Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27721-cd

Exploring Data-Driven Solutions to Predict Hydrogen Solubility in Saline Environments for Underground Hydrogen Storage

Sandrely Pereira da Silva

Department of Production Engineering, Federal University of Pernambuco (UFPE), Brazil. E-mail: sandrely.pereira@ufpe.br

Caio Bezerra Souto Maior

Center of Informatics, UFPE, Brazil. E-mail: caio.maior@ufpe.br

Isis Didier Lins

Department of Production Engineering, UFPE, Brazil. E-mail: isis.lins@ufpe.br

The global transition toward low-carbon energy systems has intensified interest in hydrogen as a sustainable energy carrier due to its high gravimetric energy density and compatibility with renewable energy sources. Underground hydrogen storage (UHS) in geological formations such as aquifers, depleted reservoirs, and salt caverns has emerged as a promising solution for large-scale and seasonal energy storage. However, accurately predicting hydrogen solubility in saline aqueous environments remains a key challenge, as it directly influences gas loss, chemical stability, and operational safety. In this study, a machine learning (ML) approach was applied to model hydrogen solubility in saline solutions, addressing the limitations of conventional thermodynamic approaches in capturing complex, non-linear interactions. A dataset comprising 255 experimental observations, including pressure, temperature, salinity, and hydrogen solubility, was analyzed through exploratory data analysis, physically motivated feature engineering, and permutation-based feature selection. Among the evaluated models, CatBoost demonstrated superior predictive performance and robustness. Using the complete dataset, the optimized CatBoost model achieved a coefficient of determination of $R^2 = 0.9973$, a root mean squared error (RMSE) of 0.0125, and a mean absolute error (MAE) of 0.0050, indicating excellent accuracy and strong generalization capability. These results highlight the effectiveness of gradient boosting–based ML methods for modeling hydrogen solubility under saline conditions and demonstrate their potential to support feasibility assessments and risk evaluation in underground hydrogen storage applications.

Keywords: Hydrogen, Underground Hydrogen Storage, Machine Learning, Data-Driven Modeling, Energy, Storage.

1. Introduction

The global transition toward low-carbon energy systems has intensified the search for sustainable energy carriers capable of supporting deep decarbonization (Hassan et al., 2024). However, global energy consumption remains predominantly fossil fuel-based, representing a major challenge for the energy transition (Maior et al., 2022). Hydrogen offers significant advantages in this context due to its high gravimetric energy density and its ability to generate energy without greenhouse gas emissions, producing only water vapor as a byproduct (Maior et al., 2025). When produced from low-carbon pathways, hydrogen can function as an effective medium for energy transfer and storage, particularly by converting surplus renewable electricity from solar and wind sources into chemical energy through water electrolysis (Osman et al., 2022). In addition, hydrogen can be stored and transported in multiple forms, including compressed gas, liquid hydrogen, and chemical carriers such as ammonia or metal hydrides (Aziz, 2021). However, the large-scale implementation of hydrogen-based systems depends on the availability of safe, efficient, and economically viable storage solutions (Ratnakar et al., 2021).

Among the various storage options, underground hydrogen storage (UHS) in geological formations, including depleted oil and gas reservoirs, deep saline aquifers, and salt caverns, has emerged as a promising option for storing large hydrogen volumes and addressing seasonal imbalances between energy supply and demand (Ramesh Kumar et al., 2023). These subsurface systems provide substantial storage capacity and utilize existing geological knowledge and infrastructure. Nevertheless, hydrogen–brine interactions introduce complex physicochemical processes that can affect storage efficiency, operational safety, and long-term reservoir performance (Gbadamosi et al., 2023).

One of the main challenges in UHS is the reliable prediction of hydrogen solubility in saline aqueous environments. Hydrogen dissolution in formation brines contributes to gas loss, alters pressure behavior, and may trigger geochemical or microbial processes that compromise storage integrity (Dehghani et al., 2024). Hydrogen solubility is governed by multiple interdependent variables, including pressure, temperature, and salinity, whose combined effects often exhibit nonlinear behavior (Mwakipunda et al., 2024). Conventional thermodynamic and empirical models, while useful under limited conditions, frequently struggle to capture the full complexity of these interactions, particularly across wide ranges of operating conditions relevant to underground storage applications (Keith et al., 2021).

In recent years, machine learning (ML) techniques have demonstrated significant potential in modeling complex, nonlinear systems in energy and geoscience applications (Nachtane et al., 2023). By learning directly from experimental data, ML-based models can uncover hidden patterns and relationships that are difficult to represent using traditional physics-based approaches alone (Maior & Silva, 2024). When appropriately designed and validated, these data-driven methods offer a powerful tool for improving predictive accuracy and reducing uncertainty in subsurface modeling tasks (Zhou et al., 2022).

Within this context, the present study investigates the application of ML techniques to predict hydrogen solubility in saline aqueous solutions relevant to underground storage scenarios. The research follows a systematic workflow composed of distinct stages. Section 2 details the methodology, including an Exploratory Data Analysis (EDA) to uncover patterns and distributions within the dataset. Subsequently, a preprocessing workflow was implemented, comprising feature engineering, standardization, and variable selection. After preprocessing, the predictive models were trained and evaluated. In Section 3, the results are presented and analyzed, emphasizing the performance of each model and their relevance to UHS applications. Finally, Section 4 provides the study's conclusions and overarching insights.

2. Methodology

The research methodology integrates data preprocessing, feature engineering, and supervised regression modeling. As illustrated in Fig. 1, the workflow comprises four main stages: (1) exploratory data analysis, (2) data preprocessing and feature engineering, (3) model training and validation, and (4) performance evaluation.

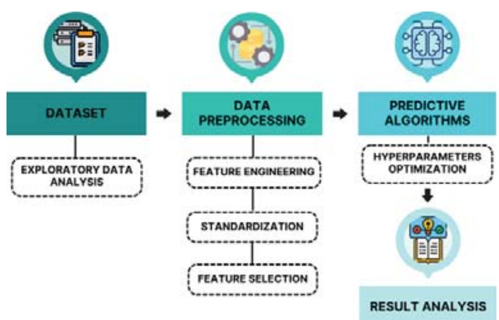


Fig. 1. Methodology of the current study.

2.1. *Data Exploratory Analysis*

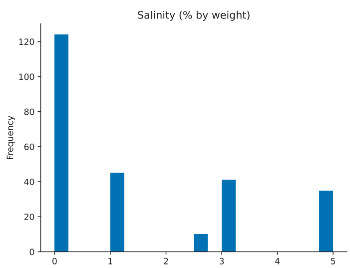
An exploratory data analysis was conducted to evaluate the distribution and key characteristics of the dataset. Descriptive statistics indicate substantial variability in pressure and temperature, a predominance of low salinity values, and a right-skewed distribution of hydrogen solubility. The dataset consists of 255 observations with no missing values. The data was extracted by the study of Vo Thanh et al. (2024). A summary of the main statistical parameters is presented in Table 1, while noninformative variables were removed prior to model development.

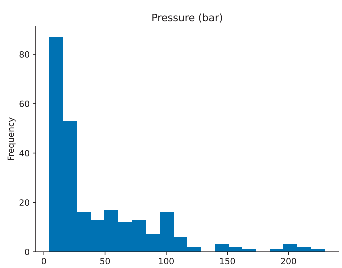
Table 1. Descriptive statistics.

Variable	Mean	Std	Min	Max
Pressure (bar)	44.57	45.22	4.60	229.72
Temperature (K)	336.95	42.26	273.15	423.15
Salinity (% by weight)	1.44	1.80	0.00	5.00
Hydrogen solubility (mole fraction)	0.19	0.24	0.00	0.98

Elaborated by the Authors (2026).

The histograms indicate that hydrogen solubility is predominantly concentrated at low values, exhibiting a pronounced right-skewed distribution. Similar skewness is observed for pressure and salinity, whereas temperature displays a more uniform distribution, as observed in Fig. 2.





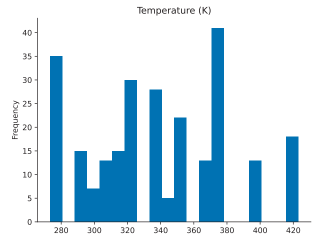
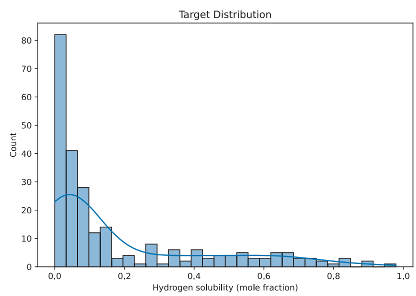
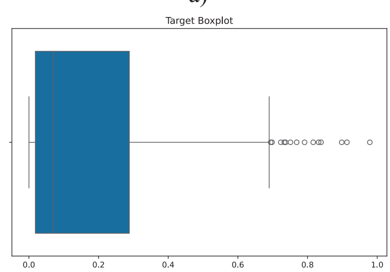


Fig. 2. Histograms of independent variables.

The boxplot analysis (Fig. 3b) further confirms the asymmetric behavior of hydrogen solubility, highlighting a limited number of high-value outliers above approximately 0.5 mole fraction, with the interquartile range primarily concentrated below 0.3.



a)



b)

Fig. 3. Analysis of the target variable.

2.2. *Data Preprocessing*

Data preprocessing was conducted to ensure data consistency and suitability for machine learning analysis.

2.2.1. *Feature Engineering*

Feature engineering was then applied to derive physically meaningful attributes from the original variables, enabling the models to capture non-linear effects and interaction mechanisms. The engineered features include polynomial terms, interaction variables, logarithmic transformations, and a pressure–temperature ratio to reflect thermodynamic behavior and improve predictive performance. A summary of the engineered features and their physical or statistical motivations is provided in Table 2.

Table 2. Description of the new variables included in the dataset.

Feature	Formula	Purpose	Physical / Statistical Justification
P_T_ratio	P / T	Ratio between pressure and temperature	Reflects thermodynamic equilibrium and hydrogen stability (Bachand et al., 2024)
logP	log(1 + P)	Reduce pressure scale	Mitigates outlier influence and stabilizes variance (Solatpour et al., 2024)
P^2	P^2	Capture non-linear pressure effects	Represents non-linear pressure–property relationship(Tiwari et al., 2025)
S_P	S × P	Pressure–salinity interaction	Captures coupled reservoir effects (Medina et al., 2024)
T_S	T × S	Temperature–salinity interaction	Represents geological interaction effects (Janjua et al., 2024)
T^2	T^2	Capture non-linear temperature effects	Models non-linear thermal behavior (Tiwari et al., 2025)
logT	log(1 + T)	Smooth temperature scale	Reduces thermal variability for improved modeling (Solatpour et al., 2024)
S^2	S^2	Capture non-linear salinity effects	Models non-linear salinity influence on hydrogen behavior (Feng et al., 2024)

Elaborated by the Authors (2026).

2.2.2. *Data Splitting*

The dataset was partitioned into an 80:20 ratio, with 0.8 for training and 0.2 for testing, to analyse model performance.

2.2.3. *Standardization*

Feature scaling was performed to avoid the dominance of variables with disparate magnitudes, which is especially critical for scale-sensitive algorithms such as Support Vector Machines. Therefore, all input features were standardized to have zero mean and unit variance using the *StandardScaler* technique, following the methodology proposed by (Ahmed et al., 2022).

2.2.4. *Feature Selection*

Feature selection was conducted using the permutation importance technique, which evaluates the contribution of each input variable by measuring the decrease in model performance when its values are randomly permuted. This approach captures both linear and non-linear effects and is well suited for ensemble-based models (Mehdiyev et al., 2025). An Extra Trees regressor was employed as the base estimator to compute feature importance scores, using the coefficient of determination (R^2) as the evaluation metric. Fig. 4 indicates the feature importance based on permutation method.

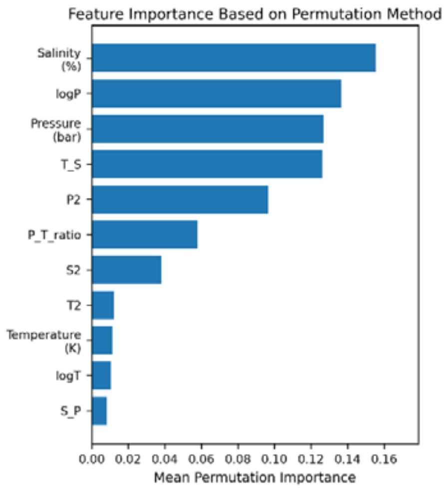


Fig. 4. Features ordered by the importance value.

Initially, models were evaluated using only the features with the highest permutation importance scores. However, improved predictive performance was consistently observed when all features exhibiting positive importance were retained. As all variables in the dataset presented positive permutation importance values, even those with relatively low contributions were preserved in the final feature set.

3. Training

To predict hydrogen solubility, a diverse set of supervised learning algorithms was implemented, encompassing ensemble methods, gradient boosting approaches, and kernel-based regression, allowing the capture of both linear and non-linear relationships among the selected features.

3.1. *Models*

The ML models utilized in the study are briefly described below:

- Random Forest: An ensemble of decision trees that aggregates predictions to reduce variance, improving robustness and generalization (Husain et al., 2023).
- Extra Trees: Similar to Random Forest but introduces more randomness in tree construction, which can further reduce overfitting (Joshi et al., 2021).
- Gradient Boosting: Sequentially builds weak learners, each correcting errors of the previous iteration, enhancing predictive accuracy (A. Deif et al., 2021).
- AdaBoost: Focuses on reweighting mispredicted observations in each iteration, improving the model's ability to capture difficult patterns (Hatwell et al., 2020).
- XGBoost: An optimized gradient boosting algorithm with parallel processing and regularization, providing high efficiency and accuracy (Zhang et al., 2022).
- CatBoost: A gradient boosting specifically designed for categorical data, reducing overfitting and accelerating convergence (Zhang & Jánošík, 2024).
- Support Vector Regressor (SVR): Fits data within a defined margin of tolerance, capturing complex relationships in high-dimensional feature spaces (Ghaddar & Naoum-Sawaya, 2018).

3.2. *Hyperparameter Optimization*

Hyperparameter tuning is essential for enhancing the predictive performance of machine learning models by identifying the most effective configuration of parameters (Aftab et al., 2025). In this study, *GridSearchCV* combined with repeated cross-validation was employed to systematically explore predefined search spaces for each model and select the optimal hyperparameters based on R^2 performance. This approach ensures that models are both accurate and robust while minimizing overfitting. The hyperparameters selected is presented in Table 3.

Table 3. Hyperparameters selected using *GridSearch.*

Model	Best Hyperparameters
Random Forest	n_estimators=300, max_depth=20, min_samples_split=2, min_samples_leaf=1, max_features='log2'
Extra Trees	n_estimators=100, max_depth=10, min_samples_split=2, min_samples_leaf=1, max_features='log2'
Gradient Boosting	n_estimators=100, learning_rate=0.2, max_depth=3, min_samples_split=2, min_samples_leaf=1, subsample=1.0

AdaBoost	n_estimators=50, learning_rate=1.0, loss='square'
XGBoost	n_estimators=500, learning_rate=0.1, max_depth=3, subsample=1.0, colsample_bytree=1.0
CatBoost	iterations=500, depth=4, learning_rate=0.2
SVR	C=10.0, epsilon=0.01, kernel='rbf', gamma='scale'

Elaborated by the Authors (2026).

4. Results

The predictive performance of the implemented ML models was evaluated using R^2, RMSE, and MAE metrics, providing complementary insights into both the accuracy and error magnitude of the predictions. Among the models tested, CatBoost achieved the highest overall performance, with a Test R^2 of 0.967, RMSE of 0.0270, and MAE of 0.0147, indicating its strong ability to capture the complex, non-linear relationships between pressure, temperature, salinity, and hydrogen solubility. Gradient Boosting also performed well, with Test $R^2 = 0.9533$, RMSE = 0.0321, and MAE = 0.0156, slightly lower than CatBoost but still demonstrating robust predictive capabilities.

Support Vector Regressor (SVR) exhibited a lower training R^2 (0.8799) compared to tree-based ensemble methods, suggesting more conservative fitting, yet its Test R^2 of 0.9431 indicates good generalization and low overfitting. XGBoost, Random Forest, and Extra Trees achieved moderate-to-high performance, with Test R^2 values ranging from 0.9069 to 0.9269, reflecting their ability to model non-linear interactions but slightly underperforming compared to CatBoost and Gradient Boosting. AdaBoost showed the lowest predictive accuracy among the models, with Test $R^2 = 0.8485$ and higher RMSE and MAE values, suggesting it may be less suited for capturing the complex relationships present in the dataset.

Overall, ensemble-based gradient boosting models (CatBoost, Gradient Boosting, XGBoost) outperformed simpler or less regularized algorithms due to their sequential learning approach and ability to correct residual errors iteratively. The results also highlight the importance of hyperparameter optimization, as models tuned via GridSearchCV consistently achieved higher R^2 and lower RMSE/MAE values than typical default configurations. The models with higher performances are presented in Table 4.

Table 4. Best results considering all data.

Model	**All R^2**	**All RMSE**
CatBoost	0.9973	0.0125
GradientBoosting	0.9951	0.0169
XGBoost	0.9942	0.0183

Elaborated by the Authors (2026).

5. Conclusion

This study evaluated the application of ML models to predict hydrogen solubility in saline aqueous solutions, with the aim of supporting underground hydrogen storage strategies. Among the models tested, CatBoost demonstrated the highest overall performance, achieving an All R^2 of 0.9973, an All RMSE of 0.0125, and an All MAE of 0.0050, indicating excellent predictive accuracy and generalization. Gradient Boosting and XGBoost also showed strong performance, though slightly lower than CatBoost, reinforcing the effectiveness of gradient boosting methods for capturing complex, non-linear relationships in the dataset. Support Vector Regressor performed well in generalization, while simpler ensemble models such as Random Forest and Extra Trees exhibited moderate performance. AdaBoost presented the lowest predictive accuracy, suggesting limited suitability for this problem.

Despite these promising results, some limitations must be noted. The dataset was relatively small, which may constrain model generalization, particularly in regions of extreme hydrogen solubility or uncommon salinity and temperature conditions. The study also did not assess the computational cost of each model, which is important for practical implementation in large-scale or real-time systems.

Future work should focus on expanding the dataset to include a wider range of salinity and temperature conditions, particularly in critical solubility regions. Alternative feature selection strategies, including those based on permutation or importance scores, could further enhance model interpretability and performance. Investigating neural networks or hybrid approaches may offer additional gains by capturing complex, non-linear interactions that tree-based models may not fully represent. Finally, evaluating computational efficiency will provide practical guidance for selecting the most suitable models for operational hydrogen storage applications.

Acknowledgement

The authors acknowledge the financial support of the Human Resources Program of the National Agency of Petroleum, Natural Gas and Biofuels – PRH-ANP, specifically PRH 38.1 entitled "Risk Management, Reliability and Sustainability in the Exploration and Production of Oil, Natural Gas and Renewable Energies", funded by ANP and managed by Fapesp (São Paulo Research Foundation), process number 2025/03497-0, the 'Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): 402761/2023-5', the 'Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE): APQ-0504-3.08/25', and 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)' - Finance Code 001 for the financial support through research grants.

References

A. Deif, M., E. Hammam, R., & Solyman, A. A. A. (2021). Gradient Boosting Machine Based on PSO for prediction of Leukemia after a Breast Cancer Diagnosis. *International Journal on Advanced Science, Engineering and Information Technology*, *11*(2), 508–515. https://doi.org/10.18517/ijaseit.11.2.12955

Aftab, M., Ahmad, T., Adeel, S., Bhatti, S. H., & Irfan, M. (2025). Hyper-parameter tuning through innovative designing to avoid over-fitting in machine learning modelling: a case study of small data sets. *Journal of Statistical Computation and Simulation*, *95*(7), 1595–1609. https://doi.org/10.1080/00949655.2025.2465787

Ahmed, H. A., Muhammad Ali, P. J., Faeq, A. K., & Abdullah, S. M. (2022). An Investigation on Disparity Responds of Machine Learning Algorithms to Data Normalization Method. *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY*, *10*(2), 29–37. https://doi.org/10.14500/aro.10970

Aziz, M. (2021). Liquid Hydrogen: A Review on Liquefaction, Storage, Transportation, and Safety. *Energies*, *14*(18), 5917. https://doi.org/10.3390/en14185917

Bachand, A., Doyon, B., & Raymond, J. (2024). Thermo-physical numerical model for hydrogen storage in underground tanks and caverns. *International Journal of Hydrogen Energy*, *66*, 66–80. https://doi.org/10.1016/j.ijhydene.2024.03.246

Dehghani, M. R., Nikravesh, H., Aghel, M., Kafi, M., Kazemzadeh, Y., & Ranjbar, A. (2024). Estimation of hydrogen solubility in aqueous solutions using machine learning techniques for hydrogen storage in deep saline aquifers. *Scientific Reports*, *14*(1), 25890. https://doi.org/10.1038/s41598-024-76850-8

Feng, X., Liu, J., Shi, J., Hu, P., Zhang, T., & Sun, S. (2024). Phase equilibrium, thermodynamics, hydrogen-induced effects and the interplay mechanisms in underground hydrogen storage. *Computational Energy Science*, *1*(1), 46–64. https://doi.org/10.46690/compes.2024.01.05

Gbadamosi, A. O., Muhammed, N. S., Patil, S., Al Shehri, D., Haq, B., Epelle, E. I., Mahmoud, M., & Kamal, M. S. (2023). Underground hydrogen storage: A critical assessment of fluid-fluid and fluid-rock interactions. *Journal of Energy Storage*, *72*, 108473. https://doi.org/10.1016/j.est.2023.108473

Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, *265*(3), 993–1004. https://doi.org/10.1016/j.ejor.2017.08.040

Hassan, Q., Viktor, P., J. Al-Musawi, T., Mahmood Ali, B., Algburi, S., Alzoubi, H. M., Khudhair Al-Jiboory, A., Zuhair Sameen, A., Salman, H. M., & Jaszczur, M. (2024). The renewable energy role in the global energy Transformations. *Renewable Energy Focus*, *48*, 100545. https://doi.org/10.1016/j.ref.2024.100545

Hatwell, J., Gaber, M. M., & Atif Azad, R. M. (2020). Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences. *BMC Medical Informatics and Decision Making*, *20*(1), 250. https://doi.org/10.1186/s12911-020-01201-2

Husain, M., Kumar, P., Ahmed, M. N., Ali, A., Rasool, M. A., Hussain, M. R., & Dildar, M. S. (2023). Harnessing Ensemble in Machine Learning for Accurate Early Prediction and Prevention of Heart Disease. *International Journal of Advanced Computer Science and Applications*, *14*(10). https://doi.org/10.14569/IJACSA.2023.0141020

Janjua, A. N., Ali, M., Murtaza, M., Patil, S., & Kamal, M. S. (2024). Effects of salinity, temperature, and pressure on H2–brine interfacial tension: Implications for underground hydrogen storage. *Journal of Energy Storage*, *95*, 112510. https://doi.org/10.1016/j.est.2024.112510

Joshi, R. C., Mishra, R., Gandhi, P., Pathak, V. K., Burget, R., & Dutta, M. K. (2021). Ensemble based machine learning approach for prediction of glioma and multi-grade classification. *Computers in Biology* *and Medicine*, *137*, 104829. https://doi.org/10.1016/j.compbiomed.2021.104829

Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K.-R., & Tkatchenko, A. (2021). Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chemical Reviews*, *121*(16), 9816–9872. https://doi.org/10.1021/acs.chemrev.1c00107

Maior, C. B. S., Macêdo, J. B., Lins, I. D., Moura, M. C., Azevedo, R. V., Santana, J. M. M., da Silva, M. F., & Magalhães, M. V. C. (2022). Bayesian prior distribution based on generic data and experts' opinion: A case study in the O&G industry. *Journal of Petroleum Science and Engineering*, *210*, 109891. https://doi.org/10.1016/j.petrol.2021.109891

Maior, C. B. S., Silva, S. P., Lins, I. D., Moura, M. J., & Droguett, E. L. (2025). Machine learning for hydrogen storage applications: An exploration of the current literature. *International Journal of Hydrogen Energy*, *151*. https://doi.org/10.1016/j.ijhydene.2025.150219

Maior, C. S., & Silva, T. (2024). Time-series failure prediction on small datasets using machine learning. *IEEE Latin America Transactions*, *22*(5), 362–371. https://doi.org/10.1109/TLA.2024.10500720

Medina, O. E., Gallego, J. F., Moncayo-Riascos, I., Lysyy, M., Benjumea, P. N., Cortés, F. B., & Franco, C. A. (2024). Salinity influence on underground hydrogen storage: Insights from molecular dynamics and pore-scale analysis. *International Journal of Hydrogen Energy*, *60*, 959–975. https://doi.org/10.1016/j.ijhydene.2024.02.073

Mehdiyev, N., Majlatow, M., & Fettke, P. (2025). Integrating permutation feature importance with conformal prediction for robust Explainable Artificial Intelligence in predictive process monitoring. *Engineering Applications of Artificial Intelligence*, *149*, 110363. https://doi.org/10.1016/J.ENGAPPAI.2025.110363

Mwakipunda, G. C., Komba, N. A., Kouassi, A. K. F., Ayimadu, E. T., Mgimba, M. M., Ngata, M. R., & Yu, L. (2024). Prediction of hydrogen solubility in aqueous solution using modified mixed effects random forest based on particle swarm optimization for underground hydrogen storage. *International Journal of Hydrogen Energy*, *87*, 373–388. https://doi.org/10.1016/j.ijhydene.2024.09.054

Nachtane, M., Tarfaoui, M., Abichou, M. amine, Vetcher, A., Rouway, M., Aâmir, A., Mouadili, H., Laaouidi, H., & Naanani, H. (2023). An Overview of the Recent Advances in Composite Materials and Artificial Intelligence for Hydrogen Storage Vessels Design. *Journal of Composites Science*, *7*(3), 119. https://doi.org/10.3390/jcs7030119

Osman, A. I., Mehta, N., Elgarahy, A. M., Hefny, M., Al-Hinai, A., Al-Muhtaseb, A. H., & Rooney, D. W. (2022). Hydrogen production, storage, utilisation and environmental impacts: a review. *Environmental Chemistry Letters*, *20*(1), 153–188. https://doi.org/10.1007/s10311-021-01322-8

Ramesh Kumar, K., Honorio, H., Chandra, D., Lesueur, M., & Hajibeygi, H. (2023). Comprehensive review of geomechanics of underground hydrogen storage in depleted reservoirs and salt caverns. *Journal of Energy Storage*, *73*, 108912. https://doi.org/10.1016/j.est.2023.108912

Ratnakar, R. R., Gupta, N., Zhang, K., van Doorne, C., Fesmire, J., Dindoruk, B., & Balakotaiah, V. (2021). Hydrogen supply chain and challenges in large-scale LH2 storage and transportation. Em *International Journal of Hydrogen Energy* (Vol. 46, Número 47, p. 24149–24168). Elsevier Ltd. https://doi.org/10.1016/j.ijhydene.2021.05.025

Solatpour, R., Babak, P., & Kantzas, A. (2024). Log-exponential transformation function for interpreting NMR relaxation measurements of hydrocarbon in organic porous media for enhancing absolute adsorption estimation. *Chemical Engineering Science*, *286*, 119607. https://doi.org/10.1016/j.ces.2023.119607

Tiwari, P., Naskar, S., & Mukhopadhyay, T. (2025). Nonlinear functionally graded metamaterials for hydrogen storage and enhanced sustainability under extreme environments. *Thin-Walled Structures*, *210*, 112901. https://doi.org/10.1016/j.tws.2024.112901

Vo Thanh, H., Zhang, H., Dai, Z., Zhang, T., Tangparitkul, S., & Min, B. (2024). Data-driven machine learning models for the prediction of hydrogen solubility in aqueous systems of varying salinity: Implications for underground hydrogen storage. *International Journal of Hydrogen Energy*, *55*, 1422–1433. https://doi.org/10.1016/j.ijhydene.2023.12.131

Zhang, C., Dong, H., Geng, Y., Liang, H., & Liu, X. (2022). Machine learning based prediction for China's municipal solid waste under the shared socioeconomic pathways. *Journal of Environmental Management*, *312*, 114918. https://doi.org/10.1016/j.jenvman.2022.114918

Zhang, L., & Jánošík, D. (2024). Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. *Expert Systems with Applications*, *241*, 122686. https://doi.org/10.1016/j.eswa.2023.122686

Zhou, Z., Nourani, P., Karimi, M., Kamrani, E., & Anqi, A. E. (2022). Relying on machine learning methods for predicting hydrogen solubility in different alcoholic solvents. *International Journal of Hydrogen Energy*, *47*(9), 5817–5827. https://doi.org/10.1016/j.ijhydene.2021.11.121