results and Sensitivity Analysis. First, we compare predicted H_2 uptake values from the developed model to the collected data to assess its accuracy. Second, the optimization results are presented to highlight the model's potential in identifying optimal hydrogen storage conditions. Finally, a sensitivity analysis is conducted by restricting SSA and, subsequently, both SSA and temperature within dataset boundaries to explore model behavior under constrained conditions.

ARTICLE IN PRESS

Figure 5. Correlation matrix of selected material properties and H_2 storage metrics. The heatmap displays the Pearson correlation coefficients between various parameters, with red indicating positive correlations and blue indicating negative correlations.

The Machine Learning Results

Two machine learning models were developed and trained: a Random Forest (RF) and a Convolutional Neural Network (CNN). Their performance was evaluated using cross-validation, with results summarized in Table 5 in terms of Root Mean Squared Error (RMSE) and the coefficient of determination (R^2). Both models demonstrated strong predictive capabilities. The CNN achieved the best result in Folder 4, with an RMSE of 0.0406 and an R^2 of 0.9353, indicating high accuracy and model fit.

The CNN was trained using the RMSprop optimizer with Mean Squared Error (MSE) as the loss function, a batch size of 16, and for 200 epochs. Figure 6 illustrates the CNN's convergence during training and validation in Folder 5, illustrating that the model effectively learns the underlying patterns in the dataset. The validation curve (in blue) closely follows the training curve (in red), both showing a decreasing oscillation trend over time. At epoch 184, the training MSE in Checkpoint was 0.0003,indicating ~~effective learning from the training data, while the validation MSE was 0.0003.~~ that the model achieves stable learning without overfitting. This convergence is important because it demonstrates that the CNN is capable of reliably predicting hydrogen uptake for unseen data, ensuring the robustness of subsequent optimization results.

ARTICLE IN PRESS

Table 5. CNN and RF performance metrics (RMSE and R^2) for each fold.

Fold	CNN		RF	
	Test	R^2	Test	R^2
1	0.0461	0.9162	0.0475	0.9113
2	0.0460	0.9167	0.0437	0.9247
3	0.0447	0.9214	0.0465	0.9148
4	0.0406	0.9353	0.0443	0.9228
5	0.0465	0.9148	0.0411	0.9337
6	0.0460	0.9166	0.0470	0.9129

Figure 6. Convergence curves of the CNN during training (red) and validation (blue) in Folder 1, using RMSprop optimizer and MSE loss over 200 epochs.

ARTICLE IN PRESS

Table 6 presents the test results, including standard deviations (STD), for both CNN and RF models. Both models achieved a mean training RMSE of 0.0450, and an R^2 value of 0.92. While the CNN model exhibited lower standard deviations for both RMSE and R^2, the RF outperformed in fold 5 (Table 5).

Table 6. RMSE and R^2 test results with standard deviation (STD) using CNN and RF.

ML	RMSE	STD	R^2	STD
CNN	0.0450	0.0021	0.9202	0.0071
RF	0.0450	0.0022	0.9200	0.0079

Validating the model's predictive accuracy ensures that it reliably captures the relationships between input features and hydrogen uptake, which is critical before using the model for optimization. The correlation between real H_2 values (from the dataset, y-axis) and model-predicted H_2 values (x-axis) using the CNN is presented in figure 7. The strong linear relationship between the two datasets, quantified by an R^2 value of 0.94, indicates that the model captures 94% of the variation in actual H_2, demonstrating the model's solid performance and predictive accuracy. Furthermore, the near overlap of the regression line (green), given by the equation $y = 1.00x + 0.00$, with the reference line (red, y=x), which represents an ideal scenario where predicted values perfectly match real values, further confirms the strong agreement between the predicted and actual hydrogen uptake. This validation is crucial because it confirms that the model can be trusted for subsequent optimization tasks, where accurate predictions are needed to identify promising material designs.

Optimization Result

Optimization aims to find the best solution(s) for one or more objectives by maximizing performance or minimizing value within defined constraints. In this study, we aimed to maximize hydrogen uptake while achieving an optimal average pore diameter (AVD), a critical parameter for effective hydrogen storage. AVD was calculated as part of the analysis, and among all

ARTICLE IN PRESS

Figure 7. Real (from dataset) versus Predicted test data of hydrogen uptake using the CNN model. The red dashed line (Reference line: y=x) represents the ideal scenario where predicted values perfectly match the real data. Blue dots (Data points) show the individual predictions, and the green line (Regression line) illustrates the best-fit trend. The displayed R^2 value and regression equation show the model's predictive accuracy on the test dataset.

trained targets, hydrogen storage capacity is the only property where higher values are always better, whereas other indicators are adjusted depending on the requirements of specific applications.

In multi-objective optimization, the outcome is a Pareto front which is a set of optimal solutions that illustrate trade-offs between conflicting objectives. Unlike mono-objective problems, which yield a single best result, multi-objective problems identify knee points on the Pareto front, where small improvements in one objective lead to significant compromises in others[73]. Figure 8 presents the Pareto front illustrating the trade-off between H_2 uptake and AVD obtained through the optimization process. To investigate this trade-off, three representative points were selected: one with the lowest AVD, one with the highest H_2 uptake, and a third near the knee point, representing a balanced compromise between the two objectives.

The corresponding values for Agent, Ratio, SSA, Vmic, Vmes, Pressure, and Temperature are summarized in Table 7. A key novelty of this study is that the optimized model generated these parameter sets, which were not present in the original experimental data, demonstrating its potential to design novel synthesis conditions. For instance, the "Knee Point" solution predicts a high H_2 uptake of 16.66 %wt at an exceptionally low AVD of 0.08 nm. As shown in the table, achieving this requires an Agent of 3.00, a Ratio of 2.81, and an SSA of 5292.48 m^2g^{-1}, among other conditions.

Table 7. Optimized parameters for three representative solutions from the Pareto front: minimizing AVD (Min. AVD), maximizing H_2 uptake (Max H_2), and near the knee point (Knee).

	Min. AVD	**Knee**	**Max. H_2**
Adsorbent	0.00	0.00	0.00
Agent	3.00	3.00	4.00
Ratio	1.96	2.81	2.81
SSA (m^2g^{-1})	5309.43	5292.48	5292.48
Vmic (cm^3g^{-1})	0.06	0.06	0.06
Vmes (cm^3g^{-1})	0.05	0.05	1.28
Pressure (bar)	32.70	111.72	26.27
Temp. (K)	7.74	112.68	121.75
H_2 (%wt)	3.11	16.66	**18.67**
AVD (nm)	**0.08**	0.08	1.01

~~Figure 9 provides a comparative visualization of the three representative Pareto solutions, normalized according to the~~

ARTICLE IN PRESS

Figure 8. Pareto front solutions illustrating the trade-off between H_2 uptake and AVD, as obtained from the optimization process. Gray circles represent Pareto-optimal solutions, the green circle indicates the minimum AVD solution, the blue square the maximum H2 solution, and the red diamond the knee (compromise) solution.

~~global optimization limits enabling the direct comparison across the different parameters.~~ To better understand how different Pareto-optimal solutions balance competing objectives, it is useful to compare their characteristics across all relevant parameters. A radar chart provides an effective visualization for this purpose, allowing the simultaneous assessment of adsorbent descriptors, structural properties, operating conditions, and performance objectives. The normalized radar chart, (figure 9), highlights how each solution prioritizes different trade-offs between adsorbent characteristics (adsorbent, agent, ration), structural properties (SSA, Vmic, Vmes), operating conditions (pressure and temperature), and performance objectives (AVD and H2 uptake). The knee solution (presented in blue) offers the most balanced trade-off, achieving the maximum normalized AVD score (1.0) while retaining a high H2 score (0.56). In contrast, the remaining solutions correspond to extreme cases: the “Max. H2” profile (in red) maximizes H2 (1.0) at the expense of AVD (0.0), whereas the “Min. AVD” profile (in green) exhibits the opposite behavior.

ARTICLE IN PRESS

Figure 9. Normalized radar chart comparing the parameters of the three representative Pareto front solutions: Min. AVD (green), Knee (blue), and Max. H2 (red). All parameters are scaled according to the global optimization limits.

To evaluate the ability of the AI-driven framework to propose novel material designs beyond existing data, it is informative to compare optimized solutions with the original dataset. This comparison demonstrates whether the model can identify promising

ARTICLE IN PRESS

candidates that surpass previously observed performance. The trade-off between the original dataset and the Pareto-optimal solutions is presented in Figure 10. ~~Figure 10 plots the optimized Pareto front against the original experimental dataset.~~ The figure clearly illustrates the model's primary capability: it identifies optimal solutions (gray circles) that lie far beyond the range of the original data (blue dots). This demonstrates that the AI-driven approach can design novel parameters, achieving significantly higher H_2 uptake at low AVD values than what was previously observed in the experimental data.

Figure 10. AVD–H2 trade-off between the original dataset and Pareto-optimal solutions. Blue dots show the original data, gray circles the Pareto front, and the red diamond the knee point.

Figure 11 compares the normalized values of the optimized solutions against the original dataset's range and the absolute extrapolation limits. This normalization scheme was designed to allow the model to extrapolate beyond the maximum values of the original experimental data (yellow diamonds). The figure clearly indicates that the optimizer pushed the values for **SSA** and **Pressure** to their maximum extrapolation limit (red stars), exceeding the dataset's original maximum. Conversely, the optimized values for other parameters, such as H_2, AVD, Ratio, and Vmic, did not reach their upper limits. This suggests that achieving H_2 values significantly higher than those in the original dataset requires exploring SSA and Pressure values far beyond existing experimental data.

ARTICLE IN PRESS

Sensitivity analysis

To assess the sensitivity of our methodology, we applied the Optuna-CNN combination using various constraints. This analysis aimed to evaluate the impact of different constraint approaches on model performance and optimization outcomes. By comparing the results from different constraints, we can identify the most effective method for enhancing the accuracy and generalization capabilities of the CNN model when predicting hydrogen adsorption in ACs.

Restricted SSA

The aim of this analysis was to provide insights into the model's ability to optimize hydrogen adsorption within realistic experimental boundaries. Figure 12 presents the Pareto front for both objectives when SSA was limited to 4300 m^2/g. The resulting Pareto front shows that the model struggles to achieve high H_2 uptake under this limitation, peaking significantly lower than the main optimization results. This contrasts with the Pareto front in Figure 10, which shows the model identifying solutions with much higher H_2 uptake (up to 18.67 %wt when not restricted by the dataset's maximum SSA values. The optimization under the SSA constraint indicates a clear trade-off: within this limited range, higher H_2 uptake is only achieved at the expense of higher AVD values.

Restricted SSA and Temperature

A second analysis constrained both SSA to 4300 $m^2 g^{-1}$ and temperature to 77 K, with the resulting Pareto front shown in Figure 13. As noted in the text, fixing the temperature at 77 K (a highly favorable condition for physisorption) allowed the model to find solutions with higher H_2 uptake compared to the SSA-only constraint scenario (Figure 12).

However, the performance in this scenario is still significantly lower than the unrestricted optimization shown in Figure 10, which achieved H_2 uptake as high as 18.67 %wt. This comparison reinforces the insight from Figure 10: the model's ability to

ARTICLE IN PRESS

Figure 11. Normalized values of the optimized variables compared to data limits. The scatter plot shows the optimized values (green) alongside the minimum (blue) and maximum (yellow) data points, as well as the defined minimum (black) and maximum (red) range for each variable.

ARTICLE IN PRESS

Figure 12. AVD–H2 trade-off between the original dataset and Pareto-optimal solutions with SSA at limit 4300 $m^2 g^{-1}$. Blue dots show the original data, gray circles the Pareto front, and the red diamond the knee point.

predict breakthrough H_2 values in the main optimization relied heavily on its freedom to extrapolate SSA values far beyond the 4300 m^2/g limit.

Overall, Figure 13 underscores the model's capability to efficiently manage these more realistic dual constraints, identifying the optimal trade-offs achievable within those given boundaries.

Figure 14 illustrates the optimization under the most severe constraints applied, limiting SSA to 2000 $m^2 g^{-1}$ while fixing temperature at 77 K. The impact of this SSA restriction is drastic when compared to the main optimization (Figure 10). While the unrestricted model predicted H_2 uptake above 18 %wt by extrapolating SSA to high values, this scenario struggles to exceed 8 %wt, demonstrating that a high SSA was the most critical factor for achieving the breakthrough results in Figure 10.

Furthermore, this scenario's performance is significantly lower than the intermediate constraint (Figure 13), which used a 4300 $m^2 g^{-1}$ SSA limit at the same 77 K temperature and achieved results around 12.5 %wt. By reducing the SSA limit from

ARTICLE IN PRESS

Figure 13. AVD–H2 trade-off between the original dataset and Pareto-optimal solutionswith a temperature limit of 77 K and an SSA limit of 4300 m^2g^{-1}. Blue dots show the original data, gray circles the Pareto front, and the red diamond the knee point.

4300 to 2000 m^2g^{-1}, the model's predicted H_2 capacity was nearly halved, confirming that SSA, not just temperature, acts as a primary bottleneck.

Overall, the Pareto front in Figure 13 is the one that most closely resembles the original experimental dataset (the blue data points). It validates the model's ability to optimize within realistic boundaries, but it also confirms that achieving high-performance H_2 storage is not possible without developing materials that substantially exceed an SSA of 2000 m^2g^{-1}.

ARTICLE IN PRESS

Figure 14. AVD–H2 trade-off between the original dataset and Pareto-optimal solutionswith a temperature limit of 77 K and an SSA limit of 2000 m^2g^{-1}. Blue dots show the original data, gray circles the Pareto front, and the red diamond the knee point.

Physical Plausibility of Optimized Designs

Table 8 summarizes the results of the sensitivity analysis, comparing the optimal compromise solution (Knee Point) across the four optimization scenarios. This comparison highlights the trade-offs between H_2 uptake and AVD under varying experimental constraints.

ARTICLE IN PRESS

The main optimization (referenced as Figure 10), which used the full normalization limits for extrapolation, identified an ideal knee point at 16.66 %wt and a derived AVD of 0.08 nm. This sub-ångström AVD should be interpreted as an extrapolation-driven, equivalent-diameter outcome rather than a physically accessible pore width. In this work, the "average pore diameter" (AVD) is an equivalent diameter computed from the ratio of total pore volume to specific surface area (Eq. 3), rather than a directly measured minimum pore width. As a consequence, when the optimizer pushes the pore-volume-to-surface-area ratio toward extreme values during unconstrained extrapolation, the derived AVD can drop to sub-Ångström values. The unconstrained knee-point solution (16.66 wt% at AVD = 0.08 nm; Table 8) should therefore not be interpreted as a physically accessible pore size in porous carbons. This value is far below the kinetic diameter of H_2 ($\sim$ 0.289 nm), implying that such "pores" would not be accessible to hydrogen molecules and lie outside the physically meaningful range of porous-carbon textural characterization[74]. In addition, commonly used pore-size classifications place ultramicropores below $\sim$ 0.7 nm, which already represents the smallest practically relevant micropore range for adsorption[75]. The constraint-based sensitivity analysis (Table 8) provides a physically meaningful interpretation of the optimization results. When the search space is limited to experimentally realistic bounds, the compromise (knee-point) solutions shift to AVD $\approx$ 0.7 nm (Constraints 1–2) and AVD $\approx$ 1.1 nm (Constraint 3), along with lower but more credible H_2 uptake values. These AVD values fall within the microporous regime and are consistent with the idea that pores around 0.6–0.7 nm can enhance H_2 uptake per unit surface area at cryogenic temperatures and elevated pressures due to overlapping adsorption fields from opposing pore walls[76]. Accordingly, we view the unconstrained solution as a theoretical upper bound enabled by extrapolation, whereas the constrained solutions in Table 8 define a physically plausible design space for porous carbons and provide more practical guidance for experimental synthesis and validation. ~~When the first constraint was applied (Figure 12), limiting only the SSA to 4300 m^2g^{-1}, the performance remained high at approximately 15.8 %wt, though at a higher AVD of $\approx$ 0.7 nm. However, when the temperature was also realistically constrained to 77 K (Figure 13), the optimal compromise dropped to approximately 12.5 %wt. Finally, the most severe constraint scenario, depicted in Figure 14 (and labeled as Constraint 3 in the table), limited the SSA to 2000 m^2g^{-1} and the temperature to 77 K. This caused a significant performance drop to $\approx$ 6.5 %wt, a value much closer to the original dataset. This comparison demonstrates that the high SSA values (enabled by extrapolation) were the most critical factor in achieving the breakthrough H_2 storage performance predicted in the main optimization.~~

Table 8. Comparison of Compromise Solutions (Knee Points) across all optimization scenarios.

Optimization Scenario	Applied Constraints	Compromise Solution (Knee Point)
Main optimization (Figure 10)	limits of normalization	16.66 %wt @ 0.08 nm
Constraint 1 (Figure 12)	SSA $\leq$ 4300 m^2g^{-1}	$\approx$ 15.8 %wt @ $\approx$ 0.7 nm
Constraint 2 (Figure 13)	SSA $\leq$ 4300 m^2g^{-1} and Temp $\leq$ 77 K	$\approx$ 12.5 %wt @ $\approx$ 0.7 nm
Constraint 3 (Figure 14)	SSA $\leq$ 2000 m^2g^{-1} and Temp $\leq$ 77 K	$\approx$ 6.5 %wt @ $\approx$ 1.1 nm

To contextualize the optimization output, we benchmark the knee-point excess H_2 uptake values in Table 8 against literature-reported ranges for gravimetric excess adsorption under cryogenic conditions. For activated carbons at 77 K, multiple studies report an apparent upper envelope near $\sim$6–7 wt% excess, even for highly activated samples, reflecting a practical saturation of cryogenic physisorption performance for conventional carbons[20]. Encouragingly, our most restrictive scenario (Constraint 3: SSA $\leq$ 2000 m^2 g^{-1} and $T \leq$ 77 K) yields a knee-point of $\sim$6.5 wt% excess, which aligns closely with this established benchmark. In contrast, relaxing the textural constraint while maintaining cryogenic operation (Constraint 2: SSA $\leq$ 4300 m^2 g^{-1} and $T \leq$ 77 K) increases the predicted compromise value to $\sim$12.5 wt% excess (Table 8). A common heuristic for cryogenic excess adsorption (often referred to as Chahine's rule) suggests roughly $\sim$1 wt% excess per $\sim$500 m^2 g^{-1} increase in surface area at 77 K, implying $\sim$8–9 wt% excess for SSA $\approx$ 4300 m^2 g^{-1}; therefore, the higher predicted value indicates that the optimization is also leveraging favorable pore-structure descriptors within the allowable bounds of the search space[77]. For comparison, benchmark high-surface-area porous materials such as MOF-177 have reported surface-excess uptakes on the order of $\sim$7–8 wt% under comparable cryogenic/high-pressure testing[20]. Overall, these comparisons indicate that the constrained optimum most representative of conventional materials (Constraint 3) falls within well-established experimental ranges, whereas the higher-uptake constrained solutions should be regarded as ambitious performance targets requiring experimental confirmation.

Study Limitations and Future Recommendations

The application of the proposed model is subject to several limitations. First, the training dataset is largely clustered around common experimental conditions, particularly temperatures near 77 K and 298 K, which are widely used in hydrogen adsorption measurements. As a result, predictions at intermediate or less-represented conditions rely on the model's interpolation capabilities and may carry additional uncertainty. Second, the optimization algorithm identified candidate solutions associated

ARTICLE IN PRESS

with specific surface area (SSA) values exceeding 5000 m^2/g, whereas the maximum SSA observed in the training dataset is approximately 4300 m^2/g. These predictions therefore represent model extrapolation beyond the available data range and should be interpreted as hypothetical material designs that require experimental validation. Future work could improve the robustness of the framework by expanding the dataset with additional experimental measurements and by integrating physics-informed constraints to ensure that optimized material parameters remain within physically realistic ranges.

Conclusion and Future Work

This research explored the application of artificial intelligence to optimize hydrogen storage in porous carbon adsorbents. By developing predictive machine learning models and integrating them into an optimization framework, we aimed to maximize hydrogen uptake (wt%) while minimizing average pore diameter (AVD).

We utilized a comprehensive dataset of 917 data points extracted from the literature to develop two models: Random Forest (RF) and Convolutional Neural Network (CNN). Both models demonstrated strong capabilities in accurately estimating hydrogen uptake based on material properties (Agent, Ratio, SSA, Vmic, Vmes) and experimental parameters (Pressure and Temperature). The CNN model, in particular, exhibited satisfactory performance with an RMSE of 0.0406 and an R^2 of 0.9353. The model demonstrated high accuracy in interpolation tasks within the dataset range. In the subsequent optimization phase, the framework explored extrapolative regions of the design space, identifying theoretical material configurations with predicted hydrogen uptake values beyond those currently reported in experimental datasets. These models were then integrated into an optimization framework to identify optimal properties and parameters that maximize hydrogen storage while minimizing large average pore diameters. The strong linear correlation between actual and predicted H_2 uptake values, with a correlation of 94%, underscores the model's predictive power. The optimized CNN model successfully predicted hydrogen uptake values without solely relying on an existing dataset, showcasing its capability to generate optimal parameters beyond the range of the experimental data. For instance, at the knee point, the model predicted a H_2 uptake of 16.66% by weight with an AVD of 0.08 nm.

Additionally, the research examined the sensitivity of the methodology to different data normalization, further validating the robustness of the Optuna-CNN approach in optimizing hydrogen storage conditions. The ability to design parameters that achieve high hydrogen uptake highlights the practical implications of this study, potentially reducing experimental time and costs in the development of porous carbon adsorbents for hydrogen storage.

Future research could explore incorporating additional material properties, such as different activation agents, to enhance predictive accuracy further. Moreover, applying these methodologies to other types of adsorbents, like metal-organic frameworks (MOFs), or different adsorbates, such as CO_2 or methane, could broaden the scope and impact of this work, potentially leading to significant advancements in gas storage and related fields.

From an industrial perspective, the developed AiDO framework can function as a virtual high-throughput screening tool for porous carbon design. Instead of physically synthesizing and characterizing a large number of candidate materials, researchers can use the trained models to evaluate thousands of potential combinations of synthesis conditions and structural parameters. This enables the identification of promising material configurations prior to experimental validation, significantly reducing the time and resources required for material discovery. The optimized parameter ranges generated by the framework can then guide targeted synthesis and experimental testing, thereby accelerating the development of high-performance hydrogen storage materials.

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Supplementary material

Abbreviations used in Table 1 (Representative ML studies on gas storage materials):

References

1. Züttel, A. Materials for hydrogen storage. *Mater. today* **6**, 24–33 (2003).
2. Sakintuna, B., Lamari-Darkrim, F. & Hirscher, M. Metal hydride materials for solid hydrogen storage: a review. *Int. J. Hydrog. Energy* **32**, 1121–1140 (2007).
3. Firlej, L., Beckner, M., Romanos, J., Pfeifer, P. & Kuchta, B. Different approach to estimation of hydrogen-binding energy in nanospace-engineered activated carbons. *The J. Phys. Chem. C* **118**, 955–961 (2014).

ARTICLE IN PRESS

Abbreviation	Meaning
DT	Decision Tree
PR	Poisson Regression
SVM/SVR	Support Vector Machine / Regression
RF	Random Forest
LR	Linear Regression
RR	Ridge Regression
ANN	Artificial Neural Network
XGBoost/XGBT	Extreme Gradient Boosting
CatBoost	Categorical Boosting
LGB	LightGBM
GBDT	Gradient Boosted Decision Trees
MGBR	Multiple Gradient Boosting Regressor
MLP/MLPNN	Multi-Layer Perceptron
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
BDTR	Bayesian Decision Tree Regressor
BLR	Bayesian Linear Regression
NNR	Nearest Neighbor Regression
CFFNN	Cascade Feed-Forward Neural Network
GRNN	Generalized Regression Neural Network
RNN	Recurrent Neural Network
KNN	k-Nearest Neighbor
ET	Extra Tree
BRANN	Bayesian Regularized Artificial Neural Network
AB	AdaBoost
GB	Gradient Boosting
LSSVM	Least Squares Support Vector Machine
ANFIS	Adaptive Neuro-Fuzzy Inference System
ELM	Extreme Learning Machine
CMIS	Convolutional Mixed-Integer System

4. Romanos, J. *et al.* High surface area carbon and process for its production. *US Pat.* **9**, 445 (2016).
5. Romanos, J. *et al.* Cycling and regeneration of adsorbed natural gas in microporous materials. *Energy & Fuels* **31**, 14332–14337 (2017).
6. Rash, T. A. *et al.* Microporous carbon monolith synthesis and production for methane storage. *Fuel* **200**, 371–379 (2017).
7. Kuchta, B. *et al.* Open carbon frameworks - a search for optimal geometry for hydrogen storage. *J. Mol. Model.* **19**, 4079–4087 (2013).
8. Romanos, J., Barakat, F. & Dargham, S. A. Nanoporous graphene monolith for hydrogen storage. *Mater. Today: Proc.* **5**, 17478–17483 (2018).
9. Bannenberg, L. *et al.* Metal (boro-) hydrides for high energy density storage and relevant emerging technologies. *Int. J. Hydrog. Energy* **45**, 33687–33730 (2020).
10. DOE. Technical targets for onboard hydrogen storage for light-duty vehicles. *energy.gov* (2017).
11. Chae, H. K. *et al.* A route to high surface area, porosity and inclusion of large molecules in crystals. *Nature* **427**, 523 (2004).
12. Kuchta, B. *et al.* Hypothetical high-surface-area carbons with exceptional hydrogen storage capacities: open carbon frameworks. *J. Am. Chem. Soc.* **134**, 15130–15137 (2012).
13. Romanos, J. *et al.* Infrared study of boron–carbon chemical bonds in boron-doped activated carbon. *Carbon* **54**, 208–214 (2013).
14. Romanos, J. *et al.* Nanospace engineering of koh activated carbon. *Nanotechnology* **23**, 015401 (2012).
15. Jordá-Beneyto, M., Suárez-García, F., Lozano-Castelló, D., Cazorla-Amorós, D. & Linares-Solano, A. Hydrogen storage on chemically activated carbons and carbon nanomaterials at high pressures. *Carbon* **45**, 293–303 (2007).

ARTICLE IN PRESS

16. Reyhani, A., Mortazavi, S. Z., Moshfegh, S. M. A. Z., Parvin, P. & Golikand, A. N. Hydrogen storage in decorated multiwalled carbon nanotubes by ca, co, fe, ni, and pd nanoparticles under ambient conditions. *The J. Phys. Chem. C* **115**, 6994–7001 (2011).

17. Yang, R. T. Hydrogen storage by alkali-doped carbon nanotubes-revisited. *Carbon* **38**, 623–641 (2000).

18. Moussa, M. *et al.* Toward sustainable hydrogen storage and carbon dioxide capture in post-combustion conditions. *J. Environ. Chem. Eng.* **5**, 1628–1637 (2017).

19. Hu, W. *et al.* Hierarchically porous carbon derived from neolamarckia cadamba for electrochemical capacitance and hydrogen storage. *ACS Sustain. Chem. & Eng.* **7**, 15385–15393 (2019).

20. Sevilla, M. & Mokaya, R. Energy storage applications of activated carbons: supercapacitors and hydrogen storage. *Energy & Environ. Sci.* **7**, 1250 (2014).

21. Sevilla, M., Fuertes, A. B. & Mokaya, R. High density hydrogen storage in superactivated carbons from hydrothermally carbonized renewable organic materials. *Energy & Environ. Sci.* **4**, 1400–1410 (2011).

22. Blankenship, L. S., Balahmar, N. & Mokaya, R. Oxygen-rich microporous carbons with exceptional hydrogen storage capacity. *Nat. Commun.* **8**, 1545 (2017).

23. Pardakhti, M., Moharreri, E., Wanik, D., Suib, S. L. & Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (mofs). *ACS Comb. Sci.* **19**, 640–645 (2017).

24. Kim, S.-Y., Kim, S.-I. & Bae, Y.-S. Machine-learning-based prediction of methane adsorption isotherms at varied temperatures for experimental adsorbents. *The J. Phys. Chem. C* **124**, 19538–19547 (2020).

25. Meng, M., Zhong, R. & Wei, Z. Prediction of methane adsorption in shale: Classical models and machine learning based models. *Fuel* **278**, 118358 (2020).

26. Abdi, J., Hadavimoghaddam, F., Hadipoor, M. & Hemmati-Sarapardeh, A. Modeling of co2 adsorption capacity by porous metal organic frameworks using advanced decision tree - based models. *Sci. Reports* **11**, 24468 (2021).

27. Yuan, X., Suvarna, M., Low, S., Lee, P. D. D. B. & andYong Sik Ok, J. L. W. Applied machine learning for prediction of co2 adsorption on biomass waste-derived porous carbons. *Environ. Sci. & Technol.* **55**, 11925–11936 (2021).

28. Maheri, M., Bazan, C., Zendehboudi, S. & Usefi, H. Machine learning to assess co2 adsorption by biomass waste. *J. CO2 Util.* **76**, 102590 (2023).

29. Alatefi, S., Agwu, O. E., Nait Amar, M. & Alkouh, A. Advancing hydrogen storage: Explainable machine learning models for predicting hydrogen uptake in metal-organic frameworks. *Results Eng.* **28**, 107304 (2025).

30. Rahnama, A., Zepon, G. & Sridhar, S. Machine learning based prediction of metal hydrides for hydrogen storage, part i: Prediction of hydrogen weight percent. *Int. J. Hydrog. Energy* **44**, 7337–7344 (2019).

31. Rahnama, A., Zepon, G. & Sridhar, S. Machine learning based prediction of metal hydrides for hydrogen storage, part ii: Prediction of material class. *Int. J. Hydrog. Energy* **44**, 7345–7353 (2019).

32. Alizadeh, S. M. S., Parhizi, Z., Alibak, A. H., Vaferi, B. & Hosseini, S. Predicting the hydrogen uptake ability of a wide range of zeolites utilizing supervised machine learning methods. *Int. J. Hydrog. Energy* **47**, 21782–21793 (2022).

33. Salehi, K., Rahmani, M. & Atashrouz, S. Machine learning assisted predictions for hydrogen storage in metal-organic frameworks. *Int. J. Hydrog. Energy* **48**, 33260–33275 (2023).

34. Meduri, S. & Nandanavanam, J. Prediction of hydrogen uptake of metal organic frameworks using explainable machine learning. *Energy AI* **12**, 100230 (2023).

35. Kusdhany, M. I. M. & Lyth, S. M. New insights into hydrogen uptake on porous carbon materials via explainable machine learning. *Carbon* **179**, 190–201 (2021).

36. Davoodi, S. *et al.* Machine-learning models to predict hydrogen uptake of porous carbon materials from influential variables. *Sep. Purif. Technol.* **316**, 123807 (2023).

37. Wang, C.-S. & Brinkerhoff, J. Predicting hydrogen adsorption and desorption rates in cylindrical metal hydride beds: Empirical correlations and machine learning. *Int. J. Hydrog. Energy* **46**, 24256–24270 (2021).

38. Anderson, G., Schweitzer, B., Anderson, R. & Gómez-Gualdrón, D. A. Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: Molecular simulation and machine learning. *The J. Phys. Chem. C* **123**, 120–130 (2019).

ARTICLE IN PRESS

ARTICLE IN PRESS

39. Batalović, K., Radaković, J., Kuzmanović, B., Ilić, M. M. & Mamula, B. P. Machine learning-based high-throughput screening of mg-containing alloys for hydrogen storage and energy conversion applications. *J. Energy Storage* **68**, 107720 (2023).

40. Rossetti, I., Ramis, G., Gallo, A. & Di Michele, A. Hydrogen storage over metal-doped activated carbon. *Int. J. Hydrog. Energy* **40**, 7609–7616 (2015).

41. Rahimi, M., Abbaspour-Fard, M. H. & Rohani, A. Machine learning approaches to rediscovery and optimization of hydrogen storage on porous bio-derived carbon. *J. Clean. Prod.* **329**, 129714 (2021).

42. Alatefi, S., Nait Amar, M., Agwu, O. E. & Alkouh, A. Accurate prediction of water activity in ionic liquid-based aqueous ternary solutions using advanced explainable artificial intelligence frameworks. *Chem. Eng. Sci.* **318**, 122218 (2025).

43. Thanh, H. V. *et al.* Hydrogen storage on porous carbon adsorbents: rediscovery by nature-derived algorithms in random forest machine learning model. *Energies* **16**, 2348 (2023).

44. Huang, C.-C., Chen, H.-M., Chen, C.-H. & Huang, J.-C. Effect of surface oxides on hydrogen storage of activated carbon. *Sep. Purif. Technol.* **70**, 291–295 (2010).

45. Sun, Y. & Webley, P. A. Preparation of activated carbons from corncob with large specific surface area by a variety of chemical activators and their application in gas storage. *Chem. Eng. J.* **162**, 883–892 (2010).

46. Fierro, V., Zhao, W., Izquierdo, M., Aylon, E. & Celzard, A. Adsorption and compression contributions to hydrogen storage in activated anthracites. *Int. J. Hydrog. Energy* **35**, 9038–9045 (2010).

47. González-Navarro, M. F., Giraldo, L. & Moreno-Piraján, J. C. Preparation and characterization of activated carbon for hydrogen storage from waste african oil-palm by microwave-induced lioh basic activation. *J. Anal. Appl. Pyrolysis* **107**, 82–86 (2014).

48. Hwang, S.-H., Choi, W. M. & Lim, S. K. Hydrogen storage characteristics of carbon fibers derived from rice straw and paper mulberry. *Mater. Lett.* **167**, 18–21 (2016).

49. Bader, N. & Ouederni, A. Optimization of biomass-based carbon materials for hydrogen storage. *J. Energy Storage* **5**, 77–84 (2016).

50. Wang, D., Geng, Z., Zhang, C., Zhou, X. & Liu, X. Effects of thermal activation conditions on the microstructure regulation of corncob-derived activated carbon for hydrogen storage. *J. energy chemistry* **23**, 601–608 (2014).

51. Liu, X., Zhang, C., Geng, Z. & Cai, M. High-pressure hydrogen storage and optimizing fabrication of corncob-derived activated carbon. *Microporous mesoporous materials* **194**, 60–65 (2014).

52. Wang, H., Gao, Q. & Hu, J. High hydrogen storage capacity of porous carbons prepared by using activated carbon. *J. Am. Chem. Soc.* **131**, 7016–7022 (2009).

53. Wang, J., Senkovska, I., Kaskel, S. & Liu, Q. Chemically activated fungi-based porous carbons for hydrogen storage. *Carbon* **75**, 372–380 (2014).

54. Üner, O., Geçgel, Ü. & Avcu, T. Comparisons of activated carbons produced from sycamore balls, ripe black locust seed pods, and nerium oleander fruits and also their h2 storage studies. *Carbon Lett.* **31**, 75–92 (2021).

55. Jin, H., Lee, Y. S. & Hong, I. Hydrogen adsorption characteristics of activated carbon. *Catal. today* **120**, 399–406 (2007).

56. Figueroa-Torres, M. Z., Robau-Sánchez, A., De la Torre-Sáenz, L. & Aguilar-Elguézabal, A. Hydrogen adsorption by nanostructured carbons synthesized by chemical activation. *Microporous mesoporous materials* **98**, 89–93 (2007).

57. Minoda, A., Oshima, S., Iki, H. & Akiba, E. Synthesis of koh-activated porous carbon materials and study of hydrogen adsorption. *J. alloys compounds* **580**, S301–S304 (2013).

58. Balathanigaimani, M. *et al.* Nanostructured biomass based carbon materials from beer lees for hydrogen storage. *J. nanoscience nanotechnology* **18**, 2196–2199 (2018).

59. Bader, N. & Ouederni, A. Functionalized and metal-doped biomass-derived activated carbons for energy storage application. *J. Energy Storage* **13**, 268–276 (2017).

60. Cheng, F., Liang, J., Zhao, J., Tao, Z. & Chen, J. Biomass waste-derived microporous carbons with controlled texture and enhanced hydrogen uptake. *Chem. Mater.* **20**, 1889–1895 (2008).

61. Arshad, S. H. M. *et al.* Preparation of activated carbon from empty fruit bunch for hydrogen storage. *J. Energy Storage* **8**, 257–261 (2016).

ARTICLE IN PRESS

ARTICLE IN PRESS

62. Zhang, C. *et al.* Microstructure regulation of super activated carbon from biomass source corncob with enhanced hydrogen uptake. *Int. J. Hydrog. Energy* **38**, 9243–9250 (2013).

63. Ramesh, T., Rajalakshmi, N. & Dhathathreyan, K. S. Synthesis and characterization of activated carbon from jute fibers for hydrogen storage. *Renew. Energy Environ. Sustain.* **2**, 4 (2017).

64. Xiao, Y. *et al.* Melaleuca bark based porous carbons for hydrogen storage. *Int. J. Hydrog. Energy* **39**, 11661–11667 (2014).

65. Choi, Y.-K. & Park, S.-J. Hydrogen storage capacity of highly porous carbons synthesized from biomass-derived aerogels. *Carbon letters* **16**, 127–131 (2015).

66. Lee, S.-Y. & Park, S.-J. Effect of platinum doping of activated carbon on hydrogen storage behaviors of metal-organic frameworks-5. *Int. journal hydrogen energy* **36**, 8381–8387 (2011).

67. Fierro, V. *et al.* Experimental evidence of an upper limit for hydrogen storage at 77 k on activated carbons. *Carbon* **48**, 1902–1911 (2010).

68. Jordá-Beneyto, M., Lozano-Castelló, D., Suárez-García, F., Cazorla-Amorós, D. & Linares-Solano, Á. Advanced activated carbon monoliths and activated carbons for hydrogen storage. *Microporous Mesoporous Mater.* **112**, 235–242 (2008).

69. Kopac, T., Kırca, Y. & Toprak, A. Synthesis and characterization of koh/boron modified activated carbons from coal and their hydrogen sorption characteristics. *Int. J. Hydrog. Energy* **42**, 23606–23616 (2017).

70. Pedicini, R. *et al.* Posidonia oceanica and wood chips activated carbon as interesting materials for hydrogen storage. *Int. J. Hydrog. Energy* **45**, 14038–14047 (2020).

71. Doğan, M., Sabaz, P., Bııcııl, Z., Kizilduman, B. K. & Turhan, Y. Activated carbon synthesis from tangerine peel and its use in hydrogen storage. *J. Energy Inst.* **93**, 2176–2185 (2020).

72. Rowlandson, J. L., Edler, K. J., Tian, M. & Ting, V. P. Toward process-resilient lignin-derived activated carbons for hydrogen storage applications. *ACS Sustain. Chem. & Eng.* **8**, 2186–2195 (2020).

73. Rocha, H. R. O. *et al.* Optimizing a machine learning design of dielectric properties in lead-free piezoelectric ceramics. *Mater. & Des.* 113053 (2024).

74. Wang, H. *et al.* Superior hydrogen separation in nanofluidic membranes by synergistic effect of pore tailoring and host–guest interaction. *Nano Lett.* **25**, 9353–9361, DOI: 10.1021/acs.nanolett.5c01736 (2025). PMID: 40434398.

75. Thommes, M. *et al.* Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (iupac technical report). *Pure applied chemistry* **87**, 1051–1069 (2015).

76. Gogotsi, Y. *et al.* Importance of pore size in high-pressure hydrogen storage by porous carbons. *Int. J. Hydrog. Energy* **34**, 6314–6319, DOI: https://doi.org/10.1016/j.ijhydene.2009.05.073 (2009).

77. Broom, D. *et al.* Concepts for improving hydrogen storage in nanoporous materials. *Int. J. Hydrog. Energy* **44**, 7768–7779, DOI: https://doi.org/10.1016/j.ijhydene.2019.01.224 (2019). A special issue on hydrogen-based Energy storage.

Acknowledgements

The authors acknowledge the support of the Research Center Bio-based Economy (KCBBE) at Hanze University of Applied Sciences, The Netherlands, and the Laboratory of Telecommunications (LabTel) at the Federal University of Espírito Santo, Brazil.

Research funding

This research was partially supported by Fundação de Amparo à Pesquisa e Inovação do Espírito Santo, Brazil (Grant Nos. 891/2023-P:2023-BDKK7 and 1194/2024-P:2024-26C0T and 2022-BWBR2) and Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (Grant Nos. 311633/2025-0 and 301349/2025-8).

Author contributions statement

H.R. designed the study, developed the methodology, wrote the software, and provided resources, while J.R., S.D., and J.S. contributed to the study design and investigations. J.R., S.D., J.S and H.W. validated the results. S.D., R.R., and J.R. conducted the formal analysis. H.R. and S.D. wrote the original draft. All authors reviewed and edited the manuscript and agreed to the published version of the manuscript.