International Journal of Hydrogen Energy 91 (2024) 1394–1406

Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

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

Predicting interfacial tension in brine-hydrogen/cushion gas systems under subsurface conditions: Implications for hydrogen geo-storage

Mostafa Hosseini [a], Yuri Leonenko [a,b,*]

[a] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
[b] Department of Geography and Environmental Management, University of Waterloo, ON, N2L 3G1, Canada

ARTICLE INFO

Keywords:
Interfacial tension
Hydrogen storage
Cushion gas
Machine learning
Gas composition
Shapley additive explanations

ABSTRACT

Underground hydrogen storage (UHS) critically relies on cushion gas to maintain pressure balance during injection and withdrawal cycles, prevent excessive water inflow, and expand storage capacity. Interfacial tension (IFT) between brine and hydrogen/cushion gas mixtures is a key factor affecting fluid dynamics in porous media. This study develops four machine learning models— Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-Layer Perceptrons (MLP)—to predict IFT under geo-storage conditions. These models incorporate variables such as pressure, temperature, molality, overall gas density, and gas composition to evaluate the impact of different cushion gases. A group-based data splitting method enhances the realism of our tests by preventing information leakage between training and testing datasets. Shapley Additive Explanations (SHAP) reveal that while the MLP model prioritizes gas composition, the RF model focuses more on operational parameters like pressure and temperature, showing distinct predictive dynamics. The MLP model excels, achieving coefficients of determination (R^2) of 0.96, root mean square error (RMSE) of 2.10 mN/m, and average absolute relative deviation (AARD) of 3.25%. This robustness positions the MLP model as a reliable tool for predicting IFT values between brine and hydrogen/cushion gas (es) mixtures beyond the confines of the studied dataset. The findings of this study present a promising approach to optimizing hydrogen geo-storage through accurate predictions of IFTs, offering significant implications for the advancement of energy storage technologies.

Parameters

AARD	Average Absolute Relative Deviation
ARD	Absolute relative deviation
m	Salt concentration, mol/kg
max_depth	Maximum depth of a tree
max_features	Maximum number of features
Mean	Mean value
n_estimator	Number of trees in an ensemble method
P	Pressure, MPa
R	Pearson correlation coefficient
R^2	Coefficient of determination
Res_i	Residual
RMSE	Root mean squared error
T	Temperature, K
t	Target value
x	Mole fraction
y	Model's output
γ	Interfacial tension, mN/m
ρ	Density, g/cm3

Abbreviations

ANN	Artificial Neural Network
IFT	Interfacial tension
CH_4	Methane
CO_2	Carbon dioxide
CV	Cross-validation
DT	Decision tree
H_2	Hydrogen
ML	Machine learning
MLP	Multilayer perceptron
N_2	Nitrogen
RF	Random forest

* Corresponding author.
E-mail address: leonenko@uwaterloo.ca (Y. Leonenko).

https://doi.org/10.1016/j.ijhydene.2024.10.254
Received 30 August 2024; Received in revised form 16 October 2024; Accepted 18 October 2024
Available online 23 October 2024

1. Introduction

In the past few decades, humans have significantly depleted natural resources in their pursuit of a better quality of life [1]. This is particularly crucial as the use and consumption of fossil fuels have led to the release of substantial amounts of greenhouse gases into the atmosphere, contributing to global warming [2–4]. In response, there has been a growing interest in the potential of renewable resources with low emissions to address the issue of (carbon dioxide) CO_2 emissions [5,6]. Hydrogen (H_2) is a key player in sustainable energy, important for developing a more eco-friendly future. Its production, involving electrical, thermal, hybrid, and biological methods from diverse feedstocks, offers adaptable opportunities for clean energy. Using hydrogen technology effectively, adapted to the available regional resources, can greatly reduce emissions and support sustainable development goals, which aim to ensure everyone has access to sustainable energy and help address climate change [7,8]. However, using H_2 as an energy carrier presents challenges related to storage and transportation due to its low density and high flammability, necessitating careful and efficient handling to ensure safe storage and transport [9].

Numerous H_2 storage techniques have been explored, including compressed gas, liquefaction, and solid-state storage, each presenting unique advantages and disadvantages. Compressed gas is stored in high-pressure tanks, which can hold significant amounts of H_2 but are costly and pose safety risks due to the high pressures involved. Liquefaction involves storing H_2 in liquid form, which offers a high energy density but requires extremely low temperatures, thus adding to operational challenges and costs. Solid-state storage, such as metal hydrides, provides safer options but at lower energy densities and higher costs due to the materials used [9,10]. Most critically, these methods are currently unable to provide the large-scale storage needed to meet long-term energy requirements [11].

To overcome this limitation, Underground Hydrogen Storage (UHS) is considered a potential solution [12]. UHS offers a viable option for medium to long-term storage, depending on specific energy needs [13]. Large volumes of H_2 gas can be effectively stored in depleted hydrocarbon reservoirs [14], empty salt caverns [15], deep aquifers [16], underground coal seams [17], or basaltic formations [18].

The fluid dynamics of hydrogen and other gases are key to determining hydrogen recovery efficiency, influenced by operational factors like temperature, pressure, and reservoir characteristics. Key properties such as interfacial tension (IFT), viscosity, density, solubility, and diffusivity significantly impact fluid movement within storage formations [19]. Brine-H_2 IFT decreases with increasing temperature and pressure, with temperature having a greater effect. The reduction in IFT due to pressure is linked to hydrogen's higher density, while the temperature effect stems from a decreased density difference between water and hydrogen [20]. Lower IFT reduces capillary forces, particularly in small pores, making viscous forces more dominant as the capillary number increases [21]. Higher temperatures also increase hydrogen diffusion rates, which can complicate gas separation during withdrawal, especially in high-temperature reservoirs [22].

CO_2 storage projects primarily aim at the long-term or permanent containment of CO_2 within geological formations [23]. In contrast, UHS projects, although capable of storing large amounts of H_2, are centered around an injection and withdrawal cycle. In these systems, the H_2 injected during the storage phase must be retrieved when needed [24]. While using cushion gas can introduce challenges, such as the need for gas separation [25], it is generally added before H_2 injection to maintain pressure throughout the injection and withdrawal cycles [26], ensure consistent extraction of the working gas [27], and prevent water intrusion during H_2 production [28]. Cushion gas also enhances storage capacity by allowing the gas phase to occupy larger pore spaces alongside water [26].

Cushion gas plays a critical role in both hydrogen injection and withdrawal, directly influencing storage capacity. The required ratio of cushion gas to working gas depends on various geological factors, including reservoir depth, trap shape, and permeability [28]. Cushion gas represents a significant cost in underground hydrogen storage operations, and accurately determining the necessary volume is vital for a reliable estimation of UHS costs [29].

The type of cushion gas greatly affects both recovery efficiency and hydrogen purity. CO_2 is preferred for its favorable physical properties in reservoir conditions [30] and its capacity to reduce greenhouse gas emissions when sequestered [31]. While CO_2 offers the highest storage capacity, lighter gases such as N_2 and CH_4 improve recovery by minimizing gravity override and viscous fingering, with CH_4 achieving up to 80% recovery efficiency [32]. CH_4 also shows high hydrogen recovery rates due to its superior wettability compared to hydrogen, which limits hydrogen penetration into the reservoir's pore spaces and allows for easier separation during production [27].

N_2 and CH_4 are often chosen for their cost-effectiveness [31]. However, CH_4's lower compressibility may lead to increased reservoir pressure, potentially compromising the system's integrity [33]. Additionally, the type of cushion gas influences the interfacial tension (IFT) between gas and brine, affecting fluid flow and geochemical reactions. For instance, CO_2 can dissolve in brine, lowering pH and promoting mineral dissolution, which may alter the reservoir's porosity and permeability [34]. N_2 is also highly effective for hydrogen recovery due to its ability to boost reservoir pressure more than other gases, improving recovery efficiency [35], and its higher wettability simplifies gas separation during production [30].

Regardless of type, cushion gas can increase recovery by up to 7%, improving UHS efficiency [32]. Another critical aspect is the chemical stability of the cushion gas. If it reacts with the working gas or reservoir matrix, it could produce unwanted by-products or damage the reservoir's integrity [32].

Understanding fluid-fluid and fluid-rock interactions is crucial for effectively evaluating subsurface formations for UHS. These interactions play a key role in determining the sealing capabilities of the caprock, the distribution and saturation of H_2 within pore spaces, and the displacement of existing fluids. The presence of cushion gas alongside H_2 in geological formations alters reservoir hydrodynamics due to inevitable gas mixing and molecular diffusion. Consequently, it is essential to understand how cushion gas impacts fluid-fluid interactions with H_2 [36], as this influences phase distribution and gas trapping within the porous medium [37]. Capillary pressure significantly affects pore throat distribution and gas trapping and is influenced by factors such as pore size distribution, wettability, and interfacial tension (IFT) between existing phases [37]. Therefore, the IFT between gas and water affects capillary pressure, which in turn influences pore throat distribution and gas trapping within the porous medium. Thus, the IFT between H_2, or the H_2-cushion gas mixture, and brine could greatly affect the overall effectiveness of the UHS process.

So far, many experimental studies have been conducted to measure the IFT between gas and brine in both H_2 and H_2+cushion gas (es) systems [38,39]. Although these experimental measurements provide accurate data, they require significant time and are not cost-efficient. Chow et al. [40,41] employed the pendant-drop method to experimentally determine the IFT in both water-H_2 and water- H_2–CO_2 systems. Their experiment covered a wide range of pressures from 0.5 to 45 MPa and temperatures between 298.15 and 448.15 K. Additionally, Hosseini et al. [38] measured the IFT of H_2-water/brine systems using the same technique. Their experiments covered pressures ranging from 2.76 MPa to 34.47 MPa, temperatures from 298.15 K to 423.15 K, and brine molalities between 0 and 4.95 mol/kg for a solution consisting of 0.864 NaCl and 0.136 KCl (aq). Furthermore, Mirchi et al. [42] evaluated the IFT of H_2/(methane) CH_4 mixtures with brine at three temperatures (22, 40, and 60) °C and a pressure of 1000 psi, using a brine solution with 2 wt% NaCl salinity. Their findings indicated that the IFT values decreased as the proportion of H_2 in the mixtures decreased, resulting in a higher CH_4 content.

As discussed, understanding H_2 storage in subsurface formations requires accurate measurements of the IFT between fluids, as well as between fluids and rocks, to accurately model fluid behavior within geological formations. While experimental methods have provided valuable insights, numerical approaches have emerged as an efficient and precise way to investigate these systems. It is essential to precisely predict parameters such as IFT, which are crucial for developing comprehensive simulation models that can help optimize UHS strategies and minimize related uncertainties, particularly in the presence of cushion gas. Consequently, the development of computational models—including empirical correlations, molecular dynamics (MD) simulations, and machine learning (ML) models—has become essential.

Many efforts have been made to develop correlations for predicting IFT [43], and these correlations generally achieve a reasonably good level of accuracy. For instance, building on experimental data, Hosseini et al. [38] developed an empirical correlation where the IFT of water + H_2 or H_2 + brine is a function of temperature, pressure, and brine molality. This correlation is applicable to temperatures between 298.15 and 423.15 K, pressures ranging from 2.76 to 34.47 MPa, and brine molalities from 0 to 4.95 mol/kg. Doan et al. [44] employed MD simulations to predict the IFT in different ternary systems consisting of H_2, cushion gases (CO_2, N_2, or CH_4), and water. The research focused on how different conditions—such as varying pressures, temperatures, and proportions of cushion gases—affect the IFT between these phases. Their simulations showed that IFT decreases as pressure and temperature increase, but it rises with a higher proportion of H_2 in the mixture. Similarly, Yang et al. [45] investigated the interfacial properties of hydrogen and water mixtures in contact with silica and kerogen, crucial for UHS. Through MD simulations, they explored how temperature (298–523 K) and pressure (1–160 MPa) affect IFT, wettability, and contact angles. Their results demonstrated that IFT generally decreases with increasing temperature and pressure, although an increase in IFT was observed at higher pressures. Contact angles in both H_2+H_2O + silica and H_2+H_2O + kerogen systems were found to increase with pressure and decrease with temperature, with smaller effects in the silica system at lower temperatures. Additionally, they employed density gradient theory DGT and the *P*C-SAFT equation of state to complement their MD simulations, providing more accurate predictions of the interfacial behaviors. These findings enhance the understanding of gas-fluid interactions in geological formations, contributing to the optimization of hydrogen storage systems.

The development of ML-based models has attracted significant attention from researchers in various fields as they offer cost-effective, easy-to-use, and accurate predictive models. Nguyen et al. [46] discuss the integration of explainable artificial intelligence into renewable energy systems, emphasizing its potential to enhance transparency and effectiveness. They highlight the benefits of explainable artificial intelligence in optimizing energy efficiency and address the ethical challenges it presents. The authors advocate for continued research to develop reliable evaluation methods and standardized datasets to advance explainable artificial intelligence applications in renewable energy, promoting sustainable practices. Furthermore, Le et al. [47] integrated advanced ML algorithms with SHapley Additive exPlanations (SHAP) to improve biochar yield predictions and analysis. They utilized linear regression, AdaBoost, and boosted regression trees to model biochar properties accurately. SHAP analysis enhanced the transparency and interpretability of these models, revealing how factors like temperature and biomass type affect biochar characteristics. This approach provided key insights for optimizing biochar production, important for its environmental applications.

More importantly, ML methods can learn from existing data to identify patterns and make predictions based on input features [48]. Unlike traditional methods, they can integrate all relevant input features for a more thorough analysis [49]. Consequently, these capabilities enhance our understanding of the UHS process. These methods have been widely applied in studies related to CO_2 geo-storage [50,51], demonstrating strong predictive performance. Recently, researchers have focused on employing ML-based models to improve our understanding of H_2 storage [52], specifically in optimizing H_2 geo-storage.

Thanh et al. [53] explored the use of machine learning (ML) models to predict the wettability of rocks/minerals in H_2 storage systems involving brine and H_2. Researchers utilized four ML algorithms—Extreme Gradient Boosting (XGBoost), Random Forest (RF), Light Gradient Boosting (LGRB), and Adaptive Boosting Decision Tree (Adaboost_DT)—to analyze data from 513 experimental samples. XGBoost emerged as the most accurate model, with the highest R^2 value and lowest error rates. The study highlights the importance of substrate types in influencing wettability, as shown by the SHAP analysis. The findings suggest that ML models, particularly XGBoost, offer a promising tool for efficiently predicting wettability and H_2 column height in geological storage, thereby reducing the need for costly and time-consuming experiments. However, the study also notes limitations related to the dependence on existing data and the models' ability to capture complex interactions.

Ng et al. [54] focused on the modeling of IFT in H_2-brine systems using machine learning (ML) techniques. They employed various ML models, including Gradient Boosting Regressor (GBR), Multilayer Perceptron (MLP) optimized with Levenberg-Marquardt (LMA) and Adaptive Moment Estimation (Adam) algorithms, and Genetic Programming (GP). These models were trained on experimental data from previous studies [38,40,41] to accurately predict IFT. Among the models, MLP-LMA showed the best performance, offering high predictive accuracy with an R^2 value of 0.9997. The study identified temperature as the most significant parameter affecting IFT, with the largest impact, followed by molality and pressure. The findings suggest that these ML-based models can be effectively integrated into UHS simulations, potentially reducing reliance on costly and time-consuming experimental procedures. They [54] compared the performance of their ML-based models with the empirical correlation developed by Hosseini et al. [38]. The comparison showed that the MLP-LMA model and the GP-based correlation provided slightly better predictions in terms of R^2, but with significantly improved accuracy when looking at Average Absolute Percentage Relative Error (AAPRE) and RMSE values, especially for the MLP-LMA model.

Behnamnia et al. [55] developed several intelligent models to predict IFT using experimental data for water and various gases, including H_2 + cushion gas, CH_4, CO_2, and N_2. The models considered IFT as a function of temperature, pressure, specific gravity, and brine salinity. Among the models, the Grey Wolf Optimizer-based Least Squares Boosting (GWO-LSBOOST) showed the highest accuracy, with an R^2 of 0.9960 and the lowest average absolute relative error of 0.8060%. Sensitivity analysis revealed that temperature had the greatest impact on IFT, followed by specific gravity and pressure, while salinity had a positive effect.

In our previous study [48], we developed machine learning (ML) methods to predict the IFT between H_2 and brine, a crucial parameter for UHS. Four different models—Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP)—were developed using experimental data on pressure, temperature, and brine molality. Among these, the Random Forest (RF) model appeared as the best performer, with the lowest RMSE (1.50) and superior prediction accuracy ($R^2 = 0.96$). Sensitivity analysis revealed that temperature had the greatest impact on IFT, with molality and pressure being the next most influential factors. The robustness of the models was tested through external validation using data not included in the initial training and testing phases, further confirming the practicality of the models in predicting IFT under different conditions.

This study aims to develop a comprehensive machine learning (ML) model to predict the IFT between brine and gas mixtures. It focuses on systems with varying salinities and gas compositions, particularly those with a non-zero molar fraction of H_2, under different subsurface conditions. Initially, we will compile and refine an extensive dataset from

Table 1
Data summary from 16 sources, detailing gas compositions, pressure and temperature ranges, salinities, and sample counts.

Study	H_2/CO_2/CH_4/N_2	P (MPa)	T (K)	Salinity	n
Muhammed et al. [57]	0.8/0.05/0.1/0.05 0.7/0.05/0.2/0.05 0.6/0.05/0.3/0.05 0.5/0.05/0.4/0.05 0.4/0.05/0.5/0.05 0.3/0.05/0.6/0.05 0.2/0.05/0.7/0.05	3.45–20.68	303.15–343.15	2, 5, 10, 15, and 20 wt% NaCl brines	1050
Muhammed et al. [30]	0.8/0.05/0.05/0.1 0.6/0.05/0.05/0.3 0.4/0.05/0.05/0.5 0.2/0.05/0.05/0.7	3.45–20.68	303.15–343.15	2, 5, 10, 15, and 20 wt% NaCl brines	600
Muhammed et al. [26]	0.8/0.1/0.05/0.05 0.7/0.2/0.05/0.05 0.6/0.3/0.05/0.05 0.5/0.4/0.05/0.05	3.45–20.68	303.15–343.15	2, 5, 10, 15, and 20 wt% NaCl brines	600
Yekta et al. [58]	1.0/0/0/0	5.50–10.00	293.15–318.15	Pure water	2
Chow et al. [40,41]	1.0/0/0/0 0.7/0.3/0/0	0.50–45.20	298.03–448.87	Pure water	79
Higgs et al. [59]	1.0/0/0/0	0.69–20.68	298.00–298.00	Distilled water as well as 1000, 2000, 5000 ppm NaCl brines	36
Hosseini et al. [38]	1.0/0/0/0	2.76–34.47	298.15–423.15	Deionized water as well as 0.864 NaCl + 0.136 KCl brines at total salt molalities of 1.05, 3.15, and 4.95 mol/kg.	64
Esfandyari et al. [60–62]	1.0/0/0/0	1.00–10.00	293.15–353.15	Distilled water and formation brine containing NaCl, KCl, CaCl2, and MgCl2	32
Mirchi et al. [42]	1.0/0/0/0 0.8/0/0.2/0 0.5/0/0.5/0 0.4/0/0.6/0 0.2/0/0.8/0	6.89–6.89	295.15–333.15	20,000 ppm NaCl brine	15
Al-mukainah et al. [63]	1.0/0/0/0	0.07–6.89	323.15–323.15	10 wt% NaCl brine	6
Isfehani [25]	0.3/0.7/0/0 0.5/0.5/0/0 0.7/0.3/0/0	3.45–20.68	323.15–353.15	0/1.05/3.15 mol/kg [0.864 NaCl + 0.136 KCl]	72
Alanazi et al. [39]	1.0/0/0/0 0.5/0/0.5/0	0.34–11.03	323.00–323.00	10 wt% NaCl brine	10
Al-yaseri et al. [64]	1.0/0/0/0	3.45–10.34	348.15–348.15	10 wt% NaCl	3

the existing literature. This dataset will undergo a thorough quality check to ensure its accuracy and reliability.

A key innovation in our approach is the implementation of a rigorous model testing procedure. While past studies employed random data splitting, our study introduces a new strategy for data splitting. By focusing on the variety of gas compositions rather than the quantity of data samples, we aim to make sure that the predictive model is flexible and applicable to gas compositions not seen during the training phase.

2. Methods

The objective of this research is to precisely predict the IFT between brine and hydrogen/cushion gas (es) mixtures under geo-storage conditions using input variables such as pressure, temperature, molality, overall gas density, and gas composition. This study explores several predictive models: Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Multi-layer Perceptrons (MLP). Utilizing the robust Scikit-Learn library in Python [56], each model offers distinct benefits: DT is non-parametric; RF enhances predictive accuracy by averaging multiple DTs and reducing overfitting; SVM is effective in regression tasks by optimally separating data points, but functions as a "black-box"; similarly, MLP, which consists of multiple neural network layers, also acts as a "black-box." Detailed descriptions of these models are not included here but are extensively discussed in our previous work [48], which provides comprehensive insights into their operational mechanisms.

3. Database development

3.1. Literature survey

The compiled database includes 2569 data samples collected from 16 different sources. Table 1 details the data from each source, including gas compositions, pressure and temperature ranges, salinities, and the number of samples. This table demonstrates that IFT data have been collected across a wide range of operational parameters. Additionally, 23 different gas compositions were identified. As previously mentioned, the focus of the study is on H_2 storage, and thus, the H_2 mole fraction is not zero in any of the compositions. Consequently, H_2 is the only pure gas present in the database.

3.2. Data splitting

Proper data splitting is essential for developing an effective ML model, as it significantly impacts the model's accuracy and ability to generalize. A well-thought-out data splitting strategy provides sufficient training data to prevent underfitting while ensuring that the test data can accurately assess the model's generalization capabilities, thereby avoiding overfitting.

We employed a group-based data splitting method as opposed to the conventional sample-based approach. This choice is crucial for ensuring the robustness and validity of the model's predictive performance, especially when dealing with data that exhibits patterns across specific groups, such as different gas compositions or experimental conditions. In traditional sample-based splitting, data is randomly divided into training and testing sets without accounting for the inherent grouping of related samples. This can lead to information leakage, where information from similar samples in the training set affects the test set, unintentionally boosting the model's performance.

Conversely, the group-based splitting approach identifies unique groups within the dataset and assigns all samples from each group solely to either the training or testing set. This strategy ensures that related samples from the same group are not divided between the sets, thus preventing the leakage of trend information. For instance, in our study, where IFT is influenced by factors such as salinity, pressure, and temperature, it is crucial that the model does not encounter parts of a group (e.g., data related to a specific gas composition under certain conditions) in both the training and testing stages. If it were to do so, the testing phase would become less realistic, giving the model an unfair advantage by having previously recognized part of the pattern during training.

The advantage of group-based splitting becomes even more apparent when considering the goal of this study: to predict not just individual IFT values but also the trends and patterns across various gas compositions and operational conditions. By employing group-based splitting, the model is required to identify patterns in the training set that can be generalized to different groups in the test set, better reflecting real-world application scenarios. This approach creates a more realistic operational environment, where the model faces entirely new groups of data, such as unfamiliar combinations of gases and conditions, which it has not seen during training.

Table 2
Allocation of gas compositions to training and testing sets, listing the specific compositions used for model training and testing.

Index	H_2	CO_2	CH_4	N_2	n	Set
1	1	0	0	0	126/68	Train/Extra
2	0.7	0.3	0	0	60	Train
3	0.8	0	0.2	0	3	Train
4	0.5	0	0.5	0	8	Train
5	0.2	0	0.8	0	3	Train
6	0.3	0.7	0	0	24	Train
7	0.5	0.5	0	0	24	Train
8	0.8	0.05	0.1	0.05	150	Train
9	0.6	0.05	0.3	0.05	150	Train
10	0.5	0.05	0.4	0.05	150	Train
11	0.2	0.05	0.7	0.05	150	Train
12	0.8	0.05	0.05	0.1	150	Train
13	0.6	0.05	0.05	0.3	150	Train
14	0.4	0.05	0.05	0.5	150	Train
15	0.2	0.05	0.05	0.7	150	Train
16	0.8	0.1	0.05	0.05	150	Train
17	0.7	0.2	0.05	0.05	150	Train
18	0.6	0.3	0.05	0.05	150	Train
19	0.4	0	0.6	0	3	Test
20	0.7	0.05	0.2	0.05	150	Test
21	0.4	0.05	0.5	0.05	150	Test
22	0.3	0.05	0.6	0.05	150	Test
23	0.5	0.4	0.05	0.05	150	Test

As previously mentioned, 23 different gas compositions were identified in the database. Of these, 5 were randomly allocated to the testing set, and the remaining 18 to the training set, achieving an approximate 20:80 ratio for testing and training. The specific compositions assigned to the training and testing sets are detailed in Table 2.

The data for each composition underwent a quality check. It was found that the data reported by Higgs et al. [59] and Esfandyari et al. [60–62] were inconsistent with the common trends observed in the other samples. Consequently, the data from these authors were assigned to the extra set to prevent the model from being fed with doubtful information and to evaluate how the trained model performs on them. Based on these considerations, 603 data samples were allocated for testing and 68 data samples for extra set, resulting in a sample-wise training/testing/extra ratio of 74:23:3.

3.3. Feature engineering and selection

Typically, in studies available in the literature on IFT modeling of gas and brine [55], researchers have considered parameters such as pressure, temperature, salinity, and gas-specific gravity as inputs, with the IFT between brine and the gas mixture (and/or gas) as the output. It is important to note that including gas-specific gravity may introduce limitations to the models and correlations. This is because it is possible to have two different gas mixtures with the same specific gravity. For example, different combinations of gases like H_2, methane, and carbon dioxide (CO_2) can result in the same specific gravities. Consequently, under the same conditions of temperature and pressure, models or correlations might inaccurately predict identical IFTs for these different mixtures. As highlighted in the study by Muhammed et al. [30], the specific composition of gases within the mixtures significantly influences IFT values. To address this concern, we have included the mole fraction composition of the gas as an input variable to observe the effect of different gas compositions. Furthermore, the overall density (ρ), as introduced below, as another input in our models:

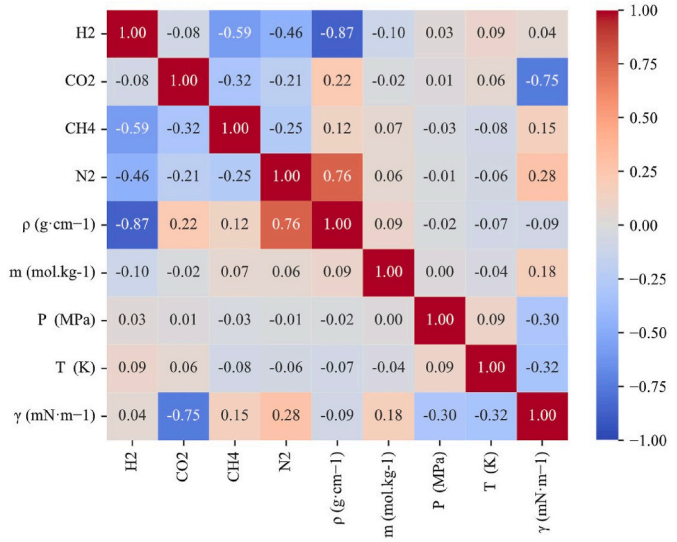


Fig. 1. Correlation matrix of features.

$$\rho = \sum_i x_i \times \rho_i \tag{1}$$

The pairwise Pearson correlation coefficient (R) for the input features is shown in Fig. 1. This coefficient measures the linear correlation between pairs of features, with values ranging from −1 (indicating perfect negative correlation) to 1 (indicating perfect positive correlation). As shown in Fig. 1, none of the features exhibit a high correlation with one another, indicating that the features are independent. This independence suggests that each feature provides unique information to the model, thereby enhancing its ability to learn and make accurate predictions.

3.4. Scaling

ML studies, feature scaling is a crucial preprocessing step, particularly for algorithms that are sensitive to the scale of input data, such as MLP and SVM. Without scaling, features with larger ranges could disproportionately influence the model, leading to skewed results. Standard scaling, which is used in this study, transforms each feature to have a mean of zero and a standard deviation of one. This normalization ensures that all features are on a similar scale, allowing distance-dependent algorithms to function effectively and thereby enhancing the model's accuracy and efficiency.

Standard scaling, the technique selected for this study, solves this problem by transforming the data to ensure each feature possesses a mean (μ) of 0 and a standard deviation (σ) of 1 This process involves subtracting the mean from each feature and then dividing by the standard deviation:

$$z = \frac{(x - \mu)}{\sigma} \tag{2}$$

Where z is the standardized value, x stands for the original value, μ denotes the mean of the feature, and σ represents the standard deviation of the feature.

It is essential to perform scaling based on the training data and then apply the same scaling parameters to the testing data. This practice ensures consistent exposure of the model to the data distribution. By computing the scaling parameters (mean and standard deviation) exclusively from the training set, we prevent data leakage from the testing set into the training process. This approach maintains the integrity of the model evaluation, offering a realistic assessment of the model's performance on unseen data. Applying the training-derived scaling parameters to the testing data ensures that both sets are on the same scale, which is crucial for accurate and fair model evaluation.

Table 3
Optimized hyperparameters for each model from a randomized search with 5-fold cross-validated RMSE after 100 trials.

Hyperparameter	Search Interval	DT	RF	SVM	MLP
max_depth	3–50	27	46	–	–
max_features	0.01–1	0.95	0.66	–	–
n_estimators	4–300	–	67	–	–
C	0.1–1000	–	–	65.71	–
Gamma	0.1–1000	–	–	0.13	–
hidden_layer_sizes	3–30	–	–	–	25
Alpha	1e-05—1.0	–	–	–	0.009
learning_rate_init	1e-05—0.1	–	–	–	0.004
Best CV RMSE	–	0.203	0.158	0.147	0.196
Elapsed time (s)	–	8.37	80.69	65.55	77.34

3.5. SHapley additive exPlanations (SHAP) method

The SHapley Additive exPlanations (SHAP) method is used to measure feature importance and examine how input features influence the system's output. Unlike conventional approaches, SHAP values are based on input features and the outputs from the ML model, not on the model's internal structure. This is especially beneficial in ensemble models, which lack specific prediction equations. Originating from coalitional game theory, SHAP offers a structured way to interpret ML model predictions, making it versatile for use with various ML models [65,66].

4. Result and discussion

4.1. Model development

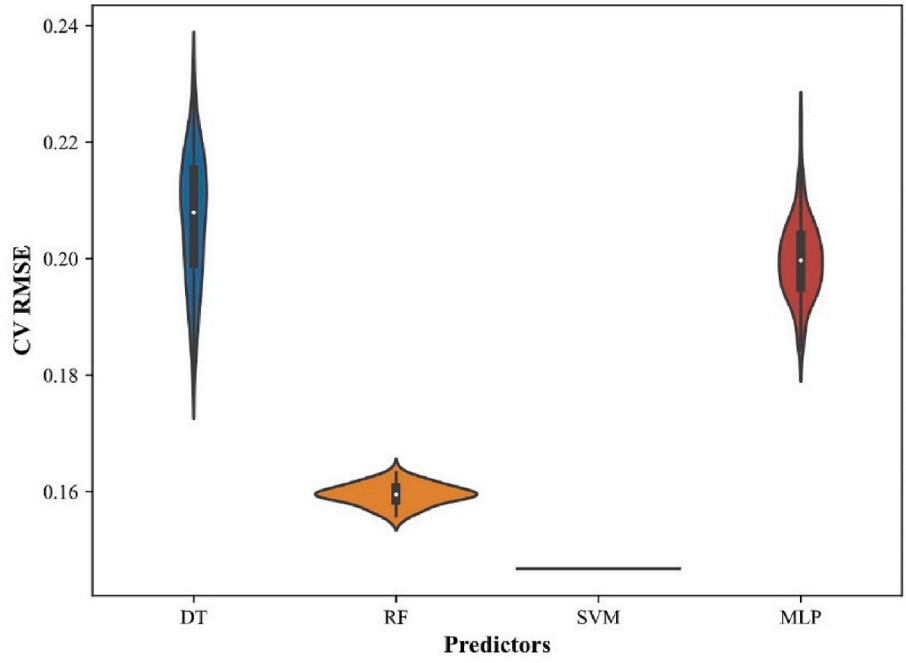
In this section, we explore the creation of ML models using the Scikit-Learn library [1], focusing on four different predictive models: DT, RF, SVM, and MLP.

The initial step is configuring the hyperparameters. Hyperparameters are crucial settings that control the learning process of each model. By carefully adjusting these hyperparameters through techniques such as randomized search, we aim to precisely tailor each model to the unique characteristics and complexities of our dataset. This fine-tuning process is essential for optimizing the models' predictive abilities, ensuring they can accurately capture and generalize the underlying patterns within the data. Table 3 presents the optimized hyperparameter values for each model, determined through a randomized search of 100 trials. A 5-fold cross-validated RMSE was used as the cost function, and the search space for each hyperparameter is also provided. Among these, SVM demonstrated the best performance.

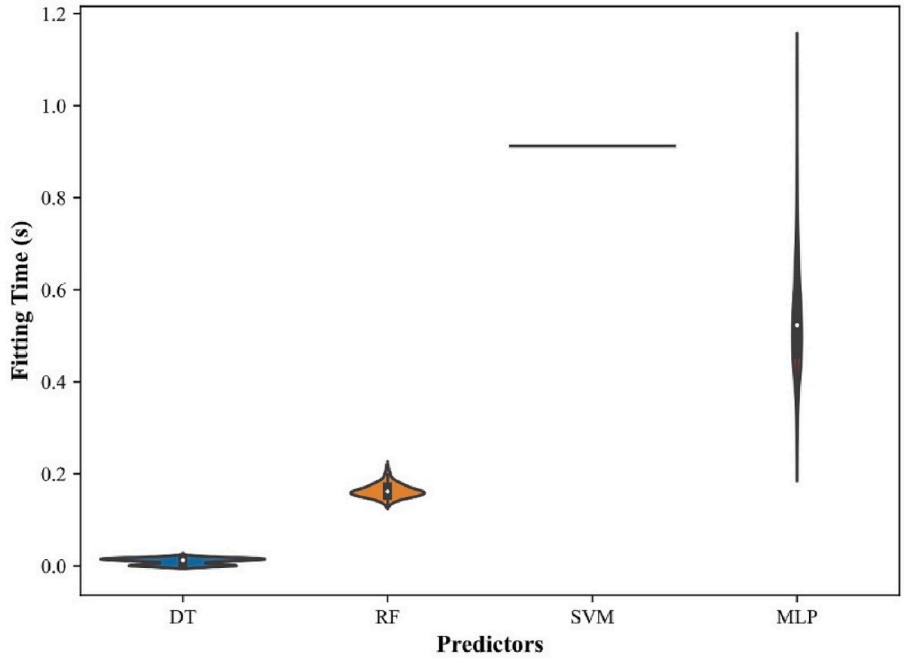
After determining the optimal structure for each model, we trained the models 100 times using different random states. The 5-fold cross-validated RMSE was utilized to evaluate the performance of the trained models and identify the best representative for each type. Fig. 2 presents violin plots illustrating (a) the 5-fold cross-validated RMSE and (b) the processing time. The height of the violin plot indicates the model's stability. As shown, the RF model has the smallest height, indicating consistent results across different random initializations. For the SVM, the violin plot reduces to a line, demonstrating that SVM is deterministic and yields consistent results every time.

4.2. Accuracy assessment

Now that the optimized structures of each model have been trained and the best predictive model based on cross-validation has been selected, it is time to apply these models to the testing data to assess their actual performance on unseen compositions. Various statistical param-



(a)



(b)

Fig. 2. Violin plots depicting (a) the 5-fold cross-validated RMSE and (b) the processing time for each model.

eters are used to measure prediction accuracy. In this study, we employ Residual (Resi), Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Average Relative Deviation (ARD), and Average Absolute Relative Deviation (AARD) to evaluate prediction performance. The formulas for these parameters are provided below:

$$Res_i = t_i - y_i \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Res_i)^2}{\sum_{i=1}^{n} (\bar{t} - t_i)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Res_i)^2} \tag{5}$$

$$ARD = \frac{100}{n} \times \sum_{i=1}^{n} \frac{Res_i}{t_i} \tag{6}$$

$$AARD = \frac{100}{n} \times \sum_{i=1}^{n} \frac{|Res_i|}{t_i} \tag{7}$$

Where y and t represent the model's output and target value,

Table 4
Statistical metrics used to evaluate the accuracy of the developed models. Metrics are calculated based on original values, with the best results for the testing data highlighted in bold.

Model	Set	AARD (%)	ARD (%)	RMSE	R^2	N
DT	Training	0.00	0.00	0.00	1.00	1898
	Testing	4.29	−1.27	3.64	0.89	603
	Extra	9.90	3.04	7.42	−1.45	68
RF	Training	0.49	0.03	0.56	1.00	1898
	Testing	4.05	−0.74	3.06	0.93	603
	Extra	**6.55**	4.27	**5.09**	**−0.15**	68
SVM	Training	1.18	0.08	1.14	0.99	1898
	Testing	13.89	−3.69	8.45	0.43	603
	Extra	8.42	7.37	6.80	−1.06	68
MLP	Training	1.86	−0.71	1.61	0.97	1898
	Testing	**3.25**	**−0.24**	**2.10**	**0.96**	603
	Extra	7.12	**3.83**	5.48	−0.34	68

respectively.

Table 4 presents an analysis of the statistical metrics used to assess the accuracy of the models developed in this study. These metrics are calculated based on the original values, not the scaled data used during the training phase, with the best values for the testing data highlighted in bold for clarity. The evaluation reveals that the MLP model excels with the lowest RMSE of 2.10 mN/m. The RF model also demonstrates a high level of accuracy, comparable to that of the MLP. These models are effective in handling the complex interactions of the features and are less prone to the influence of outliers in the data.

In contrast, the DT model shows significant overfitting. While it achieves perfect predictions on the training data, it fails to perform similarly on the testing data. This overfitting suggests that the DT model might be too sensitive to the noise in the training dataset, making it less effective on new, unseen data. The SVM model, despite having the best results based on the cross-validated RMSE, the SVM performs poorly on unseen testing data, with an RMSE of 6.80 mN/m. This discrepancy arises because the cross-validation randomly split the data into 5 folds, leading to information leakage from the validation data into the training

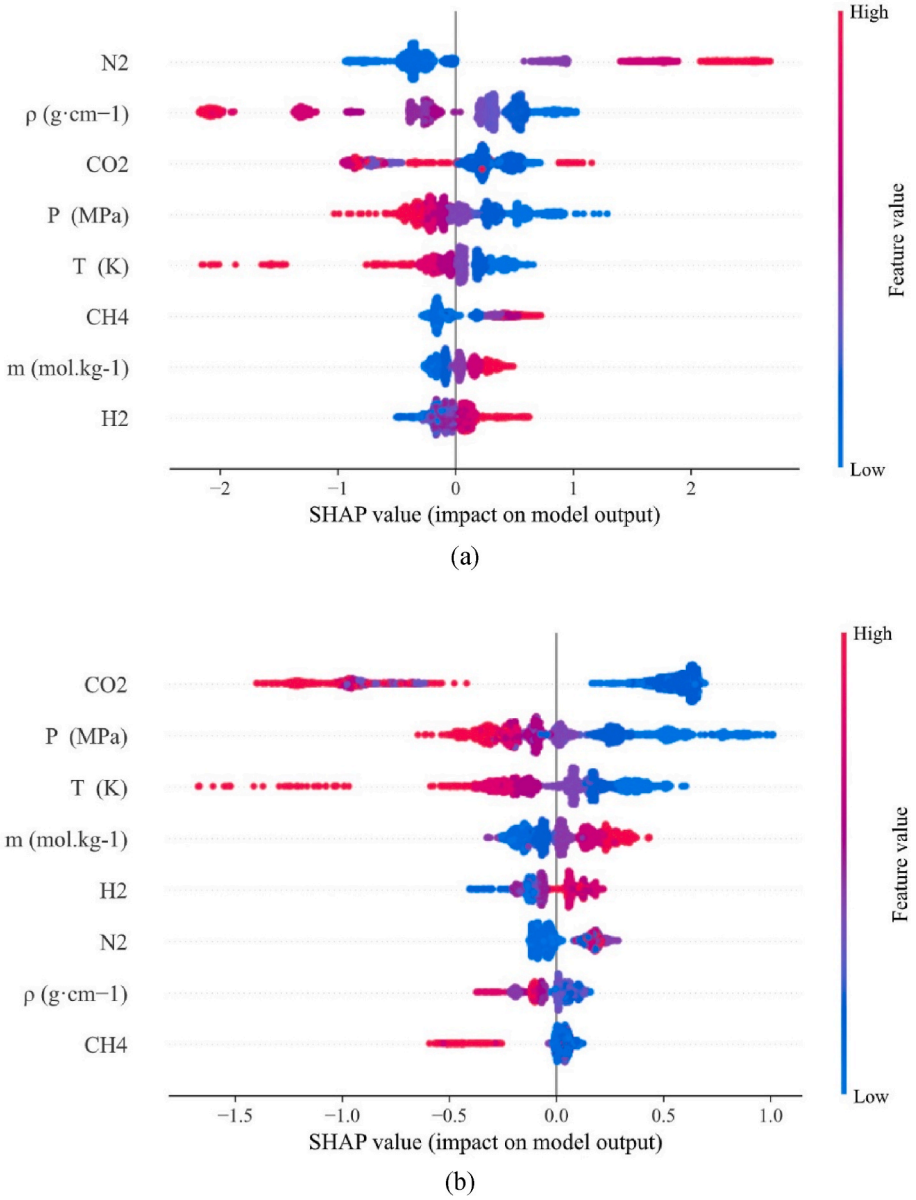


Fig. 3. SHAP beeswarm plots for (a) MLP and (b) RF models.

Table 5
Summary of the effects of key parameters on the predicted IFT, based on SHAP analysis for the MLP model.

Parameter	Impact on IFT
N_2	Higher values of N_2 (shown in red) increase IFT, as indicated by the positive SHAP values. Conversely, lower values of N_2 (blue) slightly decrease the IFT.
ρ (g/cm^3)	Higher ρ values (red) decrease IFT, as shown by the negative SHAP values. Conversely, lower ρ values (blue) have the opposite effect, slightly increasing IFT.
CO_2	Higher CO_2 values (shown in red) strongly decrease IFT, as indicated by negative SHAP values. Lower CO_2 values (shown in blue) are associated with a slight increase in IFT.
P (MPa)	Higher P values (shown in red) consistently decrease IFT, as indicated by strong negative SHAP values. Lower P values (shown in blue) slightly increase IFT.
T (K)	Higher T values (shown in red) decrease IFT, similar to P, with strong negative SHAP values. Conversely, lower T values (shown in blue) have a weaker positive effect on increasing IFT.
CH_4	Higher CH_4 values (shown in red) slightly decrease IFT, while lower CH_4 values (shown in blue) slightly increase IFT.
m (mol.kg-1)	Higher m values (shown in red) increase IFT, as indicated by positive SHAP values. Conversely, lower m values (shown in blue) slightly decrease IFT.
H_2	Higher H_2 values (shown in red) increase IFT, as indicated by positive SHAP values. Conversely, lower H_2 values (shown in blue) slightly decrease IFT.

data. Consequently, the model's cross-validation accuracy was falsely exaggerated, failing to generalize to new compositions. This emphasizes the importance and superiority of the group-based data splitting approach used in this study, which prevents information leakage and ensures a more realistic evaluation of model performance.

The model metrics for the extra data indicate that none of the trained models are capable of making reasonable predictions. While the predictions for the testing data are acceptable, this finding suggests that the experimental samples in the extra set may not conform to the patterns learned from the training data, indicating potential anomalies or variations within these samples.

While the MLP model exhibits impressive performance with the lowest RMSE of 2.10 mN/m, it is crucial to recognize its inherent black-box nature, which limits interpretability. This lack of transparency may hinder understanding of how specific features influence the model's predictions, particularly in areas like fluid dynamics where insights into the underlying physical mechanisms are critical. However, the use of SHAP helps to overcome this limitation by enabling the interpretation of individual feature contributions to the model's predictions. SHAP values provide a clearer view of how operational parameters affect the predicted IFT, thus improving the MLP model's interpretability and making it more applicable for practical use in this context.

It is valuable to compare the performance of the MLP and RF models in greater detail. Fig. 3 presents the SHAP beeswarm plots for the (a) MLP and (b) RF models. As shown, each model assigns different weights to the features when making predictions, highlighting varying feature importances. For the MLP model, the top three important features are related to the gas type. In contrast, for the RF model, the operational parameters of pressure (P) and temperature (T) are among the top three important features. Despite these differences, both models exhibit similar trends for the operational parameters. Both models indicate that increasing pressure and temperature decreases the IFT, while increasing salinity raises the IFT.

Table 5 shows the impacts of each parameter on IFT. It indicates that density (ρ) and temperature (T) significantly decrease IFT when their values increase. Similarly, pressure (P) also lowers IFT, though its impact is less noticeable than ρ and T. CO_2 moderately reduces IFT at higher concentrations. CO_2 moderately reduces IFT at higher concentrations. Conversely, molality (m) and N_2 positively affect IFT, with increases in their values leading to higher IFT. H_2 generally raises IFT, while CH_4 slightly lowers it.

As mentioned in Section 3.2, 'Data Splitting,' the samples reported by Higgs et al. [59] and Esfandyari et al. [60–62] did not show a consistent trend with the other data. Fig. 4 displays the model predictions for data collected by Higgs et al. [59], which involved 100% H_2 and deionized water at 298 K. The markers represent the experimental data, while the lines depict the model predictions. The left panel shows predictions from the MLP model, and the right panel shows predictions from the RF model. The legend in each row provides information about the gas and brine compositions as well as the temperature. The figure reveals instances where different IFT values are reported under the same conditions, exhibiting a non-uniform trend. Both models failed to provide acceptable predictions, particularly in the case where m = 0.03, as evidenced by very high RMSE values.

Fig. 5 presents the model predictions for data reported by Esfandyari et al. [60–62]. For deionized water (m = 0), an increasing trend in IFT with temperature is reported, whereas for brine, the IFT decreases with T. Additionally, the SHAP values shown in Fig. 3 indicate that IFT decreases with increasing T. For deionized water, both models predict a trend opposite to the reported data but consistent with SHAP recommendations. Even for the brine, where the trend aligns with other data and thus, SHAP recommendations, there is a significant discrepancy between the predicted and experimental values.

The modeling results from this study highlight significant potential for improvement in prediction accuracy. A key factor in achieving better predictions is the availability of new experimental data. With increasing focus on H_2 storage in recent years, the volume of experiments has grown, as evidenced by the data gathered, most of which were reported after 2020. However, ML models require even more data to effectively understand the underlying patterns.

This study also emphasizes the critical importance of conducting data quality checks before inputting data into the model. Incorrect or inconsistent data can severely disrupt the model's learning process and hinder accurate predictions.

Another crucial point highlighted by this study is the necessity of proper data splitting in ML studies. It was demonstrated that information leakage from testing or validation data into the training data can lead to overly optimistic and misleading results. Ensuring a rigorous and appropriate data split is essential for obtaining reliable and valid assessments of model performance.

5. Conclusion

This study developed and evaluated several ML models to predict the IFT between brine and gas mixtures, focusing on systems with varying salinities and gas compositions, particularly those with a non-zero molar fraction of H_2. The models explored include DT, RF, SVM, and MLP.

Key findings of this research are as follows:

- The MLP and RF models demonstrated the best predictive performance, with the MLP achieving the lowest RMSE of 2.10 mN/m. However, the DT model showed significant overfitting, and the SVM model, despite having the lowest cross-validated RMSE, failed to generalize effectively to unseen data due to information leakage during cross-validation.
- The study revealed that different models assign varying levels of importance to features. The MLP model prioritized gas composition, while the RF model focused more on operational parameters such as pressure (P) and temperature (T). Both models consistently indicated that increasing P and T decreases IFT, whereas increasing salinity leads to an increase in IFT.
- The research emphasized the necessity of rigorous data quality checks. Inconsistent data significantly impacted model performance, highlighting the importance of verifying data integrity before model training.
- The study highlighted the importance of proper data splitting to prevent information leakage. Group-based data splitting, as utilized

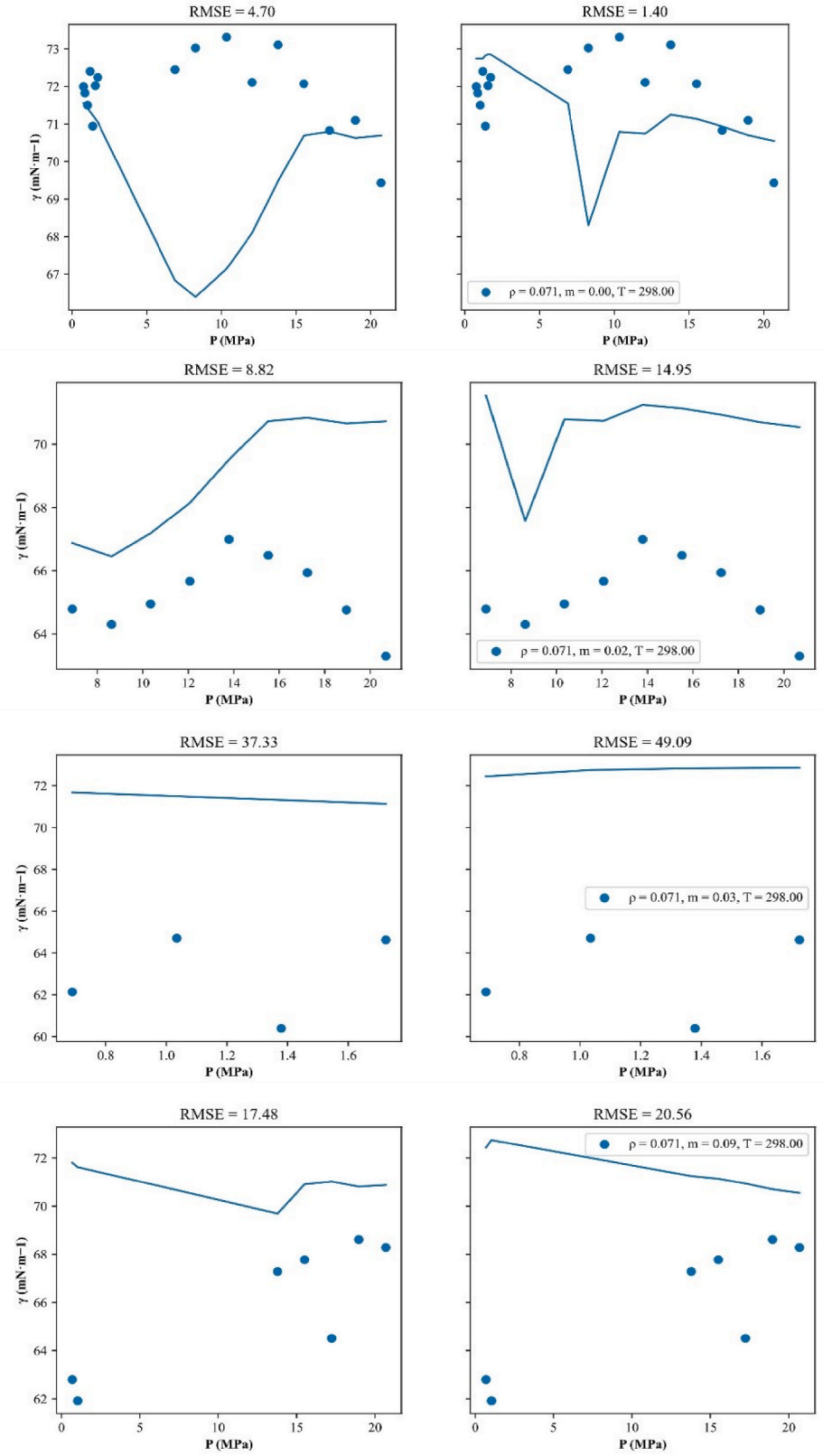


Fig. 4. Comparison of model predictions with experimental data from Higgs et al. [59], for systems containing 100% H_2 and deionized water at 298 K.

in this study, was shown to offer a more realistic evaluation of model performance compared to traditional random splitting methods.

- To scale these models for large-scale UHS, extensive and high-quality data from different geological settings are necessary. Current models rely on datasets from controlled experiments, which may not fully capture the complexity of real-world reservoirs. As more operational data from pilot and commercial UHS projects become available, model performance and generalizability can be improved by updating them with diverse datasets.

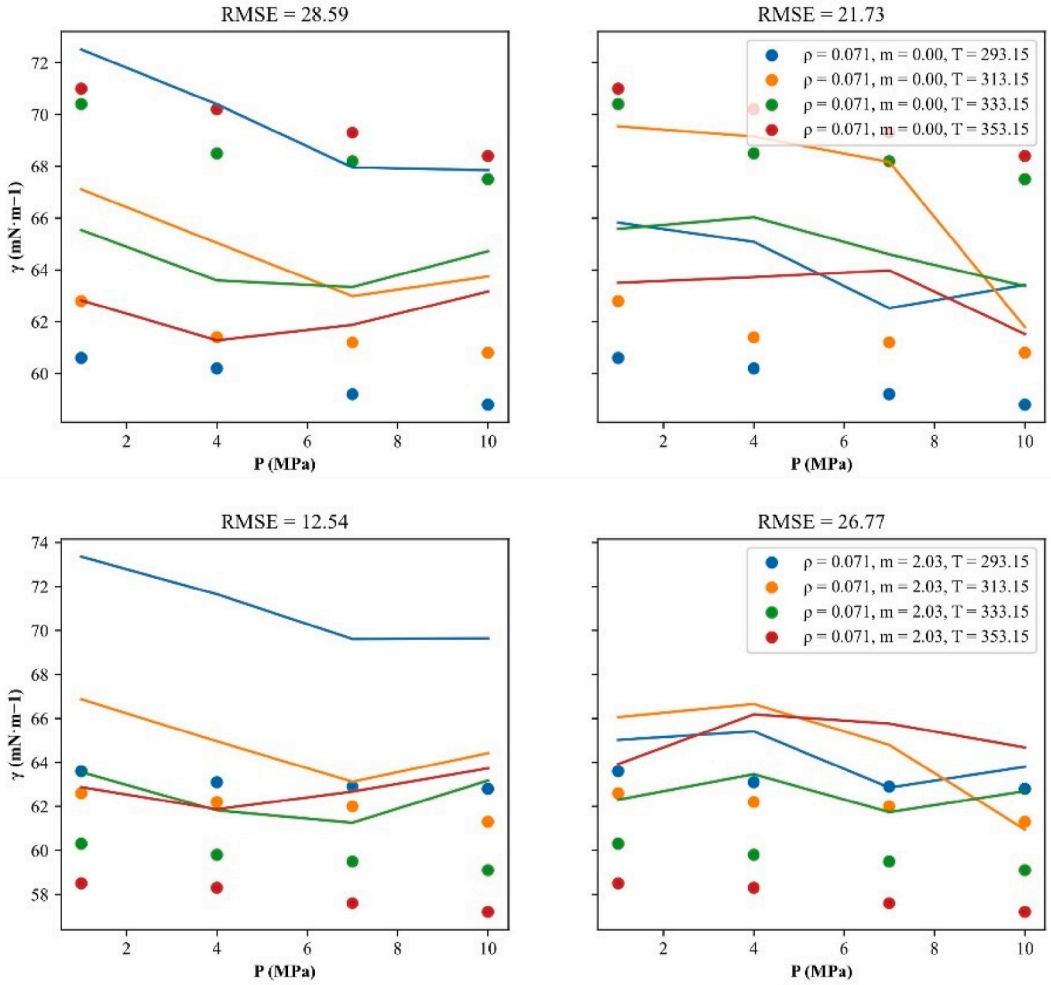


Fig. 5. Comparison of model predictions with experimental data from Esfandyari et al. [60–62], for systems containing 100% H_2 with deionized water and brine at various temperatures.

- These ML models can be integrated into existing reservoir simulation software to provide enhanced predictive capabilities, allowing operators to optimize gas injection and withdrawal cycles. However, the integration must be adapted to site-specific factors, such as reservoir heterogeneity, which may require further model tuning to account for local geological variations.
- To use these models in real-world projects, they need to be able to handle large amounts of data from monitoring systems in underground hydrogen storage. Using cloud computing and advanced machine learning methods can help manage and process this data quickly, making sure the models stay efficient even as the amount of data grows.
- The success of real-world use depends on more than just accuracy; it also requires managing costs, regulations, and long-term gas behavior. Using model predictions can help operators reduce risks like gas leaks and ensure safer hydrogen storage.

CRediT authorship contribution statement

Mostafa Hosseini: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Yuri Leonenko:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

Financial support for this work provided by Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Huo J, Peng C. Depletion of natural resources and environmental quality: prospects of energy use, energy imports, and economic growth hindrances. Res Pol 2023;86:104049.
[2] Kuşkaya S. Residential solar energy consumption and greenhouse gas nexus: evidence from Morlet wavelet transforms. Renew Energy 2022;192:793–804.
[3] Ali M, Yekeen N, Pal N, Keshavarz A, Iglauer S, Hoteit H. Influence of pressure, temperature and organic surface concentration on hydrogen wettability of caprock; implications for hydrogen geo-storage. Energy Rep 2021;7:5988–96.
[4] Dincer I, Aydin MI. New paradigms in sustainable energy systems with hydrogen. Energy Convers Manag 2023;283:116950.
[5] Hassan T, Song H, Khan Y, Kirikkaleli D. Energy efficiency a source of low carbon energy sources? Evidence from 16 high-income OECD economies. Energy 2022;243:123063.
[6] Zhu Y, Wu S, Li J, Jia Q, Zhang T, Zhang X, et al. Towards a carbon-neutral community: integrated renewable energy systems (IRES)–sources, storage, optimization, challenges, strategies and opportunities. J Energy Storage 2024;83:110663.

[7] Rasul MG, Hazrat MA, Sattar MA, Jahirul MI, Shearer MJ. The future of hydrogen: challenges on production, storage and applications. Energy Convers Manag 2022; 272:116326.
[8] Martins FP, Almaraz SD-L, Junior ABB, Azzaro-Pantel C, Parikh P. Hydrogen and the sustainable development goals: synergies and trade-offs. Renew Sustain Energy Rev 2024;204:114796.
[9] Hassan Q, Sameen AZ, Salman HM, Jaszczur M. Large-scale green hydrogen production via alkaline water electrolysis using solar and wind energy. Int J Hydrogen Energy 2023;48:34299–315.
[10] Faye O, Szpunar J, Eduok U. A critical review on the current technologies for the generation, storage, and transportation of hydrogen. Int J Hydrogen Energy 2022; 47:13771–802.
[11] Mehr AS, Phillips AD, Brandon MP, Pryce MT, Carton JG. Recent challenges and development of technical and technoeconomic aspects for hydrogen storage, insights at different scales; A state of art review. Int J Hydrogen Energy 2024;70: 786–815.
[12] Shahriar MF, Khanal A, Khan MI, Pandey R. Current status of underground hydrogen storage: perspective from storage loss, infrastructure, economic aspects, and hydrogen economy targets. J Energy Storage 2024;97:112773.
[13] Thiyagarajan SR, Emadi H, Hussain A, Patange P, Watson M. A comprehensive review of the mechanisms and efficiency of underground hydrogen storage. J Energy Storage 2022;51:104490.
[14] Hemme C, Van Berk W. Hydrogeochemical modeling to identify potential risks of underground hydrogen storage in depleted gas fields. Appl Sci 2018;8:2282.
[15] Caglayan DG, Weber N, Heinrichs HU, Linßen J, Robinius M, Kukla PA, et al. Technical potential of salt caverns for hydrogen storage in Europe. Int J Hydrogen Energy 2020;45:6793–805.
[16] Luboń K, Tarkowski R. Numerical simulation of hydrogen injection and withdrawal to and from a deep aquifer in NW Poland. Int J Hydrogen Energy 2020;45: 2068–83.
[17] Arif M, Abid HR, Keshavarz A, Jones F, Iglauer S. Hydrogen storage potential of coals as a function of pressure, temperature, and rank. J Colloid Interface Sci 2022; 620:86–93.
[18] Al-Yaseri A, Ali M, Ali M, Taheri R, Wolff-Boenisch D. Western Australia basalt-CO2-brine wettability at geo-storage conditions. J Colloid Interface Sci 2021;603: 165–71.
[19] Janjua AN, Ali M, Murtaza M, Patil S, Kamal MS. Effects of salinity, temperature, and pressure on H2–brine interfacial tension: implications for underground hydrogen storage. J Energy Storage 2024;95:112510.
[20] Ali A, Cole DR, Striolo A. Cushion gas effects on clay-hydrogen-brine wettability at conditions relevant to underground gas storage. Int J Hydrogen Energy 2024;58: 668–77.
[21] Medina OE, Gallego JF, Moncayo-Riascos I, Lysyy M, Benjumea PN, Cortés FB, et al. Salinity influence on underground hydrogen storage: insights from molecular dynamics and pore-scale analysis. Int J Hydrogen Energy 2024;60:959–75.
[22] Yang Y, Zandanel A, Liu S, Neil CW, Germann TC, Gross MR. Temperature dependence of hydrogen diffusion in reservoir rocks: implications for hydrogen geologic storage. Energy Adv 2024;3:2051–65.
[23] Eigbe PA, Ajayi OO, Olakoyejo OT, Fadipe OL, Efe S, Adelaja AO. A general review of CO2 sequestration in underground geological formations and assessment of depleted hydrocarbon reservoirs in the Niger Delta. Appl Energy 2023;350: 121723.
[24] Tarkowski R, Uliasz-Misiak B. Towards underground hydrogen storage: a review of barriers. Renew Sustain Energy Rev 2022;162:112451.
[25] Isfehani ZD, Sheidaie A, Hosseini M, Fahimpour J, Iglauer S, Keshavarz A. Interfacial tensions of (brine+ H2+ CO2) systems at gas geo-storage conditions. J Mol Liq 2023;374:121279.
[26] Muhammed NS, Haq B, Al Shehri D. CO2 rich cushion gas for hydrogen storage in depleted gas reservoirs: insight on contact angle and surface tension. Int J Hydrogen Energy 2024;50:1281–301.
[27] Prigmore S, Okon-Akan OA, Egharevba IP, Ogbaga CC, Okoye PU, Epelle E, et al. Cushion gas consideration for underground hydrogen storage. Encyclopedia 2024; 4:847–63.
[28] Heinemann N, Scafidi J, Pickup G, Thaysen EM, Hassanpouryouzband A, Wilkinson M, et al. Hydrogen storage in saline aquifers: the role of cushion gas for injection and production. Int J Hydrogen Energy 2021;46:39284–96.
[29] Luboń K, Tarkowski R. Hydrogen storage in deep saline aquifers: non-recoverable cushion gas after storage. Energies 2024;17:1493.
[30] Muhammed NS, Haq B, Al Shehri DA. Hydrogen storage in depleted gas reservoirs using nitrogen cushion gas: a contact angle and surface tension study. Int J Hydrogen Energy 2023;48:38782–807.
[31] Heinemann N, Alcalde J, Miocic JM, Hangx SJT, Kallmeyer J, Ostertag-Henning C, et al. Enabling large-scale hydrogen storage in porous media–the scientific challenges. Energy Environ Sci 2021;14:853–64.
[32] Saeed M, Jadhawar P. Optimizing underground hydrogen storage in aquifers: the impact of cushion gas type. Int J Hydrogen Energy 2024;52:1537–49.
[33] Cao C, Liao J, Hou Z, Xu H, Mehmood F, Wu X. Utilization of CO2 as cushion gas for depleted gas reservoir transformed gas storage reservoir. Energies 2020;13:576.
[34] Saeed M, Jadhawar P, Bagala S. Geochemical effects on storage gases and reservoir rock during underground hydrogen storage: a depleted North Sea oil reservoir case study, vol. 4; 2023. p. 323–37. Hydrogen.
[35] Zamehrian M, Sedaee B. Underground hydrogen storage in a partially depleted gas condensate reservoir: influence of cushion gas. J Pet Sci Eng 2022;212:110304.
[36] Gbadamosi AO, Muhammed NS, Patil S, Al Shehri D, Haq B, Epelle EI, et al. Underground hydrogen storage: a critical assessment of fluid-fluid and fluid-rock interactions. J Energy Storage 2023;72:108473.
[37] Viveros FE, Medina OE, Moncayo-Riascos I, Lysyy M, Benjumea PN, Cortés FB, et al. Hydrogen storage in depleted gas reservoirs using methane cushion gas: an interfacial tension and pore scale study. J Energy Storage 2024;98:113110.
[38] Hosseini M, Fahimpour J, Ali M, Keshavarz A, Iglauer S. H2– brine interfacial tension as a function of salinity, temperature, and pressure; implications for hydrogen geo-storage. J Pet Sci Eng 2022;213:110441.
[39] Alanazi A, Yekeen N, Ali M, Ali M, Abu-Mahfouz IS, Keshavarz A, et al. Influence of organics and gas mixing on hydrogen/brine and methane/brine wettability using Jordanian oil shale rocks: implications for hydrogen geological storage. J Energy Storage 2023;62:106865.
[40] Chow YTF, Maitland GC, Trusler JPM. Interfacial tensions of (H2O+ H2) and (H2O + CO2+ H2) systems at temperatures of (298–448) K and pressures up to 45 MPa. Fluid Phase Equil 2018;475:37–44.
[41] Chow YTF, Maitland GC, Trusler JPM. Erratum to "interfacial tensions of (H2O + H2) and (H2O + CO2 + H2) systems at temperatures of (298 to 448) K and pressures up to 45 MPa" [fluid phase equil. 475 (2018) 37–44]. Fluid Phase Equil 2020;503:112315. https://doi.org/10.1016/j.fluid.2019.112315.
[42] Mirchi V, Dejam M, Alvarado V. Interfacial tension and contact angle measurements for hydrogen-methane mixtures/brine/oil-wet rocks at reservoir conditions. Int J Hydrogen Energy 2022;47:34963–75.
[43] Hosseini M, Leonenko Y. Hydrogen-brine interfacial tension at subsurface conditions: implication for hydrogen geo-storage. In: 84th EAGE annu. Conf. Exhib., vol. 2023. European Association of Geoscientists & Engineers; 2023. p. 1–5.
[44] Doan QT, Keshavarz A, Miranda CR, Behrenbruch P, Iglauer S. A prediction of interfacial tension by using molecular dynamics simulation: a study on effects of cushion gas (CO2, N2 and CH4) for Underground Hydrogen Storage. Int J Hydrogen Energy 2024;50:1607–15.
[45] Yang Y, Nair AKN, Zhu W, Sang S, Sun S. Molecular perspectives of interfacial properties of the Hydrogen+ water mixture in contact with silica or kerogen. J Mol Liq 2023:122337.
[46] Nguyen VN, Tarełko W, Sharma P, El-Shafay AS, Chen W-H, Nguyen PQP, et al. Potential of explainable artificial intelligence in advancing renewable energy: challenges and prospects. Energy Fuels 2024;38:1692–712.
[47] Le AT, Pandey A, Sirohi R, Sharma P, Chen W-H, Pham NDK, et al. Precise prediction of biochar yield and proximate analysis by modern machine learning and SHapley additive exPlanations. Energy Fuels 2023;37:17310–27.
[48] Hosseini M, Leonenko Y. Prediction of hydrogen−brine interfacial tension at subsurface conditions: implications for hydrogen geo-storage. Int J Hydrogen Energy 2024;58:485–94. https://doi.org/10.1016/j.ijhydene.2024.01.227.
[49] Thanh HV, Zhang H, Dai Z, Zhang T, Tangparitkul S, Min B. Data-driven machine learning models for the prediction of hydrogen solubility in aqueous systems of varyingData-driven machine learning models for the prediction of hydrogen solubility in aqueous systems of varying salinity: implications for underground h. Int J Hydrogen Energy 2024;55:1422–33.
[50] Mouallem J, Raza A, Glatz G, Mahmoud M, Arif M. Estimation of CO2-brine interfacial tension using machine learning: implications for CO2 geo-storage. J Mol Liq 2024;393:123672.
[51] Pan B, Song T, Yue M, Chen S, Zhang L, Edlmann K, et al. Machine learning-based shale wettability prediction: implications for H2, CH4 and CO2 geo-storage. Int J Hydrogen Energy 2024;56:1384–90.
[52] Kanti PK, Shrivastav AP, Sharma P, Maiya MP. Thermal performance enhancement of metal hydride reactor for hydrogen storage with graphene oxide nanofluid: model prediction with machine learning. Int J Hydrogen Energy 2024; 52: 470-84..
[53] Thanh HV, Rahimi M, Dai Z, Zhang H, Zhang T. Predicting the wettability rocks/minerals-brine-hydrogen system for hydrogen storage: Re-evaluation approach by multi-machine learning scheme. Fuel 2023;345:128183.
[54] Ng CSW, Djema H, Amar MN, Ghahfarokhi AJ. Modeling interfacial tension of the hydrogen-brine system using robust machine learning techniques: implication for underground hydrogen storage. Int J Hydrogen Energy 2022;47:39595–605.
[55] Behnamnia M, Mozafari N, Monfared AD. Rigorous hybrid machine learning approaches for interfacial tension modeling in brine-hydrogen/cushion gas systems: implication for hydrogen geo-storage in the presence of cushion gas. J Energy Storage 2023;73:108995.
[56] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[57] Muhammed NS, Haq B, Al Shehri D. Role of methane as a cushion gas for hydrogen storage in depleted gas reservoirs. Int J Hydrogen Energy 2023;48:29663–81.
[58] Yekta AE, Manceau J-C, Gaboreau S, Pichavant M, Audigane P. Determination of hydrogen–water relative permeability and capillary pressure in sandstone: application to underground hydrogen injection in sedimentary formations. Transport Porous Media 2018;122:333–56.
[59] Higgs S, Da Wang Y, Sun C, Ennis-King J, Jackson SJ, Armstrong RT, et al. In-situ hydrogen wettability characterisation for underground hydrogen storage. Int J Hydrogen Energy 2022;47:13062–75.
[60] Esfandyari H, Sarmadivaleh M, Esmaeilzadeh F, Ali M, Iglauer S, Keshavarz A. Experimental evaluation of rock mineralogy on hydrogen-wettability: implications for hydrogen geo-storage. J Energy Storage 2022;52:104866.
[61] Esfandyari H, Sarmadivaleh M, Esmaeilzadeh F, Ali M, Iglauer S, Keshavarz A. Corrigendum to "Experimental evaluation of rock mineralogy on hydrogen-wettability: implications for hydrogen geo-storage" [J. Energy Storage, 52 (2022) 104866]. J Energy Storage 2023;57:106162. https://doi.org/10.1016/j.est.2022.106162.
[62] Esfandyari H, Shadizadeh SR, Esmaeilzadeh F, Davarpanah A. Implications of anionic and natural surfactants to measure wettability alteration in EOR processes. Fuel 2020;278:118392.

[63] Al-Mukainah H, Al-Yaseri A, Yekeen N, Al Hamad J, Mahmoud M. Wettability of shale–brine–H2 system and H2-brine interfacial tension for assessment of the sealing capacities of shale formations during underground hydrogen storage. Energy Rep 2022;8:8830–43.
[64] Al-Yaseri A, Abu-Mahfouz IS, Yekeen N, Wolff-Boenisch D. Organic-rich source rock/H2/brine interactions: implications for underground hydrogen storage and methane production. J Energy Storage 2023;63:106986.
[65] Yao Y, Qiu Y, Cui Y, Wei M, Bai B. Insights to surfactant huff-puff design in carbonate reservoirs based on machine learning modeling. Chem Eng J 2023;451:138022.
[66] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.