processes

MDPI

Article

Thermodynamic Variable-Augmented Ensemble Machine Learning for Feature Selection and Interfacial Tension Prediction in Underground Storage Systems

Mohammad Islam Miah [1,2,*], Md Maruf Alam [3], Ahmed Elghoul [4] and Travis Wiens [2]

[1] Department of Petroleum and Mining Engineering, Chittagong University of Engineering & Technology, Chattogram 4349, Bangladesh
[2] Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N5A9, Canada; t.wiens@usask.ca
[3] Institute of Energy Technology, Chittagong University of Engineering & Technology, Chattogram 4349, Bangladesh; mdmarufalam3120@gmail.com
[4] Division of Economics and Business, Colorado School of Mines, Golden, CO 80401, USA; elghoul_libya@hotmail.com
* Correspondence: islam.m@cuet.ac.bd or islam.miah@usask.ca; Tel.: +880-199-086-7065

Abstract

Reliable prediction of interfacial tension (IFT) is critical for optimizing engineering processes like hydrogen separation, high-density energy storage and fluid displacement in underground oil reservoir systems. The goals of this study are: (a) to assess the efficacy of optimized ensemble machine learning (EML) models and (b) parametric sensitivity analysis to interpret feature importance while predicting interfacial tension (IFT) across an extensive range of thermodynamic conditions. The dataset comprises 498 experimentally measured IFT records collected under diverse conditions of temperature (T), pressure (P), and gas compositions of hydrogen (H_2), carbon dioxide (CO_2), and methane (CH_4) in complex reservoir brine systems. The EML algorithms of extreme gradient boosting (XGB), random forest (RF), decision tree (DT), light gradient boosting (LGB) and adaptive boosting (ADB) are applied to construct the data-driven computational models to obtain IFT. The model's effectiveness and robustness are evaluated using statistical performance indices. SHapley Additive exPlanations (SHAP) analysis is adopted to interpret and rank feature importance, thereby ensuring the model's transparency and interpretability. The XGB model showed excellent performance, with high determination of coefficient (R^2, 83%) and least root mean square error (RMSE, 4.81 mN/m) compared to other models of RF (81, 5.06), DT (79.94, 5.23), LGB (77.52, 5.54) and ADB (69.32, 6.46) accordingly. Based on SHAP scores, it is revealed that T is the most significant parameter to control IFT, and we ranked them higher to lower weighty order, like T, H_2, P, CO_2 and CH_4. This study demonstrates the efficacy of EML and physics-guided computational models, coupled with feature selection techniques, in capturing IFT phenomena relevant to a subsurface hydrogen–brine geological system. It also establishes an interpretable artificial intelligent framework to uphold the project design, risk evaluation, and optimization of underground hydrogen storage systems in saline aquifers and other geologically complex formations for energy transition.

Keywords: machine learning; features selection; thermodynamic behavior; computational model; underground reservoir storage

Check for updates

Academic Editor: Federica Raganati

Received: 1 September 2026
Revised: 6 October 2026
Accepted: 16 October 2026
Published: 26 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background of Study

The global energy transition has intensified interest in underground energy storage systems as a means of bridging supply–demand gaps, stabilizing intermittent renewable energy sources, supporting large-scale decarbonization efforts, and playing an integral role in the future sustainable energy landscape. For instance, accurate interfacial tension (IFT) prediction is critical for assessing the performance and safety of underground CO_2 and H_2 storage systems because IFT governs capillary pressure, multiphase flow, fluid displacement, and trapping within underground geological porous media. Reliable IFT assessments enable more accurate predictions of injection behavior, plume migration, storage capacity, residual and capillary trapping, and leakage risk. For CO_2 storage, accurate IFT progresses assessment of plume containment and caprock sealing, whereas for H_2 storage, it is particularly significant for predicting hydrogen–brine displacement, water invasion, and recovery efficiency during cyclic injection and withdrawal. Therefore, incorporating reliable IFT predictions into reservoir and pore-scale models can reduce uncertainty and support the design of safer, more efficient, and economically viable underground CO_2 and H_2 storage systems.

The inferred IFT behavior for an underground CO_2 storage system, from surface capture and compression to subsurface injection, migration, and trapping within a geological formation, is as follows. CO_2 is injected into a porous sandstone reservoir beneath an impermeable shale caprock, where its interaction with formation brine is governed by IFT at the CO_2–brine interface. At the pore scale, IFT contributes to capillary pressure and influences CO_2 displacement, migration, and residual trapping within the pore network. For the underground CO_2 storage system, it highlights four principal trapping mechanisms—structural, residual (capillary), solubility, and mineral trapping—through which injected CO_2 can be retained over different timescales. Consequently, IFT is an important parameter for evaluating storage efficiency, injectivity, plume migration, leakage risk, and long-term storage security, highlighting the need for accurate IFT characterization under reservoir conditions. Figure 1 illustrates the role of IFT in underground H_2 storage within a porous geological formation.

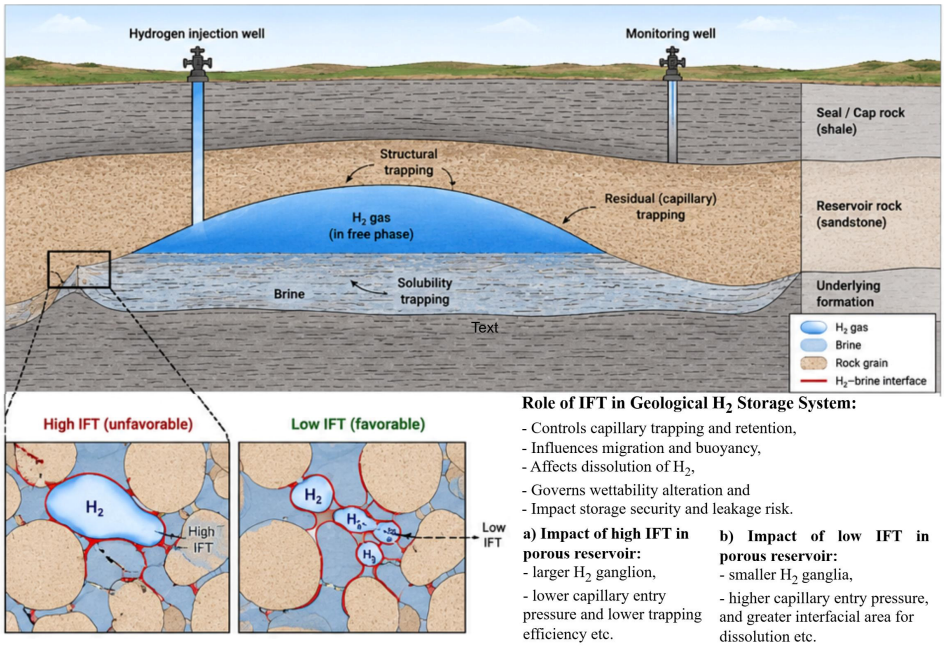


Figure 1. An illustration of a typical H_2 storage system for showing role of IFT in geological porous medium.

In particular, IFT serves as a fundamental thermophysical property that connects reservoir conditions with multiphase fluid behavior, ultimately governing the long-term integrity of subsurface hydrogen storage systems. Therefore, a lower IFT between formation brine and H_2 enhances capillary trapping, promotes hydrogen dissolution into the brine, and improves overall storage security. Consequently, IFT is a critical parameter in the design, performance evaluation, and risk assessment of geological hydrogen storage systems.

Notably, hydrogen is injected into a sandstone reservoir beneath an impermeable shale caprock, where it accumulates as a free gas phase and is retained through structural, capillary residual, and solubility trapping mechanisms. At the pore scale, IFT strongly influences the distribution and mobility of H_2 within the pore network. Under high IFT, stronger interfacial forces can restrict H_2 displacement and promote the formation of larger continuous gas clusters, which may result in less favorable injectivity and storage performance. In contrast, low IFT facilitates the breakup and redistribution of H_2 into smaller pore-scale clusters, potentially improving displacement efficiency and reducing capillary resistance. Therefore, accurate characterization of H_2–brine IFT under reservoir conditions is essential for predicting hydrogen migration, trapping efficiency, injectivity, storage capacity, and potential leakage pathways, making IFT an important parameter in the design and assessment of safe and efficient underground hydrogen storage systems.

1.2. Relevant Literature Reviews and Research Scopes

Over the decades, extensive research has been conducted to investigate IFT in subsurface hydrogen–water systems. Experimental studies have established baseline IFT values, typically ranging from 20 to 40 mN/m, while also indicating that these values are highly sensitive to variations in pressure (P); temperature (T); and fluid composition, such as methane (CH_4), hydrogen (H_2), carbon dioxide (CO_2), and salinity concentration [1–3]. Mineralogical characteristics have also been investigated for sandstone, basalt, and calcite surfaces, revealing distinct wettability and interfacial tension (IFT) behaviors for each mineral type [4,5]. Furthermore, gases, which include CO_2, CH_4, and nitrogen (N_2), have been noted to have substantial effects on IFT, which promotes multiphase flow, but they have negative effects on solubility [6–9]. In addition, molecular dynamic (MD) has validated these findings, suggesting approximate responses for these studies, which differ by approximately ±2–3 mN/m [10–12]. The most prominent limitation for these studies is that most findings have been relatively narrow in scope, and research data on gas effects have been carried out individually. Meanwhile, machine learning (ML) has also been established as a promising predictor in the field of IFT for underground porous reservoir systems. Also, ML has appeared as a powerful approach for predicting IFT in subsurface H_2–brine systems in recent years. Numerous scholars adopted hybrid ML algorithms, such as support vector regression (SVR) and deep neural network (DNN), which are proven to be effective in accurately forecasting IFT ($R^2 > 0.9$, RMSE < 0.6 mN/m) using various datasets of H_2–brine fluids [13–20]. The performance of ML in conjunction with MD simulations has also been found to be promising [21]. Although significant progress has been achieved, most ML methods are also limited, to a great extent, due to data size and decision-making mechanisms. Some of these problems have also been addressed in critical reviews and meta-analysis by different scholars [22–26]. Notably, Behnamnia et al. [13] demonstrated the possibility of achieving coefficient of determination, R^2, values of 0.95, with root mean square error (RMSE) < 0.5 mN/m, by ANN model. In turn, Hosseini & Leonenko [17] applied ANN to H_2–brine datasets, reporting $R^2 = 0.94$, with error margins within ±2 mN/m. Ibrahim [18] extended this work to DNN, achieving >90% accuracy, with RMSE < 0.6 mN/m. Janjua et al. [14] employed ensemble ML techniques for

density and IFT prediction, which exhibited high precision. The works by Kohzadvand et al. [27] and Tariq et al. [7] were able to capture wettability shifts within ±5° using SVR modeling. Further enhancements of predictive accuracy were achieved using ensemble ML approaches, such as bagging-based random forest (RF) in combination with extreme gradient boosting (XGB). Sample examples include the works of Turkson et al. [16], Sofian et al. [20] and Sagandykova et al. [28]. Hybrid frameworks that couple MD simulations with ML have emerged as a promising strategy for modeling complex interfacial phenomena. For instance, Omrani et al. [21] validated predictions within ±3 mN/m using an MD–ML integrated framework. Some recent studies on IFT prediction using ML are summarized in Table 1. These studies demonstrate that the IFT of subsurface H_2–brine systems is highly sensitive to pressure, salinity, mineralogy, and cushion gas composition, typically ranging from 20 to 40 mN/m under geological storage conditions. Experimental investigations, MD simulations, and ML models have collectively advanced the understanding of H_2–brine interfacial behavior while improving predictive capability.

Table 1. Recent ML applications for predicting IFT in H_2–brine systems.

Author (s)	Input Variables	ML Model (s)	Remarks
Omrani et al. [21]	H_2, P, T, salinity	Hybrid ANN	-ANN predicted IFT with R^2 of 0.92, and MD validated within ±3 mN/m.
Behnamnia et al. [13]	H_2, CH_4, CO_2, P, T, salinity	RF, ANN	-RF predicted IFT with R^2 of 0.95, and ANN with RMSE < 0.5 mN/m.
Janjua et al. [14]	T, P, salinity	RF, XGB	-Ensemble ML predicted IFT with R^2 of 0.93.
Hosseini & Leonenko [17]	T, P, H_2	ANN	-ANN predicted IFT with R^2 of 0.94, and error ± 2 mN/m.
Kalam et al. [29]	T, P, H_2, density	ANN	-RMSE predicted IFT with <0.5 mN/m and accuracy > 90%.
Pan et al. [15]	P, T, gas type, ionic strength	Hybrid GA	-Genetic algorithm-based automatic machine learning and symbolic regression predicted brine–gas IFT with R^2 of 0.97.
Turkson et al. [16]	H_2, CH_4, CO_2, N_2, P, T	XGB	-Gradient boosting regression model showed R^2 of 0.98 and RMSE of 0.9 mN/m.
Xie et al. [30]	H_2, CO_2, salinity	ANN	-ANN predicted IFT with R^2 of 0.93 and error ± 2 mN/m.
Sagandykova et al. [28]	P, T, H_2, specific gravity of rock/minerals	RF	-This model predicted wettability shifts with R^2 of 0.91.
Sofian et al. [20]	Salinity, P, T, density, H_2, CH_4	RF	-RF predicted IFT with R^2 of 0.92 and RMSE < 0.6 mN/m.
Ibrahim [18]	T, P, H_2, CH_4, CO_2, salinity	DNN	-DNN predicted IFT with >90% accuracy and RMSE < 0.6 mN/m.

Despite these advances, several critical challenges remain, including limited experimental datasets, the predominance of studies focusing on individual cushion gases, inadequate integration of coupled physicochemical processes, the scarcity of dynamic IFT investigations, and the lack of transparent and interpretable ensemble ML frameworks. These limitations highlight important directions for future research.

Overall, although substantial progress has been made in characterizing H_2–brine IFT, its behavior under realistic geological storage conditions remains only partially understood, necessitating more comprehensive experimental, computational, and data-driven investigations. Despite these advances, most ML applications reported in the literature remain subject to several important limitations. First, the available datasets are generally limited in size and cover relatively narrow ranges of P, temperature T, and salinity, thereby restricting the generalizability and robustness of the developed models. In addition, most existing ML models operate as "black-box" predictors, delivering high predictive accuracy while providing limited insight into the relative influence of the input variables. This lack of interpretability hinders their practical application in geological hydrogen storage, where understanding the sensitivity and physical significance of controlling parameters is as critical as achieving accurate predictions. Consequently, the development of interpretable, physics-informed, and robust ML frameworks trained on more comprehensive datasets remains a key research priority. While machine learning models show strong predictive capabilities for IFT in H_2–brine systems, most suffer from small datasets and opaque structures of decisions. The absence of an interpretability framework has further limited their applicability in operational contexts, where parameter sensitivity understanding is as crucial as prediction accuracy. Building on the above-mentioned advancements and addressing these gaps, the present study evaluates the efficacy of optimized ensemble ML algorithms, such as XGB, RF, decision tree (DT), adaptive boosting (ADB) and light gradient boosting (LGB), with interpretable SHapley Additive Explanations (SHAP) analysis. A dataset of 498 experimentally measured IFT records in diverse thermodynamic conditions with variables such as T, P, and gas composition of H_2, CO_2, and CH_4 in complex brine systems makes up the datasets used herein. The effectiveness and robustness of the models are evaluated using statistical performance indices, while SHAP and rank feature importance. By embedding the bagging and boosting-based models within explainable AI approach, not only is this study focused on achieving robust predictive accuracy, but it also allows for transparency in thermodynamic feature importance such that the gap between predictive performance and interpretability is bridged. This further advances the state of the art in IFT prediction and provides an intelligent framework that is scientifically sound and practically relevant for applications pertaining to underground CO_2 and H_2 storage.

1.3. Research Objectives and Contribution of This Study

The present study contributes to the inventive approach by exploring the potential effectiveness of the EML algorithms, including XGB, RF, DT, LGB, and ADB, for assessing the physical properties and simulating the essential properties, including variations in T, P, and gas compositions of H_2, CO_2, and CH_4. In contrast to the traditional literature, which relied on a dataset consisting of 498 experimentally observed records for the study of the IFT phenomenon at different thermodynamic conditions (e.g., T, P, and gaseous mixture variations in H_2, CO_2, and CH_4), the current study undertakes a predictive accuracy assessment alongside a novel addition, viz., the SHAP feature importance analysis methodology for improving the study limitations to a greater degree. This can help to bridge the gap for a scientific study relevant to real-life problems related to the storage of H_2 gases at underground levels. The present study aims to advance the IFT prediction with EML algorithms for subsurface H_2–brine systems. In this study, the major objectives are listed below:

- To assess the efficacy of optimized ensemble machine learning models;
- Parametric sensitivity analysis to interpret feature importance while predicting IFT across an extensive range of thermodynamic conditions.

Notably, the novelty of this study lies in the development of optimized ensemble machine learning models integrated with SHAP-based explainable artificial intelligence for predicting H_2–brine interfacial tension under diverse thermodynamic conditions. In addition to achieving high predictive accuracy, the proposed framework provides transparent interpretation of the governing physicochemical variables by quantifying both global and local feature contributions. This combination of predictive performance and interpretability offers a reliable and physically meaningful tool for interfacial property estimation in geological storage systems.

2. Materials and Methods

2.1. Dataset Integration and Preparation

A list of 498 experimental measurements of IFT between H_2 and brine is compiled to develop intelligent models using ensemble machine learning techniques. Under complex salinity conditions with deionized and saline water, all reliable datasets were compiled from peer-reviewed published studies, ensuring data traceability, reliability, and reproducibility [4,14,19,31–37]. The dataset includes a broad spectrum of thermodynamic conditions, including temperatures ranging from ambient to reservoir-scale levels and pressures extending from near-surface to deep subsurface environments, along with compositional variations such as pure H_2, H_2 mixed with cushion gases (e.g., CO_2 and CH_4), and brines of different salinity levels. This wide variability enables the model to effectively learn the complex nonlinear relationships governing IFT as a function of multiple interacting variables. Descriptive statistics are calculated to summarize the dataset's key characteristics, assess the distribution and variability of each variable, detect potential anomalies, and provide an overall understanding of the data prior to machine learning model development. In addition, all input variables are standardized using Z-score normalization to ensure consistency and comparability across heterogeneous data sources:

$$X_{scaled} = \frac{(X - \mu_X)}{(\sigma_X)} \quad (1)$$

where μ_X is the mean and σ_X is the standard deviation of raw variable X.

This transformation yields dimensionless variables, minimizes bias arising from differing measurement scales, and improves the stability and convergence of the model training process. Outliers are detected using interquartile range analysis and subsequently removed to avoid distortion in model learning. The cleaned dataset is then categorized into training (65%) and testing (35%) sets to ensure unbiased evaluation of predictive model performance and reliability.

2.2. Development of Ensemble Machine Learning Models

To capture the nonlinear and multivariate relationships governing interfacial tension (IFT), five ensemble machine learning (EML) algorithms are employed: XGB (XGB), random forest (RF), decision tree (DT), light gradient boosting machine (LGB), and adaptive boosting (ADB). Each model is chosen for its capability to effectively manage the bias–variance trade-off, accommodate heterogeneous datasets, and deliver robust predictive performance across diverse thermodynamic conditions. A general supervised ML model can be expressed as

$$\hat{y}_{IFT} = f(X\,;\,\theta) \quad (2)$$

where $\hat{y}_{IFT}$ is the predicted IFT, $X = \left[T,\ P,\ x_{H_2},\ x_{CO_2},\ x_{CH_4}\right]$ is the input feature vector (e.g., T, P, gas mole fractions), and θ represents model parameters.

A single decision tree (DT) partitions the feature space into distinct regions and assigns a constant prediction within each region. Although this structure offers interpretability, DTs are highly prone to overfitting due to their sensitivity to training data. Random forest (RF), an ensemble of decision trees trained on bootstrapped samples, addresses this limitation by aggregating (averaging) the predictions from multiple trees, which can be expressed for N number of samples as follows [38]:

$$\hat{y}_{IFT} = \frac{1}{N}\sum_{i=1}^{N} h_i X \quad (3)$$

The boosting-based ensemble methods, such as XGB, LGB, and ADB, iteratively construct weak learners, where each successive model is trained to correct the errors of its predecessors. The final prediction is obtained as a weighted combination (sum) of these weak learners, thereby improving overall accuracy and model robustness:

$$\hat{y}_{IFT} = \sum_{m=1}^{M} \alpha_m h_m(X) \quad (4)$$

where α_m is the weight assigned to learner h_m. Boosting is particularly effective for capturing subtle nonlinearities in IFT behavior.

To achieve an optimal balance among bias, variance, and interpretability, each model was tuned using its specific set of hyperparameters. The learning rate, maximum tree depth, and number of estimators are systematically tuned using a grid-search strategy coupled with 10-fold cross-validation to identify the optimal model configuration. This strategy ensured high predictive accuracy while mitigating the risk of overfitting. The use of multiple models further enabled systematic benchmarking and facilitated the identification of the most effective algorithm for IFT prediction. To facilitate comprehensive model training and testing through an iterative trial-and-error approach, the dataset is divided into training and testing subsets, with 65% (324 samples) allocated for training the EML-based models and the remaining 35% (174 samples) reserved for testing. The relatively large test set is selected to ensure statistically reliable performance evaluation while minimizing the potential risk of overfitting. Notably, the training dataset is subsequently subjected to 10-fold cross-validation, where the training data were iteratively divided into nine folds for model fitting and one fold for validation. The cross-validation results for predictive models are used to identify the optimal hyperparameters. After selecting the optimal hyperparameters, the model is retrained using the complete training dataset and evaluated only once on the independent 35% testing dataset. For the bagging-based random forest (RF) model, the optimized hyperparameters are n_estimators = 100, max_features = 1.0, min_samples_leaf = 1, and the minimum number of samples required to split an internal node = 2, while min_samples_leaf: 1, and min_samples_split: 2 for DT. The hyperparameters settings for the boosting-based model are n_estimators: 100, learning_rate: 0.1, max_depth: 3 for XGB; learning_rate: 0.1, max_depth: −1, min_child_samples: 20, min_child_weight: 0.001, n_estimators: 100, and num_leaves: 31 for LGB; learning_rate: 1.0, and n_estimators for 50 for ADB algorithm.

2.3. Model Performance Indicators

The performance of the models is evaluated using standard statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), average absolute relative error (AAPE), and mean absolute error (MAE), enabling a quantitative

assessment of predictive accuracy and robustness across different models. These metrices are mathematically expressed in Equations (5)–(8) as below [39] while predicting IFT:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(IFT_{t,i} - IFT_{p,i}\right)^2}{\sum_{i=1}^{N}\left(IFT_{t,i} - IFT_{t,mean}\right)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(IFT_{t,i} - IFT_{p,i}\right)^2} \quad (6)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|IFT_{t,i} - IFT_{p,i}\right| \quad (7)$$

$$AAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{IFT_{t,i} - IFT_{p,i}}{IFT_{t,i}}\right| \times 100 \quad (8)$$

where $IFT_{t,i}$, $IFT_{p,i}$ and $IFT_{t,mean}$ represent the measured, predicted and mean value of interfacial tension, respectively.

It is noteworthy that R^2 quantifies the proportion of variance in the experimental IFT values that is explained by the model, with values closer to 1 indicating stronger predictive performance. RMSE assigns greater weight to larger errors, making it particularly sensitive to outliers and overall prediction deviations. MAE provides a straightforward and interpretable measure of the average magnitude of prediction errors [40]. By combining these three metrics, this study ensures a more robust, balanced, and reliable assessment of predictive performance. For each model, performance is recorded on both the validation and test sets to monitor overfitting and assess generalization.

Beyond the conventional error-based evaluation metrics, a customized Taylor diagram [41] is employed to provide a graphical assessment of the agreement between the predicted and measured IFT values. The diagram simultaneously incorporates the standard deviation (STD), correlation coefficient (r), and RMSE, enabling a comprehensive evaluation of model performance. This integrated representation facilitates the assessment of the models in terms of correlation, variability, and prediction error, thereby providing a clear comparison of their overall predictive capability. It is noteworthy that direct regression prediction for R^2 and correlation-based statistics for r are computed separately using standard statistical functions.

2.4. Parametric Sensitivity Analysis and Model Interpretability

Pearson's correlation coefficient (r) is determined for the entire dataset to quantify the relationships among thermodynamic properties, identify feature interdependencies, and evaluate their respective influences on IFT. Notably, Pearson correlation measures the strength and direction of the linear association between an individual input variable and IFT. In contrast, the interpretable ensemble machine learning technique with SHapley Additive Explanations (SHAP) approach is utilized by considering nonlinear relationships and interactions with other input variables to find the importance of thermodynamic features while predicting IFT in H_2-rich brine systems. Moreover, SHAP bar plots are used to illustrate the mean absolute SHAP values of the predictor features, thereby quantifying their overall contributions to the model's predictions. A larger mean absolute SHAP value indicates a greater influence of the corresponding feature on IFT prediction, whereas a smaller value suggests a relatively weaker contribution. For a given model output, the

SHAP value quantifies the contribution of each predictor feature to the deviation of the individual prediction from the model's baseline output by Equation (9) [42,43]:

$$\mathrm{SHAP}_i = \sum_{F \subseteq Q \setminus i} \frac{|F|!(Q - |F| - 1)!}{Q!} [f(F \cup i) - f(F)] \quad (9)$$

where F consists of a set of non-zero directories and Q defines the set of all features.

The summarized methodological steps are illustrated in Figure 2 to develop the ensemble ML algorithm-based predictive model and interpretable SHAP for IFT prediction using thermodynamic properties. The framework begins with experimental data collection, including key thermodynamic parameters, followed by data processing and quality assurance to improve data consistency and reliability. The processed dataset is subsequently partitioned into training and testing subsets for intelligent data-driven model development. An appropriate EML algorithm is chosen, and the model is trained using optimized hyperparameters. The optimized model is then employed to predict IFT, while its performance is evaluated using statistical indicators. If the model accuracy does not meet the predefined criteria, the modeling process is iteratively refined through further training and an optimization process. Once satisfactory predictive performance is achieved, the final intelligent model is subjected to SHAP-based interpretability analysis to quantify the contribution of individual input features and identify the most influential parameters controlling IFT. Finally, feature selection is performed based on the interpretable model results, providing a robust and transparent framework for accurate IFT prediction and identification of the dominant thermodynamic factors in hydrogen-rich brine systems.

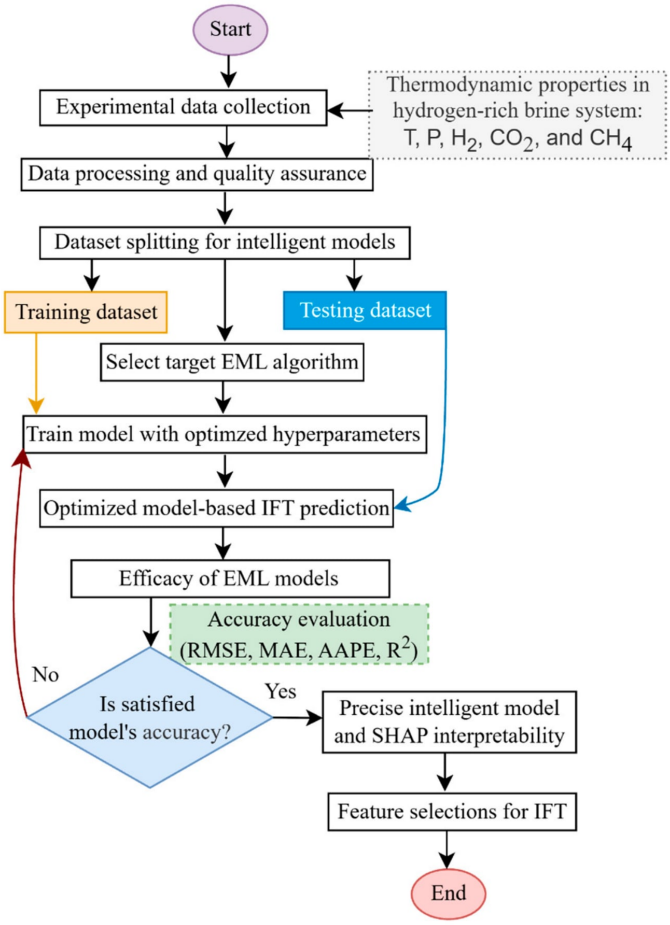


Figure 2. A flow chart to illustrate the major methodological steps with intelligent models for IFT prediction.

3. Results

3.1. Data Analysis

The final dataset comprises 498 experimental interfacial tension (IFT) measurements, encompassing diverse thermodynamic states and compositional variations to ensure comprehensive ensemble ML model development and evaluation. Table 2 presents a summary of the descriptive statistics of each input variable and output variable of IFT to assess data variability and standard deviation (STD) and ensure its suitability for subsequent predictive modeling.

Table 2. Descriptive statistics of predictor variables and IFT.

Variable	Minimum	Maximum	Mean	Median	STD
Temperature, T (°F)	71.6	348	125.97	122	58.38
Pressure, P (psi)	14.5	6554	1532.70	1000	1318.14
CO_2 (%)	0.0	70	4.80	0.00	13.58
CH_4 (%)	0.0	100	11.01	0.00	24.53
H_2 (%)	0.0	100	84.19	100	26.08
IFT ($mN{\cdot}m^{-1}$)	30.1	89	64.25	67	11.23

The output variable of IFT varied from 30.1 to 89.0, with an average of 64.25 $mN{\cdot}m^{-1}$ and a standard deviation of 11.23 with moderate variability. The predictor variable of T varied from 71.6 °F to 348.3 °F, with a mean of 125.97 °F and a standard deviation of 58.38 °F. P varied widely from 14.5 psi to 6554.0 psi, with a mean of 1532.70 psi and a standard deviation of 1318.14 psi. Gas compositions showed a dominant presence of H_2, with an average and standard deviation of 84.19 and 26.08. CO_2 and CH_4 were present in moderate concentrations of 4.80 and 11.01.

3.2. Predictive Model Performance

Five ensemble machine learning models with XGB, RF, DT, LGB, and ADB algorithms were adopted to integrate the thermodynamic variables into the predictive models. Table 3 presents the comparative metrics of all five models on how well each ensemble model performs in terms of predictive accuracy with RMSE, MAE, AAPE, and R^2 values on both training and testing datasets. Figure 3 illustrates the comparison of RMSE values among the five ML models, illustrating their respective prediction errors and providing insight into their ability to capture the nonlinear relationships between the thermodynamic input variables and IFT.

Table 3. Comparative performance metrics across EML models.

EML Model	RMSE ($mN{\cdot}m^{-1}$)		MAE ($mN{\cdot}m^{-1}$)		AAPE (%)		R^2 (%)	
	Train	Test	Train	Test	Train	Test	Train	Test
XGB	3.19	4.81	1.80	3.43	2.49	5.14	91.56	83.05
RF	3.10	5.06	1.75	3.66	3.33	5.51	90.38	81.22
DT	3.18	5.23	1.77	3.63	3.45	5.54	90.35	79.94
LGB	4.35	5.54	3.20	4.42	5.07	6.98	84.00	77.52
ADB	6.04	6.46	4.57	4.91	7.23	7.87	69.59	69.32

Among the five EML models, XGB consistently outperformed alternatives, achieving the lowest test error, MAE (3.43 $mN{\cdot}m^{-1}$), and highest test coefficient of determination, R^2 (83.05%). It is noteworthy that RF and DT followed closely with accuracy, with test R^2 values of 81.22% and 79.94% along with error of MAE (3.66 and 3.63 $mN{\cdot}m^{-1}$), respectively. Compared to bagging- and boosting-based models of RF and XGB, LGB and ADB showed

weaker generalization, with test R^2 values of 77.52% and 69.32%. Notably, the XGB model exhibits a training RMSE of 3.19 mN·m^{-1} and a testing RMSE of 4.81 mN·m^{-1}, indicating relatively good predictive performance across the two datasets. In contrast, ADB exhibits the highest RMSE, with values of 6.04 mN·m^{-1} for training and 6.46 mN·m^{-1} for testing. Overall, XGB achieves the lowest testing RMSE among the five ML models, followed by RF and DT, whereas LGB and AdaBoost exhibit comparatively higher prediction errors. The differences between training and testing RMSE provide further insight into the generalization capabilities of the models for IFT. Although XGB achieves the lowest testing error, its training–testing RMSE gap of 1.62 mN·m^{-1} indicates some variation in predictive performance between the two datasets. The findings are also responsible for establishing that it is beneficial for any algorithm, given its use of gradients, to perform tasks capturing inherent nonlinear relationships in data given during the computation of IFT prediction. The high accuracy of XGB can also be attributed to its nature as the boosting algorithm. The low R^2 recorded in ADB also confirms its sensitivity to noise and out. It should also be noted that RF and DT offered relatively comparable levels of accuracy, but as the approach focuses purely based on data-driven splits, generalization and accuracy were compromised in favor of the XGB. However, the moderate performance of approaches such as LGB and ADB indicates their greater sensitivity and limited capability in capturing complex nonlinear relationships. To comprehensively assess the predictive performance of the machine learning models beyond conventional scalar error metrics, a Taylor diagram was employed. This graphical approach simultaneously evaluates three key performance measures: the correlation coefficient (r) between the observed and predicted IFT, the STD of the predictions, and the centered RMSE, as summarized in Table 4. Collectively, these metrics provide a holistic assessment of each model's ability to reproduce the trend, variability, and overall accuracy of the target variable, as illustrated in the Taylor diagram in Figure 4. It is shown in Figure 4 that the XGB model is closest to the observed point, confirming that it is the best-performing model compared to the other intelligent models. As shown in Table 4, the XGB model exhibits near perfect agreement with the observations, achieving a correlation coefficient of 91.34% and a standard deviation (11.18) almost identical to the observed value (11.60).

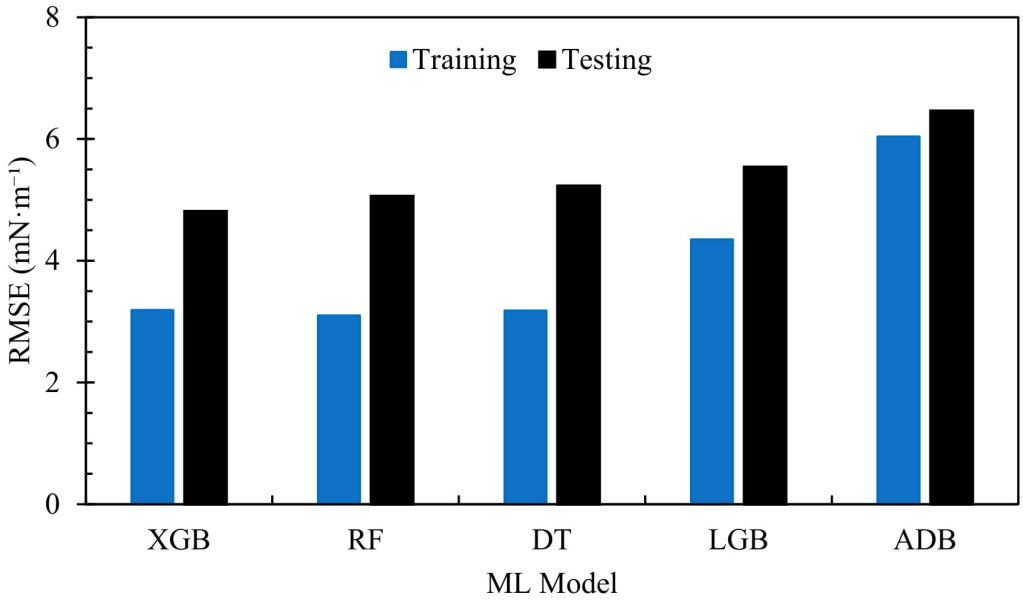


Figure 3. An illustration of RMSE values among the five ML models for IFT.

Table 4. Predictive fidelity of the machine learning models.

Intelligent Model	STD (mN·m^{-1})	Correlation Coefficient (r), %
XGB	11.18	91.34
RF	10.79	90.15
DT	11.45	89.82
LGB	10.09	88.07
ADB	9.064	83.68

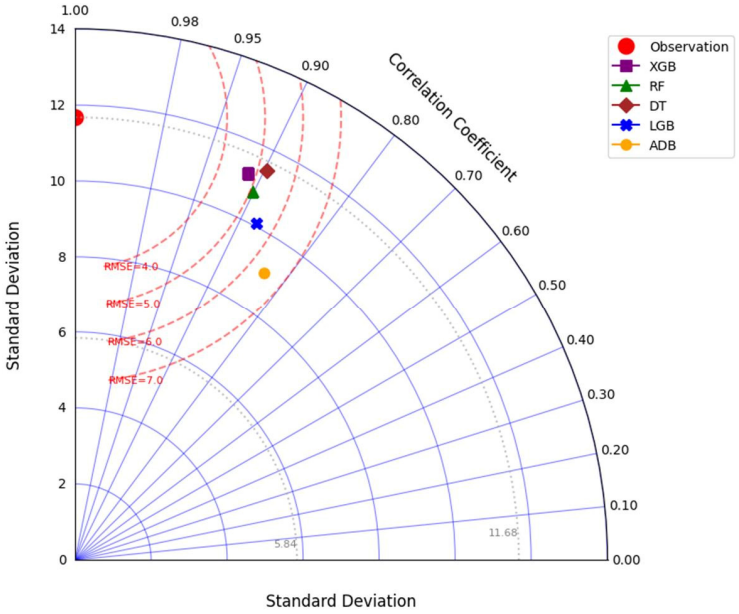


Figure 4. A Taylor diagram illustrates the efficacy of intelligent models' fidelity for testing datasets in IFT prediction.

3.3. Effects of Thermodynamic Variables for Predictive Models and SHAP-Based Interpretability

Pearson correlation coefficient (r) is computed to assess linear relationships among thermodynamic variables. Figure 5 illustrates a heatmap with the correlation matrix, exhibiting a relationship between IFT and thermodynamic variables, including temperature (T), pressure (P), methane (CH_4), hydrogen (H_2) and carbon dioxide (CO_2). Temperature showed a moderate negative correlation with IFT (r = −0.56), suggesting that increasing temperature contributes to a reduction in interfacial tension. CO_2 concentration exhibited a strong inverse linear relationship within the dataset for IFT (r = −0.57), whereas H_2 concentration displayed a moderate positive relationship (r = 0.48), suggesting that increasing H_2 content contributes to higher interfacial tension. CH_4 concentration and pressure showed relatively weak negative linear association when considered independently with IFT (r = −0.19 and −0.17, respectively).

Overall, the linear correlation analysis highlights that temperature (T) and H_2 concentration are the most influential variables governing IFT behavior. The strong negative correlation between temperature (T) and IFT, with an r = −0.56, is consistent with the well-established thermodynamic principle that increasing temperature weakens intermolecular cohesive forces at the fluid interface, thereby reducing IFT. Likewise, the strong negative correlation between CO_2 concentration and IFT, with an r = −0.57, reflects the high solubility of CO_2 in the aqueous phase, which promotes adsorption at the interface and lowers interfacial tension. In contrast, H_2 concentration displays a moderate positive correlation with IFT, with an r = 0.48, indicating that increasing H_2 content is associated with higher interfacial tension, likely due to its limited solubility and weaker interfacial activity compared with CO_2. Furthermore, the strong negative correlation between CH_4

and H_2 concentrations (r = −0.86) reflects their inverse compositional relationship, whereby increasing CH_4 content corresponds to lower H_2 fractions. Although CH_4 itself exhibits only a weak correlation with IFT, its inverse relationship with H_2 highlights compositional trade-offs that may indirectly influence interfacial behavior. These findings align with empirical studies reporting that H_2-rich systems tend to maintain higher IFT compared to CO_2-dominated regimes. To evaluate model interpretability, SHAP analysis is performed using the XGB model, and the resulting feature importance rankings were compared with those obtained from the RF and LGBM models. The corresponding SHAP bar plots are presented in Figures 6–8. The summarized results of feature importance are listed in Table 5. The SHAP analysis revealed distinct feature importance patterns across the efficient machine learning models while maintaining consistent identification of the key predictors.

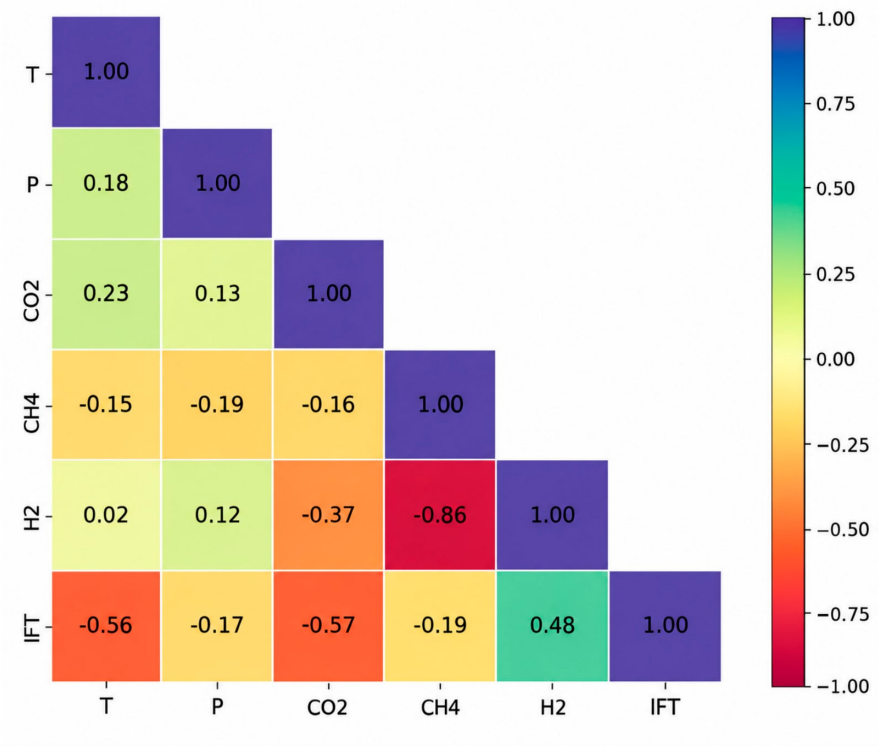


Figure 5. The heatmap illustrates the feature relationship with IFT of the studied dataset.

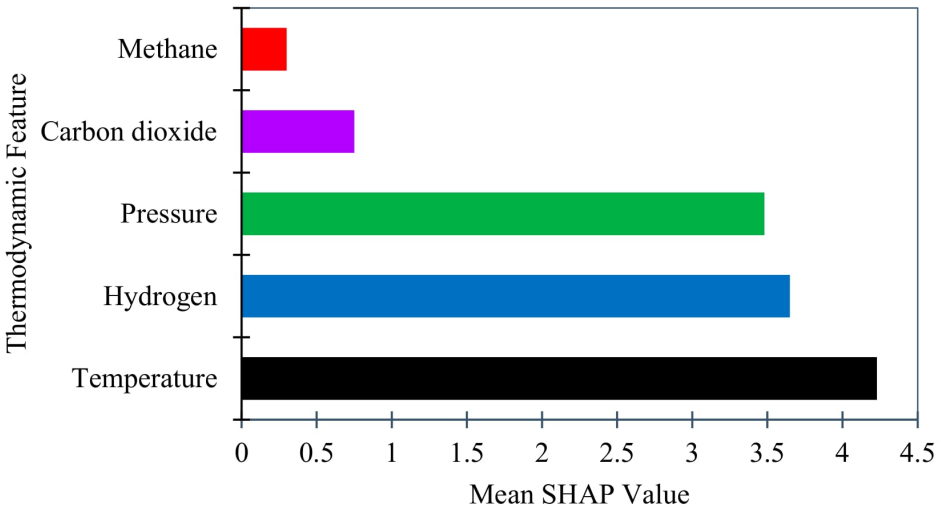


Figure 6. An illustration of a SHAP bar plot showing the relative importance of predictor features for IFT prediction using XGB.

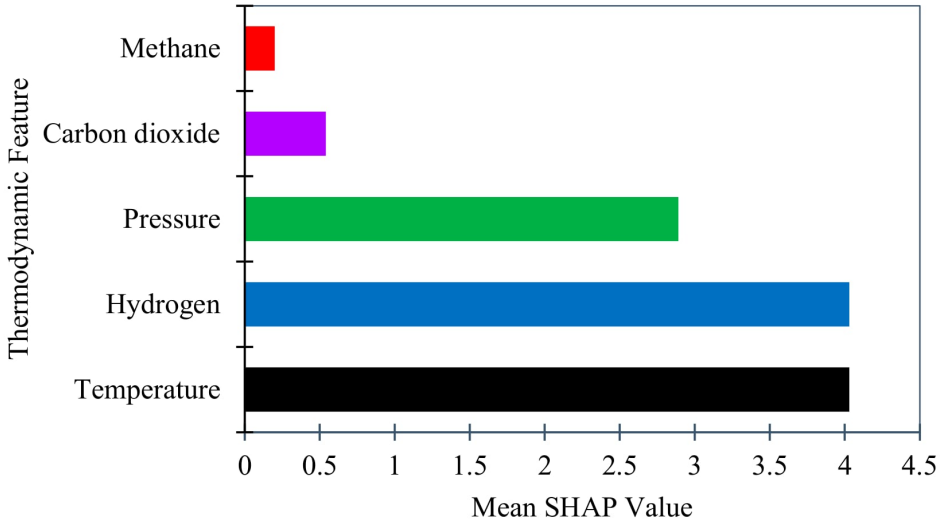


Figure 7. An illustration of a SHAP bar plot showing the relative importance of predictor features for IFT prediction using RF.

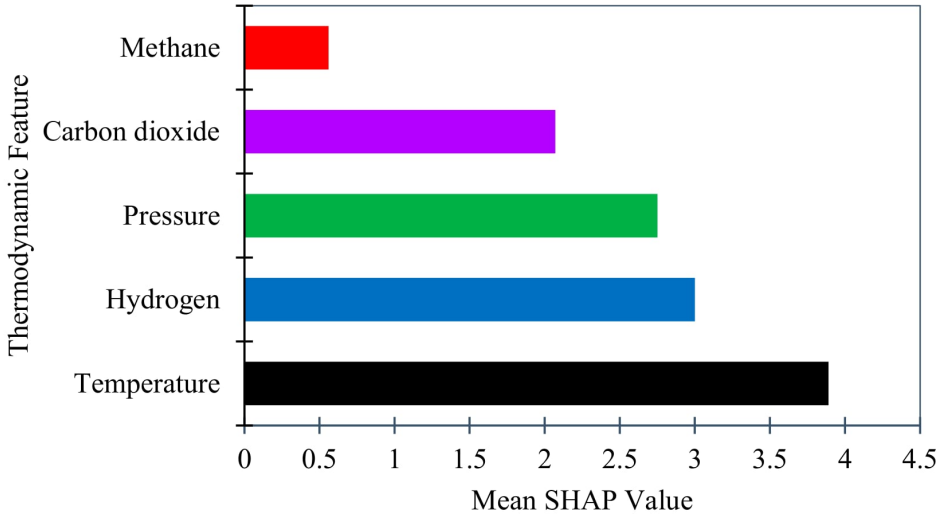


Figure 8. An illustration of a SHAP bar plot showing the relative importance of predictor features for IFT prediction using LGB.

Table 5. SHAP feature importance across models for IFT prediction in H_2-rich brine system.

Feature	XGB	RF	LGB
T (°F)	4.23	4.03	3.89
H_2 (%)	3.65	4.03	3.00
P (psi)	3.48	2.89	2.75
CO_2 (%)	0.75	0.54	2.07
CH_4 (%)	0.30	0.20	0.56

In the XGB model, temperature (T) exhibited the highest importance, with a SHAP score of 4.23, followed by H_2 concentration (3.65) and pressure (3.48), whereas CO_2 and CH_4 contributed only marginally, with SHAP scores of 0.75 and 0.30, respectively. The RF model assigned equal importance to T and H_2 (SHAP = 4.03 each), with P remaining the third most influential feature (2.89). In contrast with XGB and RF, the LGB model demonstrated a more balanced distribution of feature importance, with a SHAP score of 3.89 for T, 3.00 for H_2, 2.75 for P, and 2.07 for CO_2 concentration, all contributing substantially to model predictions. The SHAP analysis consistently revealed that temperature (T), H_2 concentration, and pressure (P) are the dominant predictors governing IFT in H_2-rich brine

systems. Based on SHAP scores, it is revealed that T is the most significant parameter to control IFT, and we ranked them higher to lower weighty order: T, H_2, P, CO_2 and CH_4.

3.4. Potential Implications for Underground H_2 and CO_2 Storage Systems

The results of this research have significant implications for subsurface hydrogen storage design and operation. The substantial influence of temperature and pressure underscores the necessity of closely monitoring and controlling in situ subsurface thermodynamic conditions, as variations in these parameters can substantially affect interfacial stability and, consequently, storage integrity. The positive relationship between H_2 concentration and interfacial tension suggests that higher hydrogen purity may help maintain interfacial stability, thereby reducing the risk of gas leakage and minimizing displacement of underground formation brines. In contrast, the strong suppressive effect of CO_2 on IFT highlights potential limitations in the use of CO_2 as a cushion gas, as its presence may weaken interfacial stability and compromise long-term containment security. Although CH_4 shows a negligible direct effect on IFT, its strong inverse relationship with H_2 signifies that it can indirectly influence system behavior in mixed-gas storage scenarios and should therefore not be overlooked in underground reservoir design and gas composition management. Overall, interfacial tension in subsurface hydrogen storage is governed by a coupled thermodynamic system in which temperature weakens interfacial cohesion, pressure modifies phase equilibrium and gas solubility, CO_2 strongly reduces IFT through enhanced solubility and interfacial adsorption, and H_2 exhibits a weak or increasing effect due to its low solubility and gas-phase preference. Collectively, these interacting variables define the interfacial free-energy landscape that controls IFT behavior under subsurface reservoir conditions.

It is noteworthy that IFT plays a critical role in the performance and long-term security of underground CO_2 storage systems by governing the interaction between injected CO_2 and formation brine within the porous reservoir. IFT directly influences capillary pressure, wettability, CO_2 displacement, plume migration, and residual trapping. Higher IFT can increase capillary forces and promote the retention of CO_2 within pore spaces, thereby enhancing storage security, whereas lower IFT may facilitate CO_2 migration and alter injectivity. Consequently, accurate characterization of IFT under reservoir temperature, pressure, and diverse salinity conditions is essential for reliable prediction of CO_2 distribution and trapping behavior. Incorporating IFT into reservoir simulations and storage assessments can improve estimates of storage capacity, injection efficiency, leakage risk, and long-term containment while also supporting appropriate site selection, monitoring strategies, and indirect risk management for geological CO_2 storage.

From an operational perspective, the superior performance of the predictive models with XGB and RF indicates that advanced ensemble learning methods are highly effective for accurate prediction of IFT. When combined with SHAP-based interpretability, this framework offers both high predictive accuracy and model transparency, thereby supporting risk-informed decision-making in safety-critical subsurface hydrogen storage applications.

3.5. Study Limitations and Future Research Scopes

Despite the strong predictive performance using limited datasets of the proposed thermodynamics-informed ensemble ML framework, certain limitations remain, which could be addressed in future studies through further investigation of interfacial tension (IFT) in underground hydrogen storage systems. Notably, the ensemble ML framework is trained on a finite dataset derived from available experimental and literature sources, which may not fully capture the complete range of underground reservoir conditions relevant to subsurface hydrogen storage. In particular, extreme thermodynamic pressure and tempera-

ture regimes and highly complex multicomponent gas mixtures may be underrepresented, potentially limiting model generalization. Notably, the limitations of the developed ML models primarily relate to the dataset coverage and applicability range, as their predictive performance may enhance when applied to conditions beyond the compositional, pressure, temperature, and IFT ranges represented in the training data, limiting their extrapolation capability. In addition, the absence of explicit physical constraints within the ML framework may reduce physical consistency under out-of-distribution conditions; therefore, the models should be applied within the validated parameter ranges and supported by additional experimental data and physics-informed ML approaches for broader applications.

Future research should explore the integration of physics-informed neural networks with big datasets to further enhance the proposed ML framework by embedding thermodynamic constraints directly into the learning process, thereby ensuring that predictions remain consistent with established physical laws. Such hybrid approaches have the potential to strengthen both model generalization and physical interpretability. Collectively, these prospective developments would provide a more robust and physically grounded foundation for predictive IFT modeling in underground hydrogen and carbon storage systems. In addition, although the predictive framework integrates key thermodynamic variables (T, P, CO_2, H_2, and CH_4), it does not explicitly account for additional environments with physicochemical effects such as salinity variations, ionic strength, p^H dependent speciation or mineral–fluid interactions, all potential sources that can influence interfacial behavior in real geological formations. Finally, the development of hybrid frameworks that integrate molecular dynamic simulations, equation-of-state modeling, and hybrid deep learning techniques could offer a more physically grounded and transferable approach for predicting IFT in subsurface energy storage systems.

Future work should incorporate established empirical correlations, Parachor methods, and equation of state-based IFT models under consistent thermodynamic conditions using the same dataset. Such comparisons would help quantify the relative predictive performance, generalizability, and practical applicability of the proposed framework for IFT.

4. Conclusions

In this study, experimental data-driven intelligent models are utilized by integration of ensemble machine learning and thermodynamic conditions for predicting interfacial tension (IFT) in subsurface hydrogen storage systems. Five ensemble machine learning (EML) models, extreme gradient boosting (XGB), random forest (RF), decision tree (DT), adaptive boosting (ADB) and light gradient boosting (LGB), are trained and tested, while optimized models are evaluated using statistical indices, including coefficient of determination (R^2), and mean absolute error (MAE). The key outcomes derived from the research findings are concluded as follows:

- The XGB is the more effective and generalized algorithm, and it exhibited a high accuracy of R^2 (83%) and least error of MAE (3.43 $mN \cdot m^{-1}$) for the testing phase compared to other EML models while predicting IFT.
- To capture nonlinear relationships with IFT, SHAP-based interpretability consistently identified temperature (T), H_2 concentration, and pressure (P) as the most influential variables governing IFT behavior in H_2-rich brine systems. These thermodynamic variables have ranked them higher to lower weighty order: T, H_2, P, CO_2 and CH_4.
- Changes in H_2, CO_2, and other thermodynamic variables may alter IFT for subsurface storage systems and, consequently, influence interfacial and capillary behavior at the storage fluid–formation fluid interface.
- The integration of intelligent framework with SHAP analysis provided clear interpretability of model predictions, confirming the consistency of feature importance

across different machine learning algorithms. These findings emphasize that both thermodynamic conditions and gas composition jointly control interfacial behavior, which is critical indirectly for the safe and efficient design of underground hydrogen storage systems.

Overall, the proposed framework demonstrates strong potential as a reliable and interpretable tool for predicting IFT under varying reservoir conditions. The combination of high predictive accuracy and explainability makes it particularly suitable for risk-informed decision-making in subsurface storage applications. Future work should focus on expanding the dataset, incorporating additional geochemical variables, and developing physics-informed hybrid computational models to further enhance generalization and physical consistency. Notably, for future studies, direct assessment of underground storage security should incorporate additional analyses of capillary entry pressure, wettability, relative permeability, multiphase fluid flow behavior, geomechanical integrity, and long-term fluid migration, supported by appropriate experimental and field-scale validation.

Author Contributions: Conceptualization, M.I.M.; methodology, M.I.M. and M.M.A.; software, M.M.A.; validation, M.M.A.; formal analysis, M.I.M.; investigation, M.I.M.; resources, T.W.; data curation, M.I.M. and M.M.A.; writing—original draft preparation, M.I.M.; writing—review and editing, M.M.A., A.E. and T.W.; visualization, A.E.; funding acquisition, T.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Data Availability Statement: The processed data presented in this study are available on request from the corresponding author, subject to privacy or ethical restrictions.

Acknowledgments: The corresponding author would like to thank the Department of Petroleum and Mining Engineering at Chittagong University of Engineering & Technology (CUET), Bangladesh, for providing the research facilities.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAPE	Average absolute relative error
ADB	Adaptive boosting
AI	Artificial intelligence
DNN	Deep neural network
DT	Decision tree
EML	Ensemble machine learning
IFT	Interfacial tension
LGB	Light gradient boosting
MAE	Mean absolute error
MD	Molecular dynamic
P	Pressure
RF	Random forest
RMSE	Root mean square error
SHAP	Shapley additive explanations
STD	Standard deviation
SVR	Support vector regression
T	Temperature
XGB	Extreme gradient boosting

References

1. Chalbaud, C.A.; Robin, M.; Lombard, J.-M. Brine/CO_2 Interfacial Properties and Effects on CO_2 Storage in Deep Saline Aquifers. *Oil Gas. Sci. Technol.* **2010**, *65*, 541–555. [CrossRef]
2. Honari, A.; Liu, K.; Graham, B.F. The effects of reservoir brine compositions on interfacial tension and reservoir wettability under reservoir conditions. *APPEA. J.* **2010**, *50*, 735. [CrossRef]
3. Hosseini, M.H.; Leonenko, Y. Hydrogen-brine interfacial tension at subsurface conditions: Implication for hydrogen geo-storage. *Proceedings* **2023**, *2023*, 1–5. [CrossRef]
4. Alanazi, A.; Yekeen, N.; Ali, M.; Ali, M.; Abu-Mahfouz, I.S.; Keshavarz, A.; Iglauer, S.; Hoteit, H. 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. [CrossRef]
5. Esfandyari, H.; Hosseini, M.H.; Mohamed Ali, M.F.B. Assessment of the interfacial properties of various mineral/hydrogen/water systems. *J. Energy Storage* **2023**, *60*, 106637. [CrossRef]
6. Doan, Q.T.; Keshavarz, A.; Miranda, C.R. A prediction of interfacial tension by using molecular dynamics simulation: A study on effects of cushion gas (CO_2, N_2 and CH_4) for Underground Hydrogen Storage. *Int. J. Hydrogen Energy* **2023**, *50*, 1607–1615. [CrossRef]
7. Tariq, Z.; Ali, M.; Yekeen, N. Enhancing wettability prediction in the presence of organics for hydrogen geo-storage through data-driven machine learning modeling of rock/H_2/brine systems. *Fuel* **2023**, *354*, 129354. [CrossRef]
8. Ali, M.; Hoteit, H. The Impact of Salinity on H_2/brine Interfacial Tension at Natural Reservoir Conditions: Implications for Hydrogen Geo-Storage. In Proceedings of the SPE International Health, Safety, Environment and Sustainability Conference and Exhibition, Abu Dhabi, United Arab Emirates, 10–12 September 2024. [CrossRef]
9. Esfandyari, H.; Safari, A.; Hassanpouryouzband, A. Interfacial Tension Dynamics in H_2+CO_2/Aqueous Systems: Unlocking Insights for Underground Hydrogen Storage and CO_2 Cushioning. In Proceedings of the SPE Proceedings, Dubai City, United Arab Emirates, 21–23 April 2025. [CrossRef]
10. Abdel-Azeim, S.; Al-Yaseri, A.; Norrman, K. Wettability of Caprock–H_2–Water: Insights from Molecular Dynamic Simulations and Sessile-Drop Experiment. *Energy Fuels* **2023**, *37*, 19348–19356. [CrossRef]
11. Chang, Q.; Dempsey, D.; Zhang, L. Molecular dynamics insights into gas-water interfacial tension: Optimizing hydrogen storage in subsurface conditions. *Int. J. Hydrogen Energy* **2024**, *64*, 896–905. [CrossRef]
12. Kalati, S.S.; Khiabani, N.P.; Ayatollahi, S. Molecular dynamics simulation of hydrogen diffusion into brine: Implications for underground hydrogen storage. *Int. J. Hydrogen Energy* **2024**, *53*, 17–28. [CrossRef]
13. Behnamnia, M.; Mozafari, N.; Monfared, A.D. 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. [CrossRef]
14. Janjua, A.N.; Patil, S.; Kamal, M.S. Machine Learning-Driven Prediction of Density and H_2-Brine Interfacial Tension: Implications for Underground Hydrogen Storage. In Proceedings of the ADIPEC, Abu Dhabi, United Arab Emirates, 4–7 November 2024. [CrossRef]
15. Pan, B.; Song, T.; Yin, X.; Jiang, Y.; Yue, M.; Hoteit, H.; Mahani, H.; Iglauer, S. A Machine Learning Based Framework for Brine-Gas Interfacial Tension Prediction: Implications for H_2, CH_4 and CO_2 Geo-Storage. In Proceedings of the SPE Gas & Oil Technology Showcase and Conference, Dubai, United Arab Emirates, 7–9 May 2024. [CrossRef]
16. Turkson, J.N.; Md Yusof, M.A.; Tackie-Otoo, B.N. Application of Ensemble Learning Paradigms in Predicting Interfacial Tension of H_2/Cushion Gas Systems and the Implications on Subsurface H_2 Storage. In Proceedings of the IPTC, Kuala Lumpur, Malaysia, 18–20 February 2024. [CrossRef]
17. Hosseini, M.; Leonenko, Y. Prediction of hydrogen−brine interfacial tension at subsurface conditions: Implications for hydrogen geo-storage. *Int. J. Hydrogen Energy* **2024**, *58*, 485–494. [CrossRef]
18. Ibrahim, A.F. A Comprehensive AI-Model for H_2-Brine Interfacial Tension Prediction: Implications for Underground Hydrogen Storage. In Proceedings of the SPE Gas & Oil Technology Showcase and Conference, Manama, Bahrain, 16–18 September 2025. [CrossRef]
19. Ahmadi, M. Physics-informed neural modeling of interfacial tension in hydrogen-rich systems using attention-based learning. *Int. J. Hydrogen Energy* **2025**, *163*, 150829. [CrossRef]
20. Sofian, M.; Abdulraheem, A.; Al Shafloot, T. Machine Learning Approach for Modeling Surface Tension Between H_2-CH_4 Gas Mixture and Brine: Implication for Underground Hydrogen Storage in Partially Depleted Gas Reservoirs. In Proceedings of the SPE Middle East Oil, Gas and Geosciences Show, Manama, Bahrain, 16–18 September 2025. [CrossRef]
21. Omrani, S.; Ghasemi, M.; Singh, M. Interfacial Tension–Temperature–Pressure–Salinity Relationship for the Hydrogen–Brine System Under Reservoir Conditions: Integration of Molecular Dynamics and Machine Learning. *Langmuir* **2023**, *39*, 12680–12691. [CrossRef]

22. Zhang, X.; Li, Q.; Simon, M. Effect of H_2S content on relative permeability and capillary pressure characteristics of acid gas/brine/rock systems: A review. *J. Rock Mech. Geotech. Eng.* **2022**, *14*, 2003–2033. [CrossRef]
23. Monge, A.; Vayssaire, A. A Review of Some Aspects of the Molecular Hydrogen Transport Behaviour in the Subsurface. *IGS Proc.* **2022**, *2022*, 1–5. [CrossRef]
24. Mouallem, J.; Arif, M.; Raza, A. Critical review and meta-analysis of the interfacial tension of CO_2-brine and H_2-brine systems: Implications for CO_2 and H_2 geo-storage. *Fuel* **2023**, *356*, 129575. [CrossRef]
25. Gbadamosi, A.O.; Muhammed, N.; Patil, S. Underground hydrogen storage: A critical assessment of fluid-fluid and fluid-rock interactions. *J. Energy Storage* **2023**, *72*, 108473. [CrossRef]
26. Shi, Y.; Guo, Y.; Tai, N. A Critical Review of Underground Hydrogen Storage in Porous Media: Focus on Fluid–Fluid and Rock–Fluid Interactions. *Energy Fuels* **2025**, *39*, 17083–17111. [CrossRef]
27. Kohzadvand, K.; Kouhi, M.M.; Barati, A. Prediction of interfacial wetting behavior of H_2/mineral/brine; implications for H_2 geo-storage. *J. Energy Storage* **2023**, *72*, 108567. [CrossRef]
28. Sagandykova, D.; Kalam, S.; Pourafshary, P. Data-Driven Modeling to Estimate Wettability in Hydrogen-Brine-Mineral Systems: Insights for Underground Hydrogen Storage. In Proceedings of the SPE Proceedings, Manama, Bahrain, 16 September 2025. [CrossRef]
29. Kalam, S.; Khan, M.R.; Arif, M. Smart Computational Algorithms for the Prediction of Interfacial Tension between Water and Hydrogen–Insights into Underground H_2 Storage. In Proceedings of the IPTC Proceedings, Dhahran, Saudi Arabia, 12 February 2024. [CrossRef]
30. Xie, M.; Zhang, M.; Jin, Z. Machine Learning-Based Interfacial Tension Equations for (H_2 + CO_2)-Water/Brine Systems Over a Wide Range of Temperature and Pressure. *Langmuir* **2024**, *40*, 5369–5377. [CrossRef]
31. Chow, Y.F.; Maitland, G.C.; Trusler, J.M. Interfacial tensions of (H_2O+ H_2) and (H_2O+ CO_2+ H_2) systems at temperatures of (298–448) K and pressures up to 45 MPa. *Fluid Phase Equilibria* **2018**, *475* , 37–44, *Erratum To Fluid Phase Equilibria* **2020**, *503*, 112315. [CrossRef]
32. Al-Mukainah, H.; Al-Yaseri, A.; Yekeen, N.; Al Hamad, J.; Mahmoud, M. Wettability of shale–brine–H_2 system and H_2-brine interfacial tension for assessment of the sealing capacities of shale formations during underground hydrogen storage. *Energy Rep.* **2022**, *8*, 8830–8843. [CrossRef]
33. Higgs, S.; Da Wang, Y.; Sun, C.; Ennis-King, J.; Jackson, S.J.; Armstrong, R.T.; Mostaghimi, P. In-situ hydrogen wettability characterisation for underground hydrogen storage. *Int. J. Hydrogen Energy* **2022**, *47*, 13062–13075. [CrossRef]
34. Hosseini, M.H.; Mohamed Ali, M.F.B.; Fahimpour, J. Basalt-H_2-brine wettability at geo-storage conditions: Implication for hydrogen storage in basaltic formations. *J. Energy Storage* **2022**, *52*, 104745. [CrossRef]
35. Hosseini, M.; Fahimpour, J.; Ali, M.; Keshavarz, A.; Iglauer, S. H_2- brine interfacial tension as a function of salinity, temperature, and pressure; implications for hydrogen geo-storage. *J. Pet. Sci. Eng.* **2022**, *213*, 110441. [CrossRef]
36. 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*, 34963–34975. [CrossRef]
37. Isfehani, Z.D.; Sheidaie, A.; Hosseini, M.; Fahimpour, J.; Iglauer, S.; Keshavarz, A. Interfacial tensions of (brine+ H_2+ CO_2) systems at gas geo-storage conditions. *J. Mol. Liq.* **2023**, *374*, 121279. [CrossRef]
38. Miah, M.I.; Abir, M.A.N.; Shuvo, M.A.I. Improved model development and feature ranking for rock permeability prediction by coupling petrophysical log data and ensemble machine learning techniques. *Earth Sci. Inform.* **2025**, *18*, 93. [CrossRef]
39. Miah, M.I.; Rahaman, M.S.; Wiens, T. Enhancing Rock Strength Prediction and Features Selection by Coupling Well Log Data and Deep Learning Approaches in Reservoir Geomechanics. *Int. J. Numer. Anal. Methods Geomech.* **2026**, *50*, 4724–4738. [CrossRef]
40. Miah, M.I.; Elghoul, A.; Butt, S.D.; Wiens, T. Effects of petrophysical parameters on sedimentary rock strength prediction: Implications of machine learning approaches. *Appl. Sci.* **2025**, *15*, 9158. [CrossRef]
41. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]
42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768 - 4777. [CrossRef]
43. Miah, M.I.; Alam, M.M.; Wiens, T. Explainable AI-Based Geophysical Logging Parameters Selection for Enhancing Rock Brittleness Prediction. *SciEn Conf. Ser. Eng.* **2026**, *4*, 549–554. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.