

Toward Estimating CO₂ Solubility in Pure Water and Brine Using Cascade Forward Neural Network and Generalized Regression Neural Network: Application to CO₂ Dissolution Trapping in Saline Aquifers

Published as part of ACS Omega virtual special issue “CO₂ Geostorage”.

Xinyuan Zou, Yingting Zhu,* Jing Lv, Yuchi Zhou, Bin Ding, Weidong Liu, Kai Xiao, and Qun Zhang*

Cite This: *ACS Omega* 2024, 9, 4705–4720

Read Online

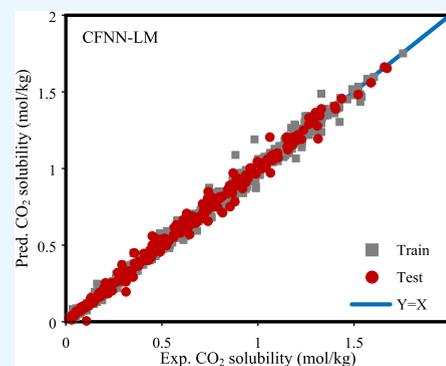
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Predicting carbon dioxide (CO₂) solubility in water and brine is crucial for understanding carbon capture and storage (CCS) processes. Accurate solubility predictions inform the feasibility and effectiveness of CO₂ dissolution trapping, a key mechanism in carbon sequestration in saline aquifers. In this work, a comprehensive data set comprising 1278 experimental solubility data points for CO₂–brine systems was assembled, encompassing diverse operating conditions. These data encompassed brines containing six different salts: NaCl, KCl, NaHCO₃, CaCl₂, MgCl₂, and Na₂SO₄. Also, this databank encompassed temperature spanning from 273.15 to 453.15 K and a pressure range spanning 0.06–100 MPa. To model this solubility databank, cascade forward neural network (CFNN) and generalized regression neural network (GRNN) were employed. Furthermore, three optimization algorithms, namely, Bayesian Regularization (BR), Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton, and Levenberg–Marquardt (LM), were applied to enhance the performance of the CFNN models. The CFNN-LM model showcased average absolute percent relative error (AAPRE) values of 5.37% for the overall data set, 5.26% for the training subset, and 5.85% for the testing subset. Overall, the CFNN-LM model stands out as the most accurate among the models crafted in this study, boasting the highest overall R² value of 0.9949 among the other models. Based on sensitivity analysis, pressure exerts the most significant influence and stands as the sole parameter with a positive impact on CO₂ solubility in brine. Conversely, temperature and the concentration of all six salts considered in the model exhibited a negative impact. All salts exert a negative impact on CO₂ solubility due to their salting-out effect, with varying degrees of influence. The salting-out effects of the salts can be ranked as follows: from the most pronounced to the least: MgCl₂ > CaCl₂ > NaCl > KCl > NaHCO₃ > Na₂SO₄. By employing the leverage approach, only a few instances of potential suspected and out-of-leverage data were found. The relatively low count of identified potential suspected and out-of-leverage data, given the expansive solubility database, underscores the reliability and accuracy of both the data set and the CFNN-LM model's performance in this survey.



1. INTRODUCTION

Despite the worldwide endeavor to curb carbon emissions in the wake of the “Paris Treaty,” the greenhouse gas emissions persist at unsustainable levels.^{1,2} Due to its profound impact on global climate change, the substantial increase in carbon dioxide (CO₂) concentration within the Earth's atmosphere is a matter of critical concern.³ The mitigation of CO₂ emissions is seen as a promising endeavor through the utilization of carbon capture and storage (CCS) techniques.^{4–6} Reliable sites for global CO₂ storage include underground geological formations like deep saline aquifers, depleted oil and gas reservoirs, and unmineable coal seams.^{7–12} Among them, deep saline aquifers offer the largest CO₂ storage capacity.¹³ CCS in saline aquifers primarily relies on several main mechanisms including dissolution (solubility) trapping, mineral trapping,

and capillary trapping. Capillary trapping physically immobilizes CO₂ in rock pores while dissolution causes it to dissolve into brine, reducing its buoyancy. Mineral trapping involves CO₂ reacting with minerals for long-term storage as stable carbonates, collectively ensuring secure carbon sequestration.^{14–16} In geological carbon storage, the injection of CO₂ results in its dissolution into the reservoir brine. This process elevates the brine's density, creating a fluid density gradient

Received: October 11, 2023
Revised: December 28, 2023
Accepted: January 4, 2024
Published: January 17, 2024



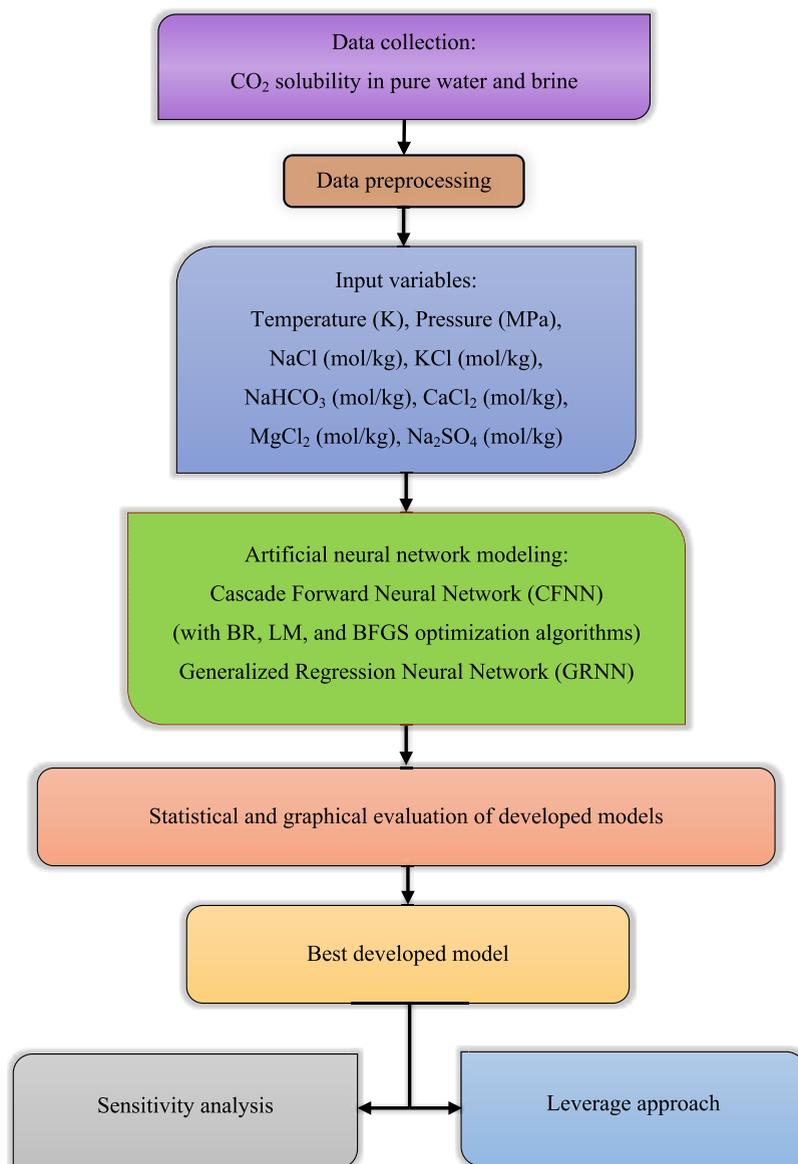


Figure 1. Research steps to model CO₂ solubility in brine.

within the formation. These gradients trigger convective mixing, accelerating the dissolution of CO₂ and effectively leading to significant CO₂ solubility trapping.^{15–17} The significance of the dissolution trapping mechanism in the success of geological carbon storage is notable. Additionally, the interplay of CO₂ and reservoir brine solubility has a profound impact on long-term storage security, as well as the intricate physical and chemical interactions among minerals and fluids within the reservoir.¹⁸ In order to accurately assess these impacts and ensure efficient carbon storage, it is imperative to possess a dependable model for forecasting the solubility of CO₂ in reservoir brines within the context of geological carbon storage conditions.

Given the crucial relevance of CO₂ solubility in both pure water and brine, numerous extensive studies have been conducted to acquire valuable insights into CO₂ dissolution in aqueous solutions. These investigations entail extensive experimental measurements of CO₂ solubility in pure water^{19–25} and brine,^{26–36} considering a range of geological pressure and temperature conditions, as well as the presence of

various salts commonly found in brine compositions. Nevertheless, due to the multitude of factors influencing solubility under real conditions, such as the presence of diverse salts in the formation of brine and variations in pressure and temperature across different geological formations, researchers are primarily focused on the development of models or correlations to estimate solubility data. A literature survey reveals that the majority of the models proposed for CO₂–brine mutual solubility rely on equations of state and activity coefficient models.^{24,25,33,37–45} The primary drawback of these models lies in their requirement for precise mixing rules, adjustable parameters, intricate iterative calculations, relatively high margin of error, and ultimately, their limitation in encompassing various parameters and delivering a comprehensive modeling approach.⁴⁶ Hence, in recent years, researchers have shifted their focus toward utilizing artificial intelligence techniques to model CO₂ solubility in aqueous solutions, recognizing them as potent tools for modeling highly intricate systems characterized by a significant degree of nonlinearity. Ali Ahmadi and Ahmadi⁴⁷ employed 54 data

Table 1. Statistical Summary of the Gathered Database

	NaCl (mol/kg)	KCl (mol/kg)	NaHCO ₃ (mol/kg)	CaCl ₂ (mol/kg)	MgCl ₂ (mol/kg)	Na ₂ SO ₄ (mol/kg)	temperature (K)	pressure (MPa)	CO ₂ solubility (mol/kg)
mean	0.94	0.26	0.03	0.14	0.16	0.02	339.27	10.07	0.58
median	0.00	0.00	0.00	0.00	0.00	0.00	323.15	6.51	0.54
mode	0.00	0.00	0.00	0.00	0.00	0.00	323.15	15.00	0.15
SD	1.68	0.88	0.14	0.68	0.71	0.15	44.29	11.34	0.41
kurtosis	1.36	11.09	34.08	35.37	28.66	119.33	-0.28	9.19	-0.91
skewness	1.66	3.49	5.85	5.80	5.24	10.68	0.78	2.31	0.39
minimum	0.00	0.00	0.00	0.00	0.00	0.00	273.15	0.06	0.01
maximum	6.00	4.50	1.00	5.21	5.00	2.00	453.15	100.00	1.76
count	1278	1278	1278	1278	1278	1278	1278	1278	1278
variable status	input	input	input	input	input	input	input	input	target

points related to CO₂–brine solubility in order to develop an LSSVM model for estimating solubility data. Mohammadian et al.⁴⁸ conducted experimental work to determine CO₂ solubility in low-salinity NaCl brines and then employed the extreme learning machine to estimate solubility data based on their findings. Menad et al.⁴⁹ employed a databank consisting of 570 data points and a radial basis function (RBF) neural network to model the solubility of CO₂ in NaCl brine. Hemmati-Sarapardeh et al.⁵⁰ employed the Least-Squares Support Vector Machine (LSSVM) method to model CO₂ dissolution in pure water under high-temperature and high-pressure states, utilizing a data set comprising 785 experimental data points. Sayahi et al.⁵¹ employed hybrid intelligence strategies to predict CO₂ solubility in aqueous brine systems. Mohammadian et al.⁵² developed a correlation for estimating CO₂ solubility in NaCl brine using a genetic algorithm and a data set of 164 data points. Nakhaei-Kohani et al.⁵³ gathered a data set of 289 data points and applied gradient boosting techniques to model CO₂–N₂ gas mixtures in aqueous electrolyte systems. Their findings demonstrated that intelligent models outperformed equations of state in terms of precision. Lv et al.⁵⁴ utilized a similar database as in Nakhaei-Kohani et al.'s study⁵³ to formulate correlations for estimating the solubility of CO₂–N₂ gas mixtures in brine. They employed the group method of data handling and gene expression programming for this purpose. Ratnakar et al.⁵⁵ introduced several machine learning models to estimate CO₂ solubility in both pure water, utilizing 137 instances, and brine solutions. Khoshraftar and Ghemei⁵⁶ collected 240 data points related to CO₂ solubility in pure water and employed response surface methodology and deep learning to develop a model for this solubility phenomenon. In 2022, Vo Thanh et al.⁵⁷ investigated CO₂ storage performance in underground saline aquifers using a databank consisting of 1509 data points. Their results showed that extreme gradient boosting (XGBoost model) achieved a root-mean-square error (RMSE) of 0.0041. In addition, Alqaness et al.⁵⁸ estimated CO₂ trapping indexes in deep saline aquifers using a Long–Short–Term–Memory (LSTM) model. The developed swarm intelligence method (AOSMA) boosts the prediction capability of the LSTM model. In 2023, Davoodi et al.^{59,60} predicted CO₂ trapping indexes in deep geological formations utilizing LSSVM, RBF, general regression neural network (GRNN), and convolutional neural network (CNN). Based on their findings, hybrid machine learning (HML) models outperform conventional machine learning (ML) models in estimating CO₂ trapping indexes. The literature review underscores the sustained research focus on CO₂ solubility in saline solutions over recent decades, with

numerous modeling approaches being developed in recent years. However, these models have primarily centered on NaCl brine and relied on limited data sets. Researchers have endeavored to broaden the models' operational scope over time, yet there remains a substantial gap in achieving a comprehensive model capable of elucidating the impact of various salts on CO₂ solubility in brine. Such a model would be instrumental in understanding the dissolution trapping mechanism and enhancing the efficiency of geological carbon storage.

In this study, a comprehensive data set comprising 1278 experimental solubility data points for CO₂–brine systems is assembled, encompassing diverse operating conditions. These data encompass brines containing six different salts: NaCl, KCl, NaHCO₃, CaCl₂, MgCl₂, and Na₂SO₄. To model this solubility databank, Cascade Forward Neural Network (CFNN) and Generalized Regression Neural Network (GRNN) are employed. Furthermore, three optimization algorithms, namely, Bayesian Regularization (BR), Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton, and Levenberg–Marquardt (LM), are applied to enhance the performance of the CFNN models. Graphical error analyses and statistical parameters are used to assess the validity of the proposed models. Moreover, sensitivity analysis is conducted to evaluate the relative influence of input parameters on CO₂ solubility as the output. Eventually, both potential suspected data and the applicability scope of the best-proposed model are investigated using the Leverage approach. The research process depicted in Figure 1 outlines the steps followed in this survey to model the solubility of CO₂ in brine.

2. DATA COLLECTION

In this work, an exhaustive collection of experimental data encompassing 1278 data points related to the solubility of CO₂ in both pure water and brine has been compiled from existing published sources.^{20,21,25,26,29–31,33–38,42,61–80} A crucial aspect of this database is its inclusion of brine solutions containing six different salts: NaCl, KCl, NaHCO₃, CaCl₂, MgCl₂, and Na₂SO₄. This stands in contrast to the majority of databases utilized in the literature as stated earlier, which predominantly rely on NaCl brine for modeling purposes. This diversity of salts enhances the realism of the brine compositions, mirroring the complexity of brine found in actual reservoirs. Within this research, the input variables taken into account encompass the concentration of each salt in terms of molality (mol/kg) in brine solution, in addition to the operational conditions of temperature (K) and pressure (MPa). The models in this

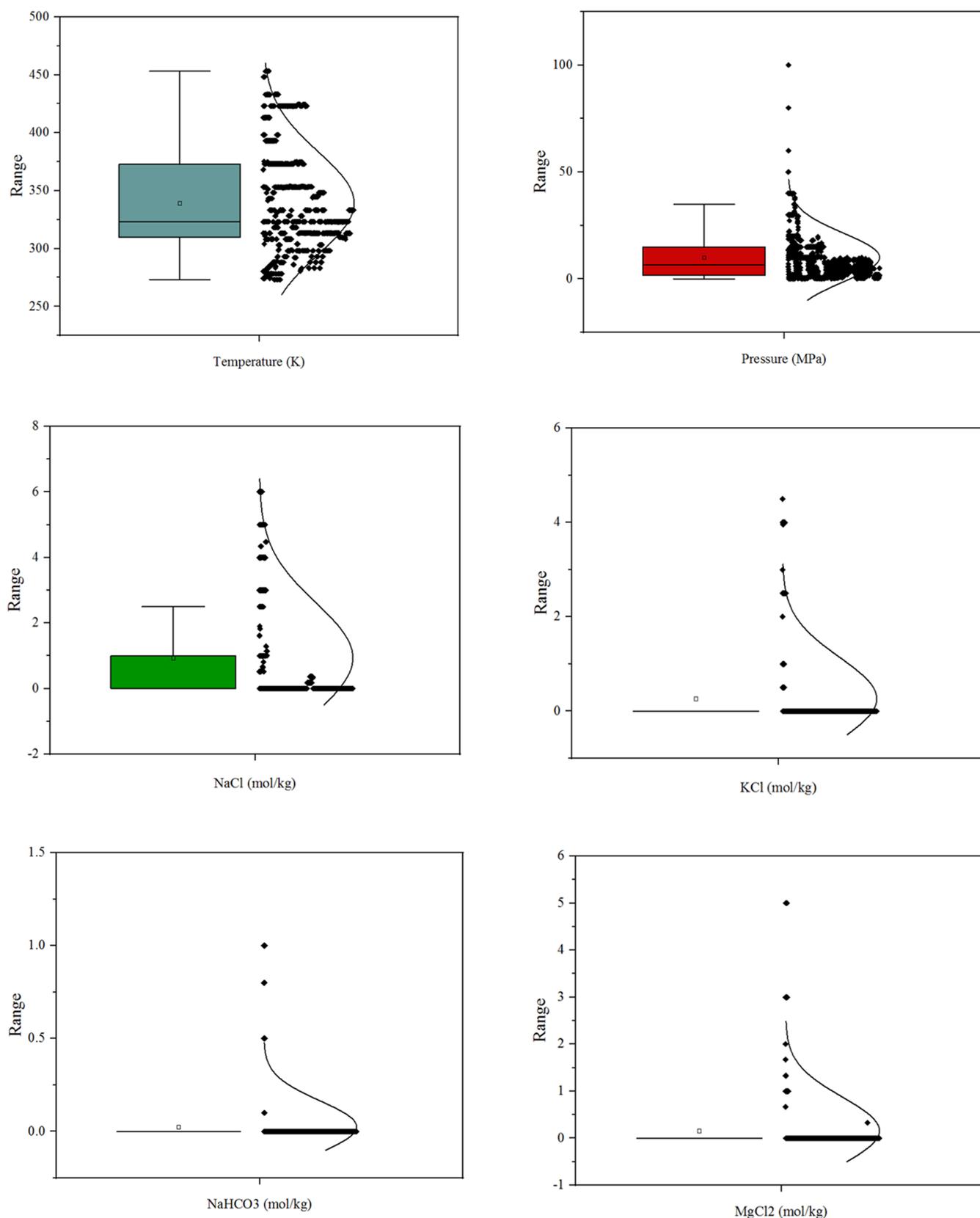


Figure 2. Box plots of all variables existing in the solubility databank.

study yielded as output the solubility of CO_2 (measured in mol/kg) in both pure water and synthetic brine solutions containing NaCl, KCl, NaHCO_3 , CaCl_2 , MgCl_2 , and Na_2SO_4 . Table 1 provides a thorough statistical analysis of the input and

target variables existing in the solubility databank that have been collected in this study. The statistical analysis reveals that the experimental data collected covers a range of CO_2 solubility values at temperatures ranging from 273.15 to

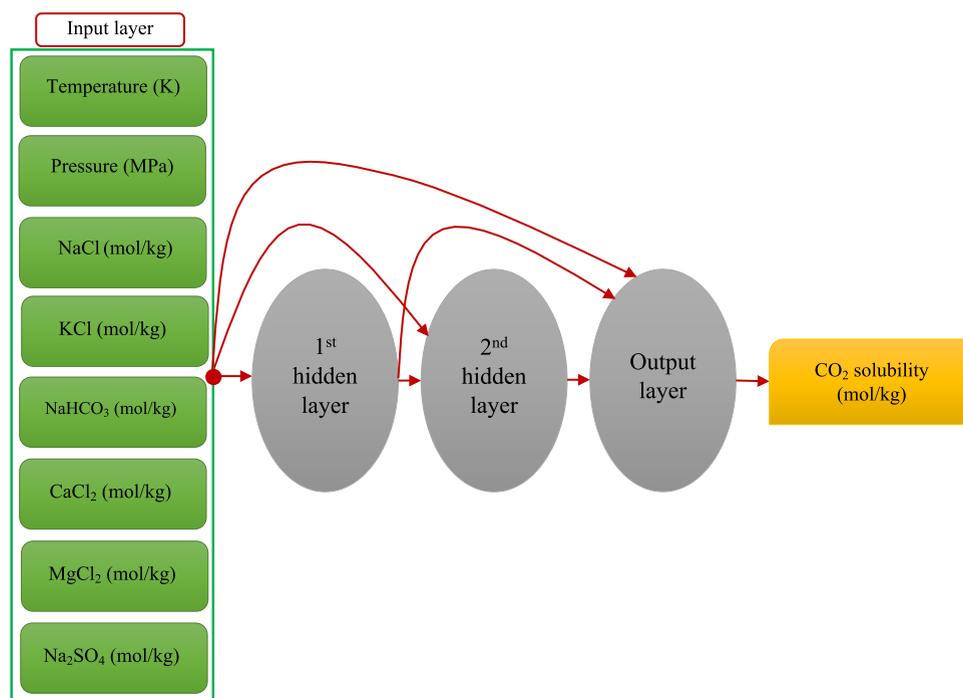


Figure 3. Schematic illustration of the CFNN model developed in this work.

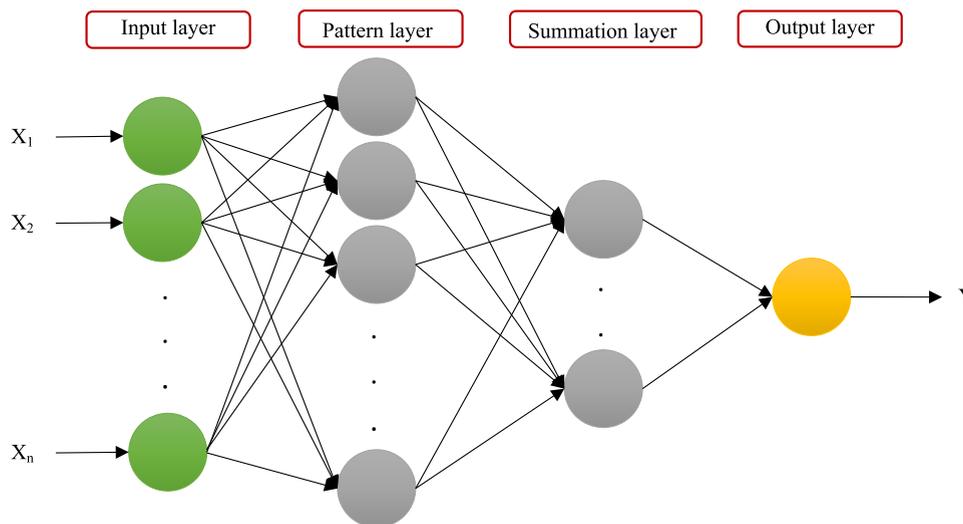


Figure 4. Schematic illustration of the GRNN architecture.

453.15 K and pressures spanning from 0.06 to 100 MPa. Here, a box plot is used to illustrate a data set's five-number summary, which includes the minimum, first quartile, median, third quartile, and maximum values. When creating a box plot, a rectangular box extends from the first quartile to the third quartile, with a horizontal line intersecting the box at the median value. Figure 2 showcases a series of box plots representing all variables contained in the database used in this study. The wide range of input parameters for the models, spanning diverse values, provides a strong foundation for developing a comprehensive predictive correlation for CO₂ solubility in brine. To partition the data for modeling, a random split of the database was employed, resulting in an 80% training set and a 20% test set. We used the data in their original format reported in the literature. In fact, we did not modify the data or select the specific data and trusted the

experimental data reported in the literature. The uncertainty of experimental data may affect our results. In addition, the models are data-driven models and their applicability domain is within the region that these models were developed. Using the data out of this range may result in higher errors.

3. METHODOLOGY

3.1. Models. **3.1.1. Cascade Forward Neural Network (CFNN).** The CFNN⁸¹ is a feed-forward network designed with the idea that increasing the connections between layers can enhance the network's ability to capture the underlying relationships between dependent and independent variables. This back-propagation neural network features a distinct structure compared to traditional feed-forward networks, specifically in terms of the count of connections between each layer. In the CFNN, every layer (excluding the input

layer) establishes connections with all preceding layers through weighted links.⁸² In the CFNN, much like in other feed-forward networks, one can observe the presence of one or multiple interconnected hidden layers and activation functions. These networks can be trained using a variety of training algorithms. Training algorithms aim to find the best weights and biases to minimize prediction errors.⁸³ While CFNN requires fewer hidden neurons for problem-solving, its higher count of weights and biases may lead to slightly slower training in comparison to conventional networks.⁸⁴ Figure 3 depicts a schematic overview of the CFNN architecture, which includes two hidden layers, as well as the connections between the input and output variables. A brief explanation of the optimization techniques, namely, Levenberg–Marquardt (LM), Bayesian regularization (BR), and Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton applied in the present study is presented in the Supporting Information file.

3.1.2. Generalized Regression Neural Network (GRNN). The GRNN concept, pioneered by Specht⁸⁵ in 1991, serves as a type of RBF network characterized by its rapid generalization, adaptable structure, and the absence of an iterative process for parameter optimization. This method relies on kernel regression, a common statistical approach, and can be conceptualized as a normalized RBF neural network with a hidden neuron positioned at the centroid of the training data.^{86,87} The GRNN constructs a function estimate directly from training data, tracing it as a probability density function, without making assumptions about a particular functional form.^{88,89} As depicted in Figure 4, The GRNN architecture comprises four distinct layers: input, pattern, summation, and output. Within the GRNN, the utilization of a normal distribution as a probability density function is a fundamental characteristic, which is formulated as follows⁹⁰

$$Y(X) = \frac{\sum_{i=1}^n Y_i \exp\left(-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right)} \quad (1)$$

Here, σ is the spread parameter, which plays a crucial role and needs to be optimized. Also, Y signifies the predicted output value corresponding to the given input X . The optimization of σ is essential for fine-tuning the GRNN's performance and achieving accurate predictions.⁹¹

4. RESULTS AND DISCUSSION

4.1. Developed Models. This study aimed to model CO₂ solubility in pure water and brine by employing a data set of 1278 data points, considering input variables like pressure, temperature, and salts concentrations. To evaluate model accuracy and reliability, the data set was randomly divided into 80% for training and 20% for testing. The trial-and-error approach was used to fine-tune the number of hidden layers and neurons in each hidden layer for CFNN models. It is also employed to optimize the spread coefficient (σ) for the GRNN model. Table 2 displays the optimized parameters and features of all of the models developed in this study.

4.2. Assessment of Developed Models. An assessment of the proposed models' performance and accuracy in both training and testing phases was conducted through statistical error analysis. This analysis involved the use of various metrics, including root-mean-square error (RMSE), determination coefficient (R^2), standard deviation (SD), average percent

Table 2. Optimized Parameters and Features of All Suggested Models

model	parameter	value
CFNN-LM	hidden layer size	[6 8]
	transfer function	tansig–logsig–purelin
CFNN-BR	hidden layer size	[6 8]
	transfer function	tansig–logsig–purelin
CFNN-BFG	hidden layer size	[6 8]
	transfer function	tansig–logsig–purelin
GRNN	spread	0.0125

relative error (APRE), and average absolute percent relative error (AAPRE). The following formulas represent these statistical criteria⁹²

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{i,\text{exp}} - Y_{i,\text{pred}})^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_{i,\text{exp}} - Y_{i,\text{pred}})^2}{\sum_{i=1}^N (Y_{i,\text{exp}} - \bar{Y}_{\text{exp}})^2} \quad (3)$$

$$\text{SD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{Y_{i,\text{exp}} - Y_{i,\text{pred}}}{Y_{i,\text{exp}}} \right)^2} \quad (4)$$

$$\text{APRE} = \frac{100}{N} \sum_{i=1}^N \left(\frac{Y_{i,\text{exp}} - Y_{i,\text{pred}}}{Y_{i,\text{exp}}} \right) \quad (5)$$

$$\text{AAPRE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{Y_{i,\text{exp}} - Y_{i,\text{pred}}}{Y_{i,\text{exp}}} \right| \quad (6)$$

where, $Y_{i,\text{exp}}$ shows the experimental CO₂ solubility in brine, $Y_{i,\text{pred}}$ is the predicted CO₂ solubility in brine by the suggested models, and N stands for the number of data points.

In Table 3, you can find the calculated statistical errors for the models suggested in this survey. The results in Table 3 highlight that out of the four models developed, the CFNN-LM model consistently exhibits the lowest error values across various metrics, including RMSE, SD, APRE, and AAPRE. More specifically, the CFNN-LM model showcases AAPRE values of 5.37% for the overall data set, 5.26% for the training subset, and 5.85% for the testing subset. Additionally, the R^2 value assesses the model's goodness of fit, where a value of 1 signifies a perfect fit. In this regard, the CFNN-LM model demonstrates outstanding accuracy, boasting the highest overall R^2 value of 0.9949 among the other models. Consequently, following thorough statistical analyses, the CFNN-LM model stands out as the most accurate among the models crafted in this study, anticipating CO₂ solubility in both pure water and brine. Overall, the models can be ranked in order of performance as follows, from best to least: CFNN-LM, CFNN-BR, CFNN-BFGS, and GRNN.

Furthermore, to assess the performance of the proposed models in this survey, graphical error analyses were employed. Figure 5 presents cross-plots of these models used to anticipate CO₂ solubility in brine, plotting model-estimated values against corresponding experimental data. The precision of a model can be gauged by examining the concentration of data points in proximity to the $Y=X$ line. A denser cluster of data points near this line signifies a more accurate model. Notably,

Table 3. Statistical Criteria of the Suggested Models

	status	CFNN-LM	CFNN-BR	CFNN-BFGS	GRNN
APRE, %	train	-0.79	-0.92	-0.14	-8.88
	test	-0.61	-0.81	0.11	-2.56
	total	-0.75	-0.90	-0.09	-7.62
AAPRE, %	train	5.26	5.47	7.22	14.11
	test	5.85	6.27	7.82	17.93
	total	5.37	5.63	7.34	14.87
RMSE	train	0.028	0.027	0.040	0.035
	test	0.034	0.050	0.055	0.099
	total	0.029	0.033	0.043	0.054
SD	train	0.103	0.109	0.131	0.478
	test	0.109	0.119	0.147	0.367
	total	0.104	0.111	0.134	0.458
R ²	train	0.9953	0.9956	0.9903	0.9925
	test	0.9931	0.9846	0.9832	0.9425
	total	0.9949	0.9935	0.9888	0.9824

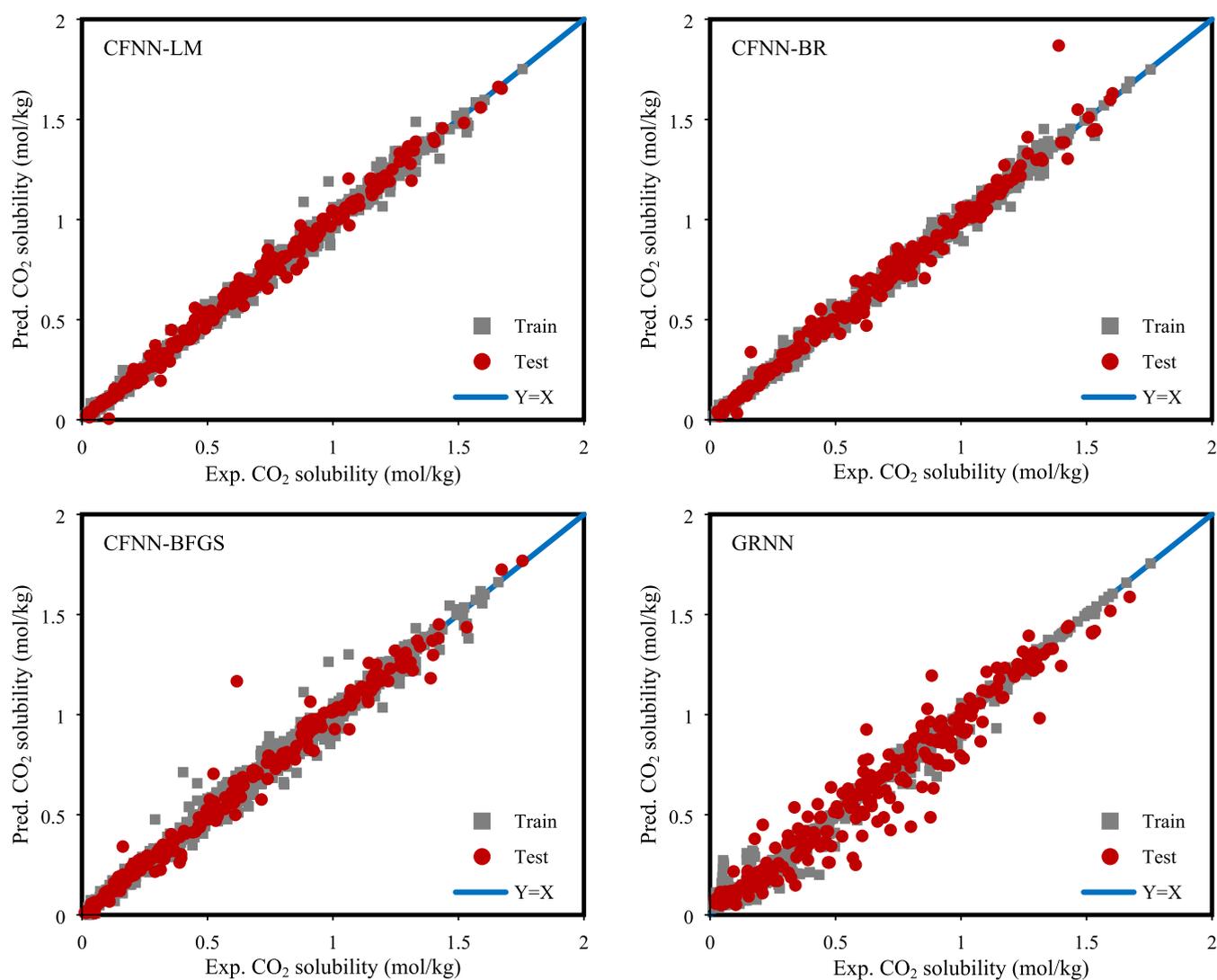


Figure 5. Cross-plots of developed models in the current work.

the close alignment of training and testing data points along the $Y=X$ line indicates that all models excel at accurately predicting experimentally measured CO_2 solubility data. While this holds true for all models, the CFNN-LM model, in particular, stands out for its superior predictive accuracy.

Moving forward, Figure 6 provides a visual representation of the distribution curves depicting relative discrepancies in both data sets of train and test. In this graphic analysis, we focus on relative errors, which are quantified using the subsequent mathematical formula⁹³

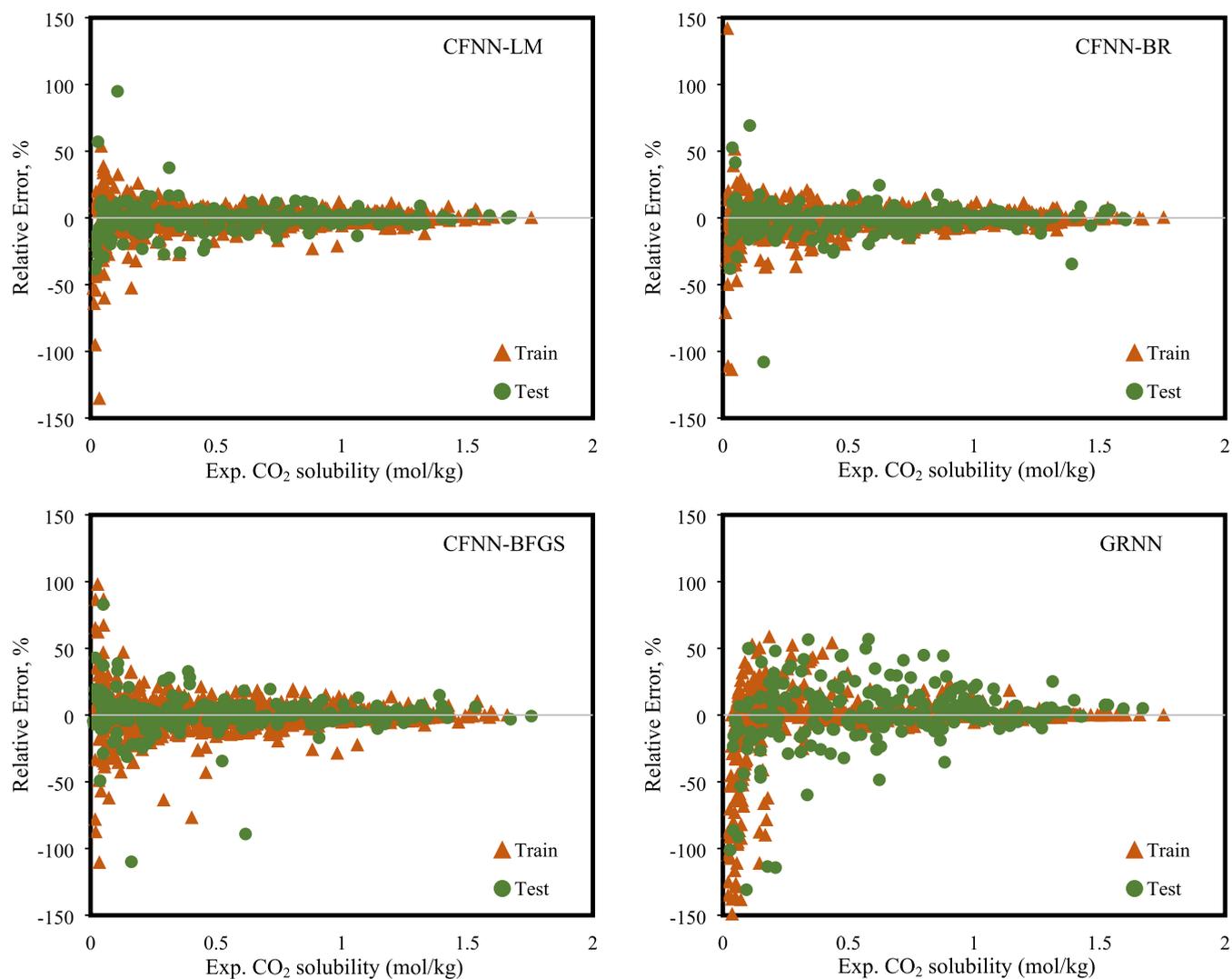


Figure 6. Relative error distribution graphs of developed models.

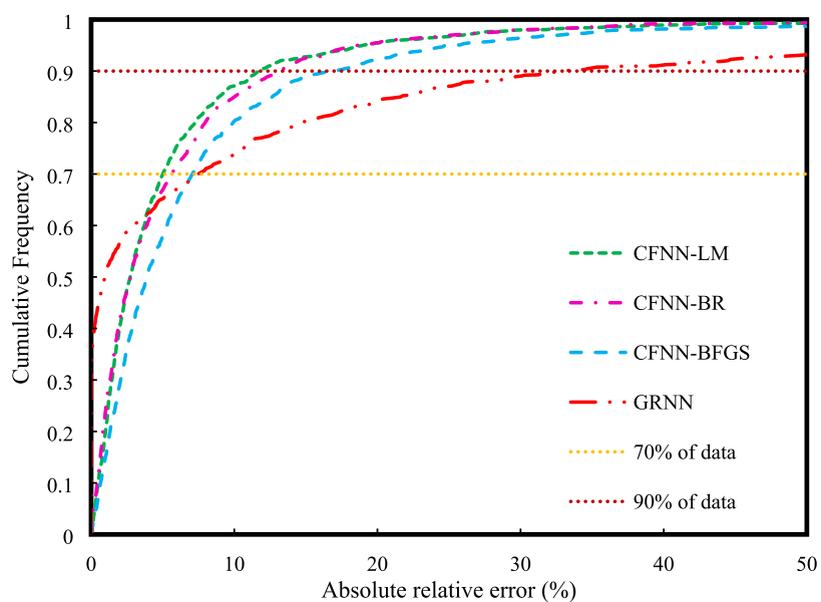


Figure 7. Cumulative frequency graph of developed models.

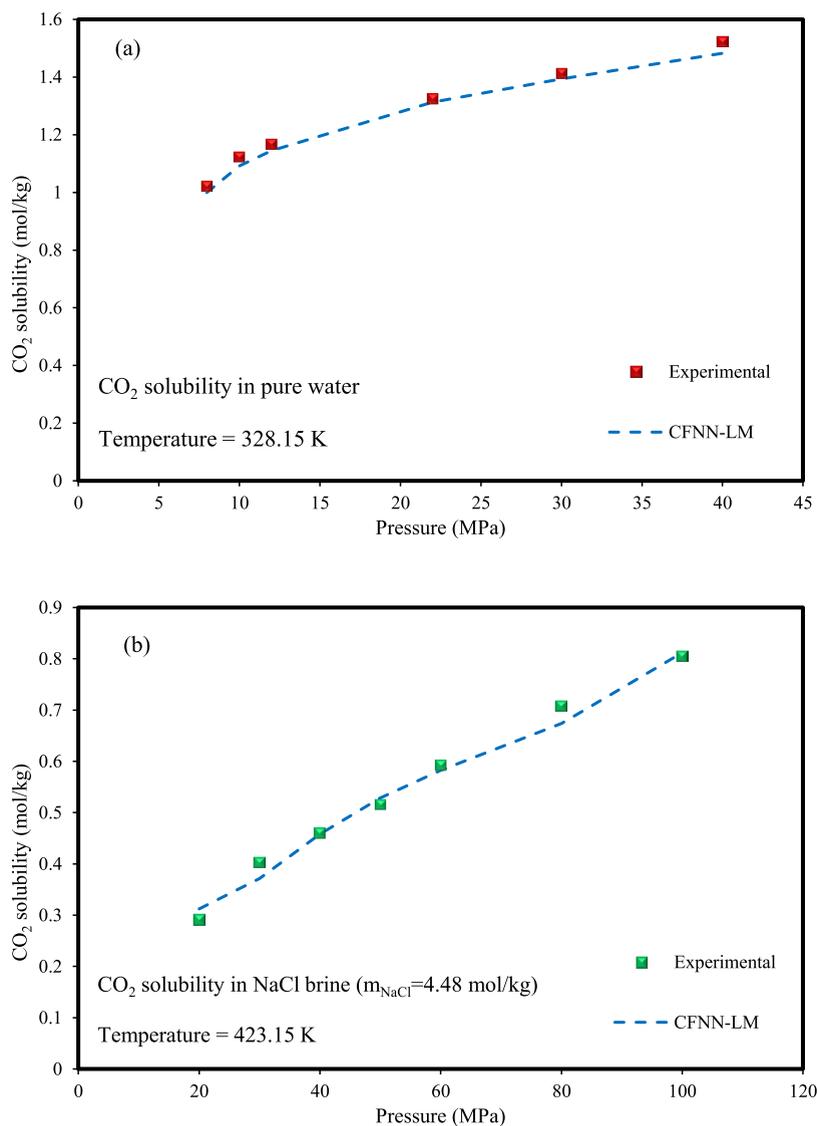


Figure 8. Experimental data on CO₂ solubility in (a) pure water⁷³ and (b) NaCl brine²⁹ versus pressure and at constant temperatures, along with CFNN-LM predictions.

$$\text{relative error} = \left| \frac{Y_{i,\text{exp}} - Y_{i,\text{pred}}}{Y_{i,\text{exp}}} \right| \times 100 \quad i = 1, 2, 3, \dots, n \quad (7)$$

In Figure 6, it becomes evident that a conspicuous absence of error trends was discernible across all models, a significant point within the realm of modeling. Particularly noteworthy is the CFNN-LM model, wherein a substantial proportion of data points elegantly coalesce into compact clusters along the zero-error axis, thereby encapsulating the full spectrum of experimental CO₂ solubility in brine. This underscores a significant alignment between the CFNN-LM model's forecasts and the experimental data set.

Following that, the next graphical analysis entails a cumulative frequency plot illustrating the cumulative frequency of data in relation to the absolute relative error. The calculation of the absolute relative error is based on the formula presented below⁹³

$$\text{absolute relative error} = \left| \frac{Y_{i,e} - Y_{i,p}}{Y_{i,e}} \right| \times 100 \quad i = 1, 2, 3, \dots, n \quad (8)$$

This assessment provides insight into the models' capacity to accurately estimate different portions of the data. When the cumulative frequency is elevated, and the curve closely approaches the Y-axis, it signifies that the models exhibit lower errors when predicting a substantial share of the data set. Figure 7 illustrates the cumulative frequency plots for these models. The CFNN-LM model, notably, excels in predicting around 70% of the data with an absolute relative error below 5%. Furthermore, it achieves accurate predictions for 90% of the data with errors below 12%. However, the remaining models exhibit higher errors for specific portions of the data set when compared to the CFNN-LM model. Therefore, the ELM model outperforms others in forecasting CO₂ solubility data.

4.3. Trend Analysis. Following a comprehensive examination of the models' accuracy through both statistical and graphical analyses, it became evident that the CFNN-LM

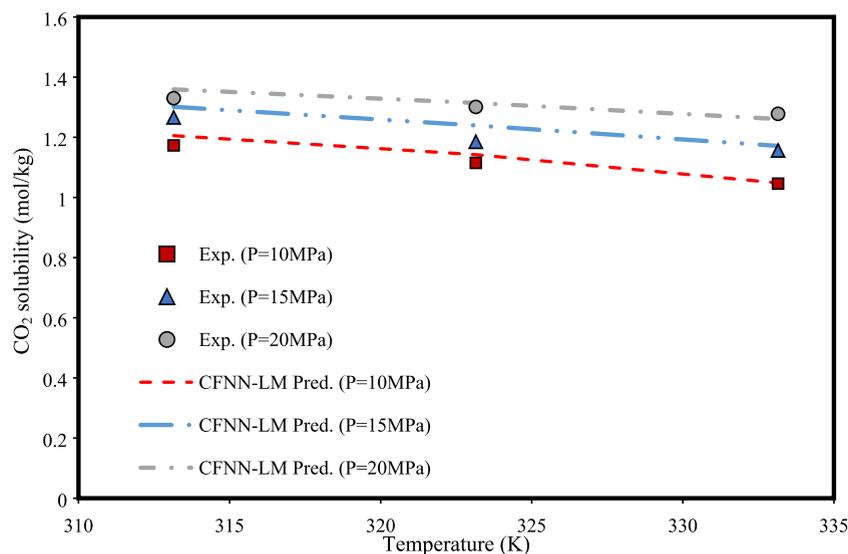


Figure 9. Experimental data³⁵ on CO₂ solubility in pure water at various temperatures and constant pressures, along with CFNN-LM predictions.

model outperforms others in terms of precision. Consequently, this section now centers its attention on exploring changes within the physical phenomenon of CO₂ solubility in pure water and brine, while also evaluating the predictive capacity exhibited by the CFNN-LM model. To start, an assessment of the impact of pressure on CO₂ solubility was conducted in both pure water, as documented in relevant literature experiments,⁷³ and in NaCl brine,²⁹ similarly based on experimental findings reported in the literature. This evaluation was carried out under consistent temperature conditions. As illustrated in Figure 8, there is a notable increase in the solubility of CO₂ in aqueous solutions as the pressure increases. The solubility of CO₂ in both water and brine increases with higher pressure due to Henry's Law.⁹⁴ This law dictates that the concentration of a gas in a liquid is directly proportional to the gas's partial pressure above the liquid, resulting in more CO₂ dissolving into the liquid as the pressure increases. As depicted in Figure 8, CFNN-LM accurately captures the increasing trend in CO₂ solubility data at various pressure levels.

Continuing with trend analysis, this section explores the influence of temperature on CO₂ solubility data. As Figure 9 illustrates, the impact of temperature on CO₂ solubility in pure water is examined across various constant pressures, drawing from experimental findings reported in the literature.³⁵ In this example, it is observed that the solubility of CO₂ in water decreases slightly as the temperature increases while maintaining constant pressure. Once again, the CFNN-LM model demonstrated its ability to predict solubility data with remarkable accuracy. The decrease in the solubility of CO₂ in pure water with increasing temperature is primarily driven by increased kinetic energy, disrupting the equilibrium between the gas and liquid phases and weakening intermolecular interactions. These factors collectively promote the escape of CO₂ molecules from the liquid phase to the gas phase, reducing its solubility as the temperature increases.⁹⁵ Furthermore, it can be elucidated through Le Chatelier's principle that when a gas dissolves in a liquid, which is inherently an exothermic process, an increase in temperature results in a reduction in solubility.⁹⁶ Observing the mentioned trend in most instances, it is worth highlighting that the

literature survey^{73,97} reveals other dramatic trends in CO₂ solubility with temperature. At low pressures ($P \lesssim 20$ MPa), CO₂ solubility shows a gradual decrease as the temperature increases. However, at higher pressures ($P \gtrsim 20$ MPa), the behavior is more complex; initially decreasing with an increase in temperature, it then reverses, showing an increase in solubility as the temperature continues to increase.^{73,97}

Finally, the study delved into the influence of salinity on CO₂ solubility in brine. Figure 10a illustrates CO₂ solubility in NaCl brine with different concentrations, maintaining a constant temperature of 273.15 K and pressures of 10 and 40 MPa, based on experimental findings from the literature.⁷⁸ Additionally, Figure 10b presents CO₂ solubility in MgCl₂ brine with different concentrations, maintaining a constant pressure of 15 MPa and temperatures of 373 and 423 K, as experimentally investigated in the literature.³⁶ In Figure 10, it is evident that across the entire spectrum of temperature and pressure, CO₂ solubility in the solutions experiences a notable decrease as the salt concentration increases. This phenomenon is associated with what is known as the salting-out effect.^{34,98} This effect can be logically explained by the notion that as ions become solvated, a portion of the water becomes inaccessible for the solute, causing it to be effectively salted out from the aqueous phase.⁷³ Again, the CFNN-LM model demonstrated exceptional accuracy in anticipating the solubility trends of CO₂ in brine under various scenarios, aligning closely with the actual trends observed.

4.4. Sensitivity Analysis. The assessment of input variables' impact on the CFNN-LM model's response, specifically CO₂ solubility in brine, involved the use of a specific metric called the relevancy factor (r). A greater r -value assigned to an input parameter signifies its increased significance and influence on CO₂ solubility in brine.^{82,99} The relevancy factor is determined through the application of the following equation^{82,100}

$$r(\ln p_i, Y) = \frac{\sum_{j=1}^n (\ln p_{i,j} - \ln p_{m,i})(Y_j - Y_m)}{\left(\sum_{j=1}^n (\ln p_{i,j} - \ln p_{m,i})^2 \sum_{j=1}^n (Y_j - Y_m)^2 \right)^{0.5}} \quad (9)$$

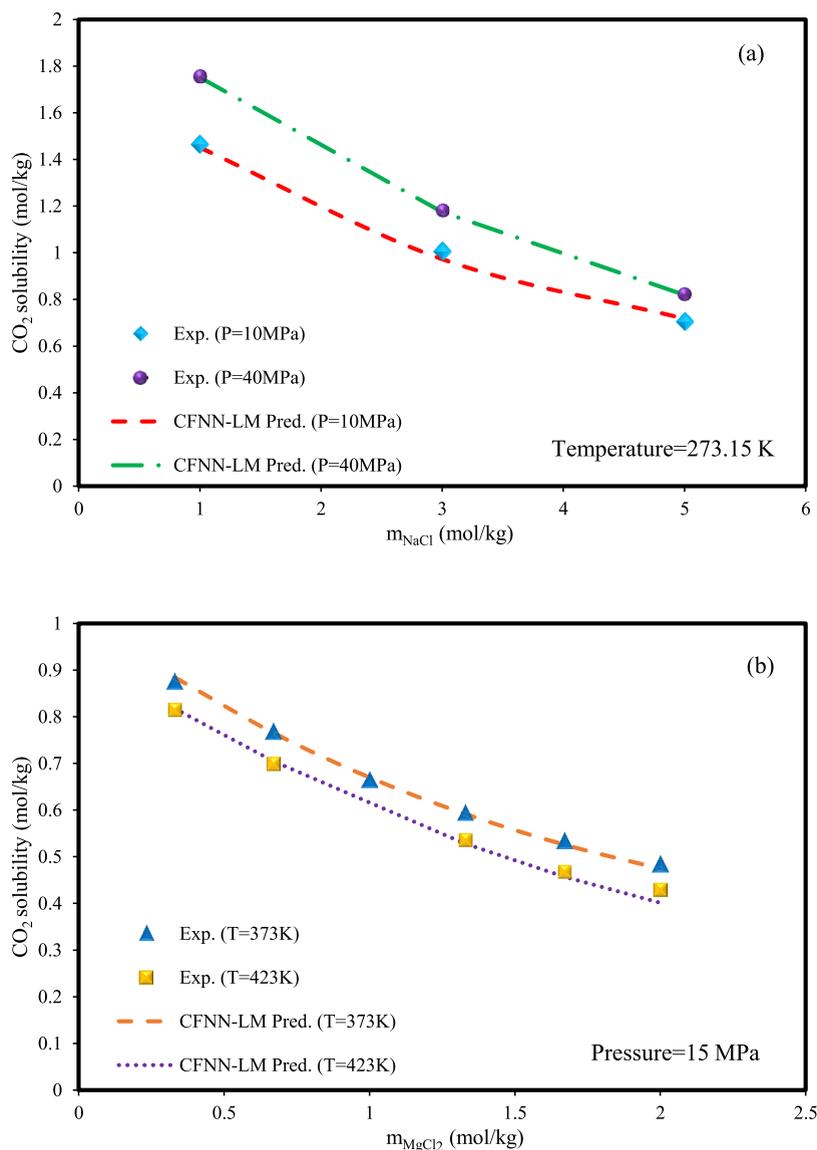


Figure 10. Experimental data on CO₂ solubility in (a) NaCl brine⁷⁸ and (b) MgCl₂ brine³⁶ at various temperatures and constant pressures, along with CFNN-LM predictions.

in this context, $\ln p_{m,i}$ signifies the mean value of the i th input variable, and $\ln p_{i,j}$ denotes the j th value of the same input variable. The variable i may encompass any of the input parameters employed in the modeling process. In addition, Y_m represents the mean of the estimated CO₂ solubility, while Y_j stands for the j th value of the estimated CO₂ solubility.

Figure 11 provides a sensitivity analysis, demonstrating how various inputs impact CO₂ solubility in brine as predicted by the CFNN-LM model. As the results indicate, pressure exerts the most significant influence and stands as the sole parameter with a positive impact on CO₂ solubility in brine. Conversely, temperature and the concentration of all six salts considered in the model exhibited a negative impact. This means that as they increase, CO₂ solubility in brine decreases. Furthermore, it is worth noting that while the temperature generally exhibits a negative impact, as observed in most cases, especially at low pressures, it can demonstrate a different trend at higher pressures as explained earlier in the trend analysis section. All salts exert a negative impact on CO₂ solubility due to their salting-out effect,^{34,73,79} with varying degrees of influence. The

salting-out effects of these salts can be ranked as follows, from the most pronounced to the least: MgCl₂ > CaCl₂ > NaCl > KCl > NaHCO₃ > Na₂SO₄. It seems that salts containing ions with higher valencies tend to have a stronger salting-out effect. For example, divalent ions (like Mg²⁺ and Ca²⁺) can have a more pronounced impact on reducing CO₂ solubility compared to monovalent ions (like Na⁺ and K⁺). However, it is important to note that the specific behavior can vary depending on the nature of the ions, the concentration of the salt, and the conditions of temperature and pressure. Therefore, the effect of the valency of salt on CO₂ solubility in brine is part of a broader set of factors that influence this phenomenon.

4.5. Leverage Approach. This research utilized the leverage approach^{101,102} to assess the applicability scope of the CFNN-LM model and to detect any potentially questionable data points within the CO₂ solubility data set. Within this approach, standardized residuals (SR) were employed to measure the disparities between the model's

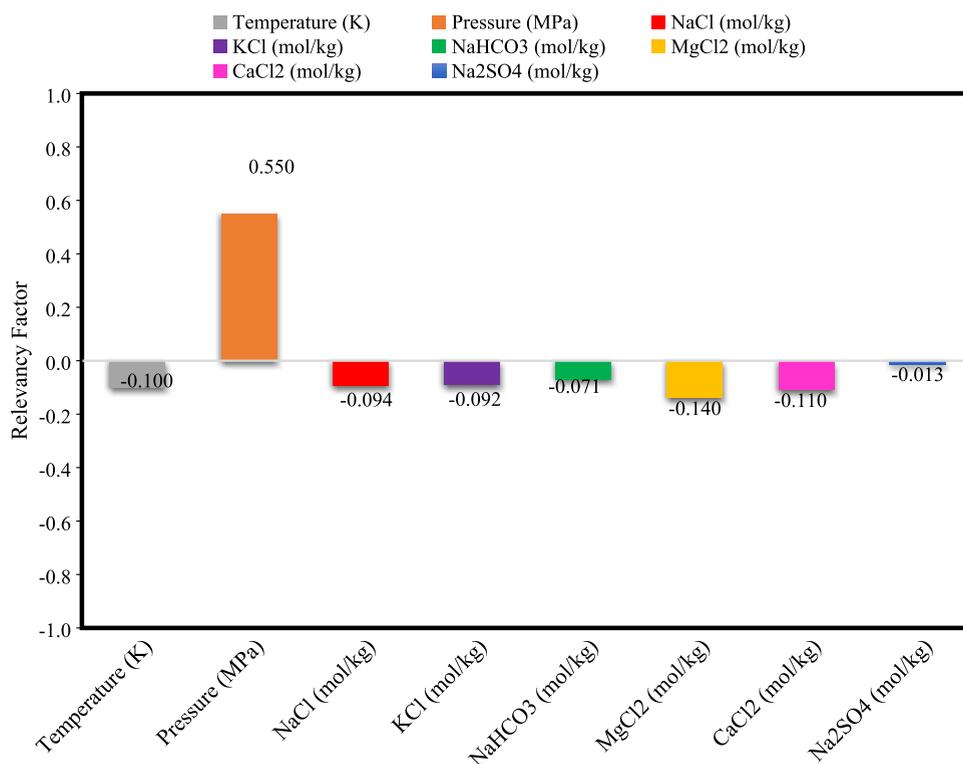


Figure 11. Sensitivity analysis using the result of the CFNN-LM model.

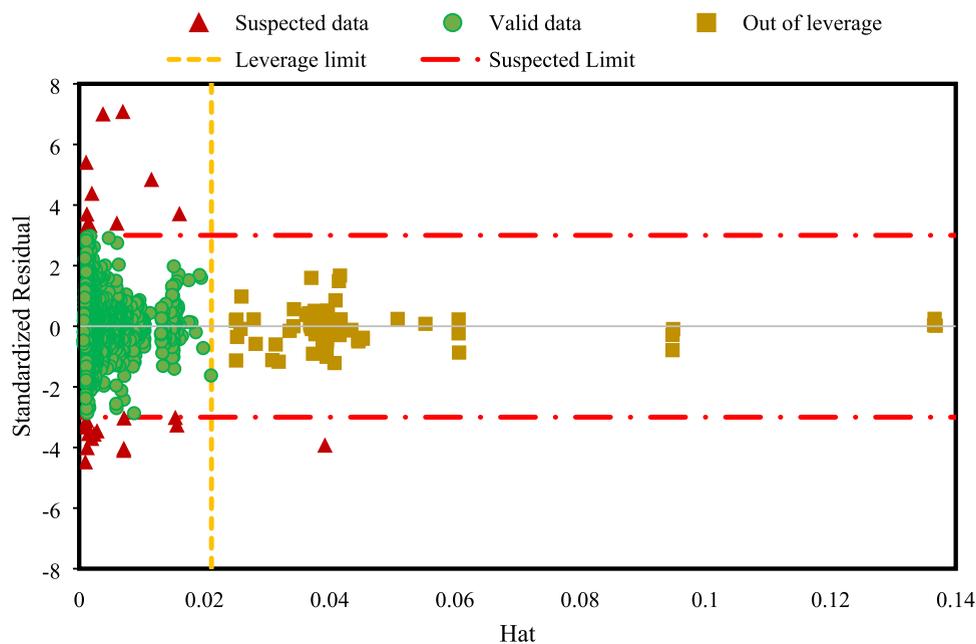


Figure 12. Williams plot for the CFNN-LM model obtained by the leverage approach.

predictions and the actual data. The SR values are computed as below^{103,104}

$$SR_i = \frac{e_i}{[MSE(1 - H_i)]^2} \quad (10)$$

Here, the leverage value for the i th data point is represented as H_i while the corresponding error value is denoted as e_i . Additionally, the mean square of error is given by MSE. Next, to calculate the Hat or leverage indices, one should calculate the H matrix as shown below¹⁰³

$$H = M(M^T M)^{-1} M^T \quad (11)$$

In the provided context, the symbol M^T denotes the transpose of a matrix referred to as M . Matrix M is characterized by its dimensions as $(n \times v)$, signifying n rows corresponding to data points and v columns representing the input variables. Additionally, the critical leverage (H^*), which is database-specific, can be determined using the subsequent formula¹⁰²

$$H^* = \frac{3 \times (\nu + 1)}{n} \quad (12)$$

The Williams plot is a visual method employed for scrutinizing a data set, pinpointing possible out-of-leverage and suspected data. Its purpose is to gauge the model's relevance by plotting SR values against Hat values. When a substantial proportion of data clusters within the $0 \leq H \leq H^*$ and $-3 \leq SR \leq 3$ intervals, it signifies both the trustworthiness of the data set and the accuracy of the model's predictions from a statistical standpoint.¹⁰⁵ As depicted in Figure 12, a total of 30 data points were identified as potentially doubtful data, suggesting the absence of dubious laboratory observations in the solubility database. Furthermore, 72 data points were identified as out of leverage, indicating that while the model's predictions were accurate, these data points fell outside the model's applicable range and statistically deviated from the bulk of the data set. The fact that only a limited number of potential suspected and out-of-leverage data were identified, considering the extensive solubility database used in this study, underscores the trustworthiness and precision of both the data set and the performance of the CFNN-LM model within this context. These models can be confidently used for predicting CO₂ solubility in saline aquifer, which is very important for CO₂ sequestration and storage

5. CONCLUSIONS

In this study, a data set consisting of 1278 experimental data points was employed to construct models for CO₂ solubility in both pure water and brine. This encompassed a temperature span from 273.15 to 453.15 K and a pressure range spanning 0.06–100 MPa. The modeling endeavor employed two robust ANNs, namely, CFNN and GRNN, coupled with three optimization algorithms, namely, LM, BR, and BFGS. Drawing from the findings of this study, the following conclusions can be made.

1. The CFNN-LM model showcases AAPRE values of 5.37% for the overall data set, 5.26% for the training subset, and 5.85% for the testing subset. Also, the CFNN-LM model demonstrates outstanding accuracy, boasting the highest overall R^2 value of 0.9949 among the other models. Consequently, following thorough statistical analyses, the CFNN-LM model stands out as the most accurate among the models crafted in this study.
2. Overall, the models can be ranked in order of performance as follows, from best to least: CFNN-LM, CFNN-BR, CFNN-BFGS, and GRNN.
3. The trend analysis showed a noteworthy positive impact of pressure on CO₂ solubility in brine, whereas salinity and temperature were identified as factors causing a decrease in CO₂ solubility. Notably, the CFNN-LM model exhibited remarkable predictive prowess, accurately estimating CO₂ solubility using experimental data under diverse salinity, pressure, and temperature conditions, yielding satisfactory results.
4. Based on sensitivity analysis, pressure exerts the most significant influence and stands as the sole parameter with a positive impact on CO₂ solubility in brine. Conversely, temperature and the concentration of all six salts considered in the model exhibited a negative impact. All salts exert a negative impact on CO₂ solubility due to their salting-out effect, with varying

degrees of influence. The salting-out effects of the salts can be ranked as follows, from the most pronounced to the least: MgCl₂ > CaCl₂ > NaCl > KCl > NaHCO₃ > Na₂SO₄.

5. Utilizing the leverage approach, 30 data points were pinpointed as potential suspected data, indicating the data set's credibility. Additionally, 72 data points were categorized as out-of-leverage data. The relatively low count of identified potential suspected and out-of-leverage data, given the expansive solubility database, underscores the reliability and accuracy of both the data set and the CFNN-LM model's performance in this context.
6. The solubility of CO₂ in brine plays an important role in monitoring the CO₂ sequestration process. CO₂ solubility in brine is regarded as the principal parameter in the CO₂ sequestration in saline aquifers. Thus, precise estimation is necessary to mitigate CO₂ emissions into the atmosphere.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c07962>.

Description of all optimization techniques used in this study (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yingting Zhu – Research Institute of Petroleum Exploration & Development, Beijing 100083, China; Key Laboratory of Oilfield Chemistry of CNPC, Beijing 100083, China; orcid.org/0000-0002-4648-9898; Email: zhuyingting@petrochina.com.cn

Qun Zhang – State Key Laboratory of Enhanced Oil Recovery, Research Institute of Petroleum Exploration and Development, CNPC, Beijing 100083, China; Research Institute of Petroleum Exploration & Development, Beijing 100083, China; Email: Zhangqun1980@petrochina.com.cn

Authors

Xinyuan Zou – State Key Laboratory of Enhanced Oil Recovery, Research Institute of Petroleum Exploration and Development, CNPC, Beijing 100083, China; Research Institute of Petroleum Exploration & Development, Beijing 100083, China

Jing Lv – Research Institute of Petroleum Exploration & Development, Beijing 100083, China; Key Laboratory of Oilfield Chemistry of CNPC, Beijing 100083, China

Yuchi Zhou – Oil and Gas engineering research Institute, Petrochina Jilin Oilfield Company, Songyuan 138000, China

Bin Ding – Research Institute of Petroleum Exploration & Development, Beijing 100083, China; Key Laboratory of Oilfield Chemistry of CNPC, Beijing 100083, China

Weidong Liu – Research Institute of Petroleum Exploration & Development, Beijing 100083, China; Key Laboratory of Oilfield Chemistry of CNPC, Beijing 100083, China

Kai Xiao – State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China

Complete contact information is available at:

https://pubs.acs.org/10.1021/acsomega.3c07962

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was financially supported by the Open Fund Projects of Research Institute of Petroleum Exploration and Development (No. 2022-KFKT-32) and PetroChina Scientific Research and Technology Development Projects (No. 2021DQ0408).

REFERENCES

- (1) Rogelj, J.; Den Elzen, M.; Höhne, N.; Fransen, T.; Fekete, H.; Winkler, H.; Schaeffer, R.; Sha, F.; Riahi, K.; Meinshausen, M. Paris Agreement climate proposals need a boost to keep warming well below 2 °C. *Nature* **2016**, *534* (7609), 631–639.
- (2) Liu, Z.; Deng, Z.; Davis, S. J.; Giron, C.; Ciais, P. Monitoring global carbon emissions in 2021. *Nat. Rev. Earth Environ.* **2022**, *3* (4), 217–219.
- (3) Boot-Handford, M. E.; Abanades, J. C.; Anthony, E. J.; Blunt, M. J.; Brandani, S.; Mac Dowell, N.; Fernández, J. R.; Ferrari, M.-C.; Gross, R.; Hallett, J. P.; et al. Carbon capture and storage update. *Energy Environ. Sci.* **2014**, *7* (1), 130–189.
- (4) Gibbins, J.; Chalmers, H. Carbon capture and storage. *Energy Policy* **2008**, *36* (12), 4317–4322.
- (5) Pires, J. C. M.; Martins, F. G.; Alvim-Ferraz, M. C. M.; Simões, M. Recent developments on carbon capture and storage: An overview. *Chem. Eng. Res. Des.* **2011**, *89* (9), 1446–1460.
- (6) Bui, M.; Adjiman, C. S.; Bardow, A.; Anthony, E. J.; Boston, A.; Brown, S.; Fennell, P. S.; Fuss, S.; Galindo, A.; Hackett, L. A.; et al. Carbon capture and storage (CCS): the way forward. *Energy Environ. Sci.* **2018**, *11* (5), 1062–1176.
- (7) Ismail, I.; Gaganis, V. Carbon Capture, Utilization, and Storage in Saline Aquifers: Subsurface Policies, Development Plans, Well Control Strategies and Optimization Approaches—A Review. *Clean Technol.* **2023**, *5* (2), 609–637.
- (8) Ringrose, P. S.; Furre, A.-K.; Gilfillan, S. M.; Krevor, S.; Landrø, M.; Leslie, R.; Meckel, T.; Nazarian, B.; Zahid, A. Storage of carbon dioxide in saline aquifers: physicochemical processes, key constraints, and scale-up potential. *Annu. Rev. Chem. Biomol. Eng.* **2021**, *12*, 471–494.
- (9) Metz, B.; Davidson, O.; De Coninck, H.; Loos, M.; Meyer, L. *IPCC Special Report on Carbon Dioxide Capture and Storage*; Cambridge University Press: Cambridge, 2005.
- (10) OECD, I. *Energy and Air Pollution: World Energy Outlook Special Report 2016*, 2016.
- (11) Raza, A.; Gholami, R.; Rezaee, R.; Bing, C. H.; Nagarajan, R.; Hamid, M. A. Preliminary assessment of CO₂ injectivity in carbonate storage sites. *Petroleum* **2017**, *3* (1), 144–154.
- (12) Raza, A.; Gholami, R.; Rezaee, R.; Bing, C. H.; Nagarajan, R.; Hamid, M. A. Well selection in depleted oil and gas fields for a safe CO₂ storage practice: A case study from Malaysia. *Petroleum* **2017**, *3* (1), 167–177.
- (13) Michael, K.; Arnot, M.; Cook, P.; Ennis-King, J.; Funnell, R.; Kaldi, J.; Kirste, D.; Paterson, L. CO₂ storage in saline aquifers I—Current state of scientific knowledge. *Energy Procedia* **2009**, *1* (1), 3197–3204.
- (14) Saadatpoor, E.; Bryant, S. L.; Sepehrmoori, K. New trapping mechanism in carbon sequestration. *Transp. Porous Media* **2010**, *82*, 3–17.
- (15) Iglauer, S. *Dissolution Trapping of Carbon Dioxide in Reservoir Formation Brine—a Carbon Storage Mechanism*; INTECH Open Access Publisher: London, UK, 2011.
- (16) Zhang, D.; Song, J. Mechanisms for geological carbon sequestration. *Procedia IUTAM* **2014**, *10*, 319–327.
- (17) Irfan, M. F.; Bisson, T. M.; Bobicki, E.; Arguelles-Vivas, F.; Xu, Z.; Liu, Q.; Babadagli, T. CO₂ storage in saline aquifers by dissolution

and residual trapping under supercritical conditions: An experimental investigation. *Colloids Surf., A* **2018**, *548*, 37–45.

(18) Rosenbauer, R.; Thomas, B. Carbon dioxide (CO₂) sequestration in deep saline aquifers and formations. In *Developments and Innovation in Carbon Dioxide (CO₂) Capture and Storage Technology*; Elsevier, 2010; pp 57–103.

(19) Findlay, A.; Howell, O. R. XXXII.—The solubility of carbon dioxide in water in the presence of starch. *J. Chem. Soc., Trans.* **1915**, *107*, 282–284.

(20) Martín, Á.; Pham, H. M.; Kilzer, A.; Kareth, S.; Weidner, E. Phase equilibria of carbon dioxide+poly ethylene glycol+water mixtures at high pressure: Measurements and modelling. *Fluid Phase Equilib.* **2009**, *286* (2), 162–169.

(21) Siqueira Campos, C. E. P.; Villardi, H. G. D. A.; Pessoa, F. L. P.; Uller, A. M. C. Solubility of Carbon Dioxide in Water and Hexadecane: Experimental Measurement and Thermodynamic Modeling. *J. Chem. Eng. Data* **2009**, *54* (10), 2881–2886.

(22) Wiebe, R.; Gaddy, V. Vapor phase composition of carbon dioxide-water mixtures at various temperatures and at pressures to 700 atm. *J. Am. Chem. Soc.* **1941**, *63* (2), 475–477.

(23) Yeh, S. Y.; Peterson, R. E. Solubility of carbon dioxide, krypton, and xenon in aqueous solution. *J. Pharm. Sci.* **1964**, *53* (7), 822–824.

(24) Teng, H.; Yamasaki, A.; Chun, M.-K.; Lee, H. Solubility of liquid CO₂ in water at temperatures from 278 to 293 K and pressures from 6.44 to 29.49 MPa and densities of the corresponding aqueous solutions. *J. Chem. Thermodyn.* **1997**, *29* (11), 1301–1310.

(25) Chapoy, A.; Mohammadi, A.; Chareton, A.; Tohidi, B.; Richon, D. Measurement and modeling of gas solubility and literature review of the properties for the carbon dioxide–water system. *Ind. Eng. Chem. Res.* **2004**, *43* (7), 1794–1802.

(26) Kiepe, J.; Horstmann, S.; Fischer, K.; Gmehling, J. Experimental determination and prediction of gas solubility data for CO₂+H₂O mixtures containing NaCl or KCl at temperatures between 313 and 393 K and pressures up to 10 MPa. *Ind. Eng. Chem. Res.* **2002**, *41* (17), 4393–4398.

(27) Findlay, A.; Williams, T. LXXI.—The influence of colloids and fine suspensions on the solubility of gases in water. Part III. Solubility of carbon dioxide at pressures lower than atmospheric. *J. Chem. Soc., Trans.* **1913**, *103*, 636–645.

(28) Yasunishi, A.; Yoshida, F. Solubility of carbon dioxide in aqueous electrolyte solutions. *J. Chem. Eng. Data* **1979**, *24* (1), 11–14.

(29) Takenouchi, S.; Kennedy, G. C. The solubility of carbon dioxide in NaCl solutions at high temperatures and pressures. *Am. J. Sci.* **1965**, *263* (5), 445–454.

(30) dos Santos, P. F.; André, L.; Ducouso, M.; Lassin, A.; Contamine, F.; Lach, A.; Parmentier, M.; Cézac, P. An improved model for CO₂ solubility in aqueous Na⁺–Cl[–]–SO₄^{2–} systems up to 473.15 K and 40 MPa. *Chem. Geol.* **2021**, *582*, No. 120443, DOI: 10.1016/j.chemgeo.2021.120443.

(31) Han, X.; Yu, Z.; Qu, J.; Qi, T.; Guo, W.; Zhang, G. Measurement and Correlation of Solubility Data for CO₂ in NaHCO₃ Aqueous Solution. *J. Chem. Eng. Data* **2011**, *56* (4), 1213–1219.

(32) Aggelopoulos, C.; Robin, M.; Vizika, O. Interfacial tension between CO₂ and brine (NaCl+CaCl₂) at elevated pressures and temperatures: The additive effect of different salts. *Adv. Water Resour.* **2011**, *34* (4), 505–511.

(33) Zhao, H.; Dilmore, R. M.; Lvov, S. N. Experimental studies and modeling of CO₂ solubility in high temperature aqueous CaCl₂, MgCl₂, Na₂SO₄, and KCl solutions. *AIChE J.* **2015**, *61* (7), 2286–2297.

(34) Liu, Y.; Hou, M.; Yang, G.; Han, B. Solubility of CO₂ in aqueous solutions of NaCl, KCl, CaCl₂ and their mixed salts at different temperatures and pressures. *J. Supercrit. Fluids* **2011**, *56* (2), 125–129.

(35) Bando, S.; Takemura, F.; Nishio, M.; Hihara, E.; Akai, M. Solubility of CO₂ in aqueous solutions of NaCl at (30 to 60) °C and (10 to 20) MPa. *J. Chem. Eng. Data* **2003**, *48* (3), 576–579.

- (36) Koschel, D.; Coxam, J.-Y.; Rodier, L.; Majer, V. Enthalpy and solubility data of CO₂ in water and NaCl (aq) at conditions of interest for geological sequestration. *Fluid Phase Equilib.* **2006**, *247* (1–2), 107–120.
- (37) Bastami, A.; Allahgholi, M.; Pourafshary, P. Experimental and modelling study of the solubility of CO₂ in various CaCl₂ solutions at different temperatures and pressures. *Pet. Sci.* **2014**, *11*, 569–577.
- (38) Hou, S.-X.; Maitland, G. C.; Trusler, J. M. Phase equilibria of (CO₂+ H₂O+ NaCl) and (CO₂+ H₂O+ KCl): Measurements and modeling. *J. Supercrit. Fluids* **2013**, *78*, 78–88.
- (39) Yan, Y.; Chen, C.-C. Thermodynamic modeling of CO₂ solubility in aqueous solutions of NaCl and Na₂SO₄. *J. Supercrit. Fluids* **2010**, *55* (2), 623–634.
- (40) Spycher, N.; Pruess, K. A phase-partitioning model for CO₂-brine mixtures at elevated temperatures and pressures: application to CO₂-enhanced geothermal systems. *Transp. Porous Media* **2010**, *82*, 173–196.
- (41) Duan, Z.; Møller, N.; Weare, J. H. Equation of state for the NaCl-H₂O-CO₂ system: prediction of phase equilibria and volumetric properties. *Geochim. Cosmochim. Acta* **1995**, *59* (14), 2869–2882.
- (42) Rumpf, B.; Nicolaisen, H.; Öcal, C.; Maurer, G. Solubility of carbon dioxide in aqueous solutions of sodium chloride: experimental results and correlation. *J. Solution Chem.* **1994**, *23*, 431–448.
- (43) Appelo, C.; Parkhurst, D. L.; Post, V. Equations for calculating hydrogeochemical reactions of minerals and gases such as CO₂ at high pressures and temperatures. *Geochim. Cosmochim. Acta* **2014**, *125*, 49–67.
- (44) Springer, R. D.; Wang, Z.; Anderko, A.; Wang, P.; Felmy, A. R. A thermodynamic model for predicting mineral reactivity in supercritical carbon dioxide: I. Phase behavior of carbon dioxide–water–chloride salt systems across the H₂O-rich to the CO₂-rich regions. *Chem. Geol.* **2012**, *322–323*, 151–171.
- (45) Mutailipu, M.; Song, Y.; Yao, Q.; Liu, Y.; Trusler, J. M. Solubility and interfacial tension models for CO₂-brine systems under CO₂ geological storage conditions. *Fuel* **2024**, *357*, No. 129712.
- (46) Mohammadi, M.-R.; Hadavimoghaddam, F.; Pourmahdi, M.; Atashrouz, S.; Munir, M. T.; Hemmati-Sarapardeh, A.; Mosavi, A. H.; Mohaddespour, A. Modeling hydrogen solubility in hydrocarbons using extreme gradient boosting and equations of state. *Sci. Rep.* **2021**, *11* (1), No. 17911.
- (47) Ali Ahmadi, M.; Ahmadi, A. Applying a sophisticated approach to predict CO₂ solubility in brines: application to CO₂ sequestration. *Int. J. Low-Carbon Technol.* **2016**, *11* (3), 325–332.
- (48) Mohammadian, E.; Motamedi, S.; Shamsirband, S.; Hashim, R.; Junin, R.; Roy, C.; Azdarpour, A. Application of extreme learning machine for prediction of aqueous solubility of carbon dioxide. *Environ. Earth Sci.* **2016**, *75*, No. 215.
- (49) Menad, N. A.; Hemmati-Sarapardeh, A.; Varamesh, A.; Shamsirband, S. Predicting solubility of CO₂ in brine by advanced machine learning systems: Application to carbon capture and sequestration. *J. CO₂ Util.* **2019**, *33*, 83–95.
- (50) Hemmati-Sarapardeh, A.; Amar, M. N.; Soltanian, M. R.; Dai, Z.; Zhang, X. Modeling CO₂ solubility in water at high pressure and temperature conditions. *Energy Fuels* **2020**, *34* (4), 4761–4776.
- (51) Sayahi, T.; Tatar, A.; Rostami, A.; Anbaz, M. A.; Shahbazi, K. Determining solubility of CO₂ in aqueous brine systems via hybrid smart strategies. *Int. J. Comput. Appl. Technol.* **2021**, *65* (1), 1–13.
- (52) Mohammadian, E.; Liu, B.; Riazi, A. Evaluation of different machine learning frameworks to estimate CO₂ solubility in NaCl brines: implications for CO₂ injection into low-salinity formations. *Lithosphere* **2022**, *2022* (Special 12), No. 1615832.
- (53) Nakhaei-Kohani, R.; Taslimi-Renani, E.; Hadavimoghaddam, F.; Mohammadi, M.-R.; Hemmati-Sarapardeh, A. Modeling solubility of CO₂-N₂ gas mixtures in aqueous electrolyte systems using artificial intelligence techniques and equations of state. *Sci. Rep.* **2022**, *12* (1), No. 3625.
- (54) Lv, Q.; Zhou, T.; Zheng, R.; Nakhaei-Kohani, R.; Riazi, M.; Hemmati-Sarapardeh, A.; Li, J.; Wang, W. Application of group method of data handling and gene expression programming for predicting solubility of CO₂-N₂ gas mixture in brine. *Fuel* **2023**, *332*, No. 126025.
- (55) Ratnakar, R. R.; Chaubey, V.; Dindoruk, B. A novel computational strategy to estimate CO₂ solubility in brine solutions for CCUS applications. *Appl. Energy* **2023**, *342*, No. 121134.
- (56) Khoshraftar, Z.; Ghaemi, A. Prediction of CO₂ solubility in water at high pressure and temperature via deep learning and response surface methodology. *Case Stud. Chem. Environ. Eng.* **2023**, *7*, No. 100338.
- (57) Vo Thanh, H.; Yasin, Q.; Al-Mudhafar, W. J.; Lee, K.-K. Knowledge-based machine learning techniques for accurate prediction of CO₂ storage performance in underground saline aquifers. *Appl. Energy* **2022**, *314*, No. 118985.
- (58) Al-qaness, M. A. A.; Ewees, A. A.; Thanh, H. V.; AlRassas, A. M.; Dahou, A.; Elaziz, M. A. Predicting CO₂ trapping in deep saline aquifers using optimized long short-term memory. *Environ. Sci. Pollut. Res.* **2023**, *30* (12), 33780–33794.
- (59) Davoodi, S.; Thanh, H. V.; Wood, D. A.; Mehrad, M.; Rukavishnikov, V. S. Combined machine-learning and optimization models for predicting carbon dioxide trapping indexes in deep geological formations. *Appl. Soft Comput.* **2023**, *143*, No. 110408.
- (60) Davoodi, S.; Thanh, H. V.; Wood, D. A.; Mehrad, M.; Rukavishnikov, V. S.; Dai, Z. Machine-learning predictions of solubility and residual trapping indexes of carbon dioxide from global geological storage sites. *Expert Syst. Appl.* **2023**, *222*, No. 119796.
- (61) Al Ghafri, S. Z. S.; Forte, E.; Galindo, A.; Maitland, G. C.; Trusler, J. P. M. Experimental and Modeling Study of the Phase Behavior of (Heptane + Carbon Dioxide + Water) Mixtures. *J. Chem. Eng. Data* **2015**, *60* (12), 3670–3681.
- (62) Anderson, G. K. Solubility of Carbon Dioxide in Water under Incipient Clathrate Formation Conditions. *J. Chem. Eng. Data* **2002**, *47* (2), 219–222.
- (63) Bamberger, A.; Sieder, G.; Maurer, G. High-pressure (vapor+liquid) equilibrium in binary mixtures of (carbon dioxide+water or acetic acid) at temperatures from 313 to 353 K. *J. Supercrit. Fluids* **2000**, *17* (2), 97–110.
- (64) Dalmolin, I.; Skovroinski, E.; Biasi, A.; Corazza, M. L.; Dariva, C.; Oliveira, J. V. Solubility of carbon dioxide in binary and ternary mixtures with ethanol and water. *Fluid Phase Equilib.* **2006**, *245* (2), 193–200.
- (65) Dell'Era, C.; Uusi-Kyyny, P.; Pokki, J.-P.; Pakkanen, M.; Alopaeus, V. Solubility of carbon dioxide in aqueous solutions of diisopropanolamine and methyldiethanolamine. *Fluid Phase Equilib.* **2010**, *293* (1), 101–109.
- (66) Han, J. M.; Shin, H. Y.; Min, B.-M.; Han, K.-H.; Cho, A. Measurement and correlation of high pressure phase behavior of carbon dioxide+water system. *J. Ind. Eng. Chem.* **2009**, *15* (2), 212–216.
- (67) Hou, S.-X.; Maitland, G. C.; Trusler, J. P. M. Measurement and modeling of the phase behavior of the (carbon dioxide+water) mixture at temperatures from 298.15K to 448.15K. *J. Supercrit. Fluids* **2013**, *73*, 87–96.
- (68) Liu, Y.; Hou, M.; Ning, H.; Yang, D.; Yang, G.; Han, B. Phase equilibria of CO₂ + N₂ + H₂O and N₂ + CO₂ + H₂O + NaCl + KCl + CaCl₂ systems at different temperatures and pressures. *J. Chem. Eng. Data* **2012**, *57*, 1928–1932.
- (69) Muromachi, S.; Shijima, A.; Miyamoto, H.; Ohmura, R. Experimental measurements of carbon dioxide solubility in aqueous tetra-n-butylammonium bromide solutions. *J. Chem. Thermodyn.* **2015**, *85*, 94–100.
- (70) Ruffine, L.; Trusler, J. P. M. Phase behaviour of mixed-gas hydrate systems containing carbon dioxide. *J. Chem. Thermodyn.* **2010**, *42* (5), 605–611.
- (71) Serpa, F. S.; Vidal, R. S.; Filho, J. H. B. A.; Nascimento, J. F. d.; Ciambelli, J. R. P.; Figueiredo, C. M. S.; Salazar-Banda, G. R.; Santos, A. F.; Fortuny, M.; Franceschi, E.; Dariva, C. Solubility of Carbon

- Dioxide in Ethane-1,2-diol–Water Mixtures. *J. Chem. Eng. Data* **2013**, *58* (12), 3464–3469.
- (72) Servio, P.; Englezos, P. Effect of temperature and pressure on the solubility of carbon dioxide in water in the presence of gas hydrate. *Fluid Phase Equilib.* **2001**, *190* (1), 127–134.
- (73) Tang, Y.; Bian, X.; Du, Z.; Wang, C. Measurement and prediction model of carbon dioxide solubility in aqueous solutions containing bicarbonate anion. *Fluid Phase Equilib.* **2015**, *386*, 56–64.
- (74) Valtz, A.; Chapoy, A.; Coquelet, C.; Paricaud, P.; Richon, D. Vapour–liquid equilibria in the carbon dioxide–water system, measurement and modelling from 278.2 to 318.2K. *Fluid Phase Equilib.* **2004**, *226*, 333–344.
- (75) Zhang, G.; Wu, Y.; Ma, P.; Wu, G.; Li, D. Measurement and correlation of solubility of carbon monoxide and other gases solubility in phenol. *Huagong Xuebao (Chin. Ed.)* **2005**, *56*, 2039–2045.
- (76) Nighswander, J. A.; Kalogerakis, N.; Mehrotra, A. K. Solubilities of carbon dioxide in water and 1 wt % sodium chloride solution at pressures up to 10 MPa and temperatures from 80 to 200.degree.C. *J. Chem. Eng. Data* **1989**, *34* (3), 355–360.
- (77) Hoballah, R. *On the Solubility of Acid and Sour Gases in Water and Brines under Reservoir Conditions*, **2017**.
- (78) Guo, H.; Huang, Y.; Chen, Y.; Zhou, Q. Quantitative Raman spectroscopic measurements of CO₂ solubility in NaCl solution from (273.15 to 473.15) K at p = (10.0, 20.0, 30.0, and 40.0) MPa. *J. Chem. Eng. Data* **2016**, *61* (1), 466–474.
- (79) Tong, D.; Trusler, J. M.; Vega-Maza, D. Solubility of CO₂ in aqueous solutions of CaCl₂ or MgCl₂ and in a synthetic formation brine at temperatures up to 423 K and pressures up to 40 MPa. *J. Chem. Eng. Data* **2013**, *58* (7), 2116–2124.
- (80) Malinin, S.; Saveleva, N. Experimental investigations of CO₂ solubility in NaCl and CaCl₂ solutions at temperatures of 25, 50 and 75 degrees and elevated CO₂ pressure. *Geokhimiya* **1972**, No. 6, 643.
- (81) Fahlman, S.; Lebiere, C. The cascade-correlation learning architecture. *Adv. Neural Inf. Process. Syst.* **1989**; Vol. 2.
- (82) Mohammadi, M.-R.; Hemmati-Sarapardeh, A.; Schaffie, M.; Husein, M. M.; Ranjbar, M. Application of cascade forward neural network and group method of data handling to modeling crude oil pyrolysis during thermal enhanced oil recovery. *J. Pet. Sci. Eng.* **2021**, *205*, No. 108836.
- (83) Nami, F.; Deyhimi, F. Prediction of activity coefficients at infinite dilution for organic solutes in ionic liquids by artificial neural network. *J. Chem. Thermodyn.* **2011**, *43* (1), 22–27.
- (84) Hedayat, A.; Davilu, H.; Barfrosh, A. A.; Sepanloo, K. Estimation of research reactor core parameters using cascade feed forward artificial neural networks. *Prog. Nucl. Energy* **2009**, *51* (6–7), 709–718.
- (85) Specht, D. F. A general regression neural network. *IEEE Trans. Neural Networks* **1991**, *2* (6), 568–576.
- (86) Asante-Okyere, S.; Xu, Q.; Mensah, R. A.; Jin, C.; Ziggah, Y. Y. Generalized regression and feed forward back propagation neural networks in modelling flammability characteristics of polymethyl methacrylate (PMMA). *Thermochim. Acta* **2018**, *667*, 79–92.
- (87) Rooki, R. Application of general regression neural network (GRNN) for indirect measuring pressure loss of Herschel–Bulkley drilling fluids in oil drilling. *Measurement* **2016**, *85*, 184–191.
- (88) Cigizoglu, H. K. Application of generalized regression neural networks to intermittent flow forecasting and estimation. *J. Hydrol. Eng.* **2005**, *10* (4), 336–341.
- (89) Feng, Y.; Peng, Y.; Cui, N.; Gong, D.; Zhang, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* **2017**, *136*, 71–78.
- (90) Ding, L.; Rangaraju, P.; Poursaee, A. Application of generalized regression neural network method for corrosion modeling of steel embedded in soil. *Soils Found.* **2019**, *59* (2), 474–483.
- (91) Mohammadi, M.-R.; Hemmati-Sarapardeh, A.; Schaffie, M.; Husein, M. M.; Karimian, M.; Ranjbar, M. On the evaluation of crude oil oxidation during thermogravimetry by generalised regression neural network and gene expression programming: application to thermal enhanced oil recovery. *Combust. Theory Modell.* **2021**, *25* (7), 1268–1295.
- (92) Mohammadi, M.-R.; Hadavimoghaddam, F.; Atashrouz, S.; Abedi, A.; Hemmati-Sarapardeh, A.; Mohaddespour, A. Modeling the solubility of light hydrocarbon gases and their mixture in brine with machine learning and equations of state. *Sci. Rep.* **2022**, *12* (1), No. 14943.
- (93) Liu, B.; Mohammadi, M.-R.; Ma, Z.; Bai, L.; Wang, L.; Xu, Y.; Ostadhassan, M.; Hemmati-Sarapardeh, A. Evolution of porosity in kerogen type I during hydrous and anhydrous pyrolysis: Experimental study, mechanistic understanding, and model development. *Fuel* **2023**, *338*, No. 127149.
- (94) Mackay, D.; Shiu, W. Y. A critical review of Henry's law constants for chemicals of environmental interest. *J. Phys. Chem. Ref. Data* **1981**, *10* (4), 1175–1199.
- (95) Zhang, J.; Pan, Z.; Liu, K.; Burke, N. Molecular simulation of CO₂ solubility and its effect on octane swelling. *Energy Fuels* **2013**, *27* (5), 2741–2747.
- (96) Quílez-Pardo, J.; Solaz-Portolés, J. J. Students' and teachers' misapplication of Le Chatelier's principle: Implications for the teaching of chemical equilibrium. *J. Res. Sci. Teach.* **1995**, *32* (9), 939–957.
- (97) Duan, Z.; Sun, R. An improved model calculating CO₂ solubility in pure water and aqueous NaCl solutions from 273 to 533 K and from 0 to 2000 bar. *Chem. Geol.* **2003**, *193* (3–4), 257–271.
- (98) Görgényi, M.; Dewulf, J.; Van Langenhove, H.; Héberger, K. Aqueous salting-out effect of inorganic cations and anions on non-electrolytes. *Chemosphere* **2006**, *65* (5), 802–810.
- (99) Ansari, S.; Mohammadi, M.-R.; Bahmaninia, H.; Hemmati-Sarapardeh, A.; Schaffie, M.; Norouzi-Apourvari, S.; Ranjbar, M. Experimental measurement and modeling of asphaltene adsorption onto iron oxide and lime nanoparticles in the presence and absence of water. *Sci. Rep.* **2023**, *13* (1), No. 122.
- (100) Chen, G.; Fu, K.; Liang, Z.; Sema, T.; Li, C.; Tontiwachwuthikul, P.; Idem, R. The genetic algorithm based back propagation neural network for MMP prediction in CO₂-EOR process. *Fuel* **2014**, *126*, 202–212.
- (101) Leroy, A. M.; Rousseeuw, P. J. *Robust Regression and Outlier Detection*, **1987**.
- (102) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26* (5), 694–701.
- (103) Rousseeuw, P. J.; Leroy, A. M. *Robust Regression and Outlier Detection*; John Wiley & Sons, **2005**.
- (104) Hadavimoghaddam, F.; Mohammadi, M.-R.; Atashrouz, S.; Nedeljkovic, D.; Hemmati-Sarapardeh, A.; Mohaddespour, A. Data-driven modeling of H₂ solubility in hydrocarbons using white-box approaches. *Int. J. Hydrogen Energy* **2022**, *47* (78), 33224–33238.
- (105) Lv, Q.; Rashidi-Khaniabadi, A.; Zheng, R.; Zhou, T.; Mohammadi, M.-R.; Hemmati-Sarapardeh, A. Modelling CO₂ diffusion coefficient in heavy crude oils and bitumen using extreme gradient boosting and Gaussian process regression. *Energy* **2023**, *275*, No. 127396.