AIMS Geosciences

AIMS Geosciences, 11(1): 201–227.
DOI: 10.3934/geosci.2025009
Received: 25 November 2024
Revised: 23 February 2025
Accepted: 28 February 2025
Published: 18 March 2025

https://www.aimspress.com/journal/geosciences

*Research article*

# ANN prediction of the $CO_2$ solubility in water and brine under reservoir conditions

**Shuo Yang, Dong Wang*, Zeguang Dong, Yingge Li and Dongxing Du***

Geo-Energy Research Institute, College of Electromechanical Engineering, Qingdao University of Science and Technology, Gaomi 261550, China

*** Correspondence:** Email: wangdong@mails.qust.edu.cn; du-dongxing@qust.edu.cn.

**Abstract:** Having accurate knowledge on $CO_2$ solubility in reservoir liquids plays a pivotal role in geoenergy harvest and carbon capture, utilization, and storage (CCUS) applications. Data-driven works leveraging artificial neural networks (ANN) have presented a promising tool for forecasting $CO_2$ solubility. In this paper, an ANN model was developed based on hundreds of documented data to predict $CO_2$ solubility in both pure water and saline solutions across a broad spectrum of temperatures, pressures, and salinities in reference to underground formation conditions. Multilayer perceptron (MLP) models were constructed for each system, and their prediction results were rigorously validated against the the literature data. The research results indicate that the ANN model is suitable for predicting the solubility of carbon dioxide under different conditions, with root mean square errors (RMSE) of 0.00108 and 0.00036 for water and brine, and a coefficient of determination ($R^2$) of 0.99424 and 0.99612, which indicates robust prediction capacities. It was observed from the ANN model that the saline water case could not be properly expanded to predict the $CO_2$ solubility in pure water, underscoring the distinct dissolution mechanisms in polar mixtures. It is expected that this study could provide a valuable reference and offer novel insights to the prediction of $CO_2$ solubility in complex fluid systems.

**Keywords:** $CO_2$ solubility; brine; water; artificial neural network; multilayer perceptron

## 1. Introduction

As global climate change intensifies, the search for and development of clean, sustainable energy solutions have become increasingly critical [1]. The expansion of renewable energy sources and the implementation of carbon capture, utilization, and storage (CCUS) technologies within various industries are deemed essential strategies for mitigating $CO_2$ emissions [2–8]. Geothermal energy, as a nearly carbon-neutral renewable energy source, is playing an increasingly prominent role in the global energy transition [9–13]. Concurrently, carbon dioxide ($CO_2$) geological storage, as an effective means to mitigate climate change, has attracted widespread attention for its potential when combined with geothermal energy development.

The $CO_2$-$H_2O$ and $CO_2$-NaCl-$H_2O$ fluid systems are of paramount importance in Earth's systems and a variety of engineering geological processes, such as $CO_2$-enhanced oil recovery (EOR), and $CO_2$-enhanced geothermal systems [14–19]. The phase behavior of $CO_2$-$H_2O$ as well as $H_2O$-NaCl-$CO_2$ systems are therefore of significant industrial relevance [20–24]. The processes of dissolution, precipitation, and ion exchange within these fluid systems can markedly influence the phase equilibrium behavior in the geological media [25–27].

To predict the $CO_2$ solubility in reservoir liquids, thermodynamic simulation stands as an effective branch in addition to the experimental studies. In thermodynamic simulation works involving $CO_2$, however, traditional cubic equations of state (EOS) and van der Waals mixing rules are commonly used for gas-liquid equilibrium calculations, and in most cases overlook the water phase's effects. Given the importance of predicting $CO_2$ solubility, researchers have extensively studied with thermodynamic models, such as those proposed by Duan et al. (2006) [28], Spycher et al. (2003) [29], and Z. Liu et al. (2022) [30]. However, these models are limited to specific temperature, pressure, and salinity ranges. Therefore, developing a predictive model capable of handling polar mixtures and capturing complex nonlinear relationships still remains a challenge in terms of complexity and predictive accuracy.

In recent years, data-driven models based on artificial neural networks (ANN) offer a new approach for predicting $CO_2$ solubility. Neural networks are large parallel systems composed of many highly interconnected components (called nodes or neurons). The structure of the connections between neurons and their arrangement effectively determine the system's operation. The strength of the connections and how the network responds to instructions are determined by its learning algorithm. Compared to traditional thermodynamic models, ANN models can automatically learn complex nonlinear relationships through large-scale data training, avoiding cumbersome parameter adjustments, and provide greater flexibility and accuracy under various temperature and pressure conditions [31].This makes ANN-based predictive methods a powerful tool for addressing the current challenges of predicting $CO_2$ solubility.

ANNs are typically divided into two main categories: feedforward neural networks and recurrent neural networks [32]. The multilayer perceptron (MLP) is a classic example of a feedforward neural network. MLP consists of an input layer, one or more hidden layers, and an output layer. The input layer receives external information, the hidden layers process the information, and the output layer generates the network's response [33]. Each layer contains multiple neurons connected by weights. Each neuron receives input signals from the previous layer, applies weighted summation, and generates

an output signal through a nonlinear transformation, which is then passed to the next layer. This feedforward mechanism enables the network to perform complex classification and regression tasks.

To optimize network parameters for specific tasks, training data is used to optimize these parameters. The backpropagation algorithm is the most commonly used training method. It iteratively adjusts network parameters to minimize the error between predicted and actual values, improving model accuracy and reliability [34]. An ANN's parallel processing capabilities allow it to extract essential features and underlying patterns from input data. ANNs have achieved remarkable results in fields such as pattern recognition, image processing, and data mining. Classic network models like MLPs, convolutional neural networks (CNN), and recurrent neural networks (RNN) have performed exceptionally well in these areas [35].

In the prediction of $CO_2$ solubility, advancements in artificial intelligence provide new solutions. The nonlinear fitting ability and self-learning properties of ANNs give them significant advantages in modeling complex systems. Compared to traditional methods, ANN-based solubility prediction models offer improvements in accuracy, applicability, and computational efficiency. By designing appropriate network structures and training them with extensive experimental data, ANNs can establish complex nonlinear mapping relationships between input variables (e.g., temperature, pressure, and salinity) and output ($CO_2$ solubility). This data-driven modeling approach does not require in-depth analysis of complex physicochemical mechanisms between variables but allows ANNs to autonomously learn data patterns, making them highly applicable and robust [36]. In recent studies, machine learning models have been extensively applied to predict the solubility of gases in hydrocarbons, which is crucial for carbon capture and storage (CCS) and enhanced oil recovery (EOR) technologies in the oil and chemical industries. For instance, Madani [37] utilized five machine learning models, including CatBoost, random forest, LightGBM, k-nearest neighbors (k-NN), and XGBoost, along with five equations of state (EOS), to forecast the solubility of nitrogen ($N_2$) in normal alkanes, revealing that the CatBoost model demonstrated the best predictive accuracy. Mohammadi [38] focused on the solubility of hydrogen in hydrocarbons, and through a comparison of five machine learning models, found that the XGBoost model stood out in terms of prediction accuracy. Additionally, Nakhaei-Kohani [39] employed six intelligent models to predict the solubility of $CO_2$-$N_2$ mixtures in aqueous solutions, with the random forest model showing the most accurate forecasting capabilities. Mahmoudzadeh [40] developed two intelligent models, gradient boosting (GBoost) and the light gradient boosting machine (LightGBM), to predict the solubility of pure $CO_2$ and impure $CO_2$ in water, with the GBoost model exhibiting superior performance in both scenarios.

This paper establishes an ANN-based method for studying the high-pressure phase behavior of gas-water and gas-saline solutions. Two systems, $CO_2$+$H_2O$ and $CO_2$+$H_2O$+NaCl, are investigated, and the accuracy of the models is verified through comparison with literature data. Additionally, the study examines whether the saline model is applicable for predicting $CO_2$ solubility in pure water with zero salinity. A database is constructed from extensive gas-liquid equilibrium data for the $CO_2$+$H_2O$ and $CO_2$+$H_2O$+NaCl systems, covering a wide range of pressure, temperature, and molar concentration of $CO_2$ and salts.

## 2. Methodology

### *2.1. Data acquistion*

Literature-reported data on $CO_2$ solubility in pure water systems [20,22,41–43] were collected, covering various conditions of temperature (273.15–573.15 K), pressure (0.19–120 MPa), and pure water. Table 1 provides a summary of the experimental studies, including the pressures of the respective applications, the amount of data and temperature ranges employed, and the methods used. The dataset size (number of samples) is 265, and the unit of the output variable ($CO_2$ solubility) is the mole fraction of $CO_2$. After excluding outliers, the data were divided into training, testing, and validation sets in a ratio of 16:4:5, with the validation set accounting for 20% (which represents 20% of the training set, or 16% of the total data). Therefore, the actual distribution of the data is as follows: 64% for the training set (80% of the training set × 80% of the total data), 16% for the validation set, and 20% for the test set. The specific values in the table refer to Table A1.

**Table 1.** Literature experiment studies on the determination of $CO_2$ solubility in water.

| References | Temperture (K) | Pressure (MPa) | Usage Method | Data Volume |
|---|---|---|---|---|
| Bamberger et al., 2004 | 273.15–573.15 | 10–120 | Quantitative Raman spectroscopy | 131 |
| Chapoy et al., 2004 | 278.22–318.23 | 0.465–7.933 | Gas-liquid equilibrium experiment | 47 |
| Guo et al., 2014 | 323.2–353.1 | 4.05–14.11 | Improved Peng-Robinson equation of state | 29 |
| Y. Liu et al., 2011 | 308.15–328.15 | 2.10–15.99 | Laboratory equipment measurement | 31 |
| Valtz et al., 2004 | 274.14–351.31 | 0.190–5.261 | VPT-EoS | 27 |

**Table 2.** Experimental studies in the literature on the determination of $CO_2$ solubility in brine.

| References | Salinity (mol/Kg) | Temperture (K) | Pressure (MPa) | Usage Method | Data Volume |
|---|---|---|---|---|---|
| Carvalho et al., 2015 | 1–6 | 313–433 | 0.5–30 | Søreide-Whitson models | 190 |
| Chabab et al., 2019 | 0–6 | 278.2–318.2 | 15 | Stainless steel reactors | 6 |
| Guo et al., 2016 | 1–6 | 313–353 | 30–40 | High pressure equilibrium equipment | 5 |
| Z. Liu et al., 2022 | 0.25–2.0 | 293.08–353.13 | 1.02–14.29 | Synthetic measurements | 9 |
| Yan et al., 2011 | 1.13–3.01 | 323.03–373.41 | 2.10–15.99 | Devices for “static analysis” methods | 5 |
| Zhaoa et al., 2015 | 1–5 | 273.15–473.15 | 10–40 | Quantitative Raman spectroscopy | 25 |

Experimental data on $CO_2$ solubility in brine systems published in the literature [21,27,30,44–46] were collected, covering various conditions of temperature (273.15–453.15 K), pressure (0.46–40 MPa), and salinity (0.25–6 mol/kg NaCl). Table 2 provides a summary of the experimental studies, including the pressures of the respective applications, the amount of data and temperature ranges employed, and the methods used. The dataset size (number of samples) is 240, and the unit of the output variable ($CO_2$ solubility) is the mole fraction of $CO_2$. After excluding outliers, the data were

divided into training, testing, and validation sets in a ratio of 16:4:5. The specific values in the table refer to Table A2.

### *2.2. Modeling technique*

This study utilized a multilayer perceptron (MLP) model to predict the solubility of $CO_2$ in both pure water and saline water. Although both models share a similar basic structure, adjustments were made to the design details to accommodate the differing tasks of solubility prediction.

Both models' input layers contain 3 neurons, representing temperature and pressure. The water model introduces a third variable, the interaction term between temperature and pressure, to better capture their complex nonlinear relationship [47]. This is important because the effects of temperature and pressure on gas solubility are typically not independent, and there may be interactions between them. By introducing interaction terms, this complex relationship can be more effectively modeled [40]. In contrast, the saline water model includes an additional important variable: salt concentration. The data are normalized using MinMaxScaler, which scales the value of each feature to between [0, 1]. The purpose of this is to make different features have the same magnitude and to avoid the impact of too-large values for some features on the model training. In the prediction of pure water solubility, solubility is primarily influenced by temperature and pressure, and is mainly determined by these two factors. Including irrelevant or low-impact features may not have practical significance and may actually reduce the interpretability of the model. Introducing additional features could increase the complexity of the model, but it does not necessarily provide useful extra information and may even lead to overfitting. Both models use ReLU activation functions in their hidden layers to enhance the nonlinear expressive capabilities of the model. Moreover, to prevent overfitting, dropout regularization was applied, and an early stopping mechanism was introduced to control the training process. Due to different physical and chemical complexities, the two models show significant differences in architecture and optimization strategies. The pure water solubility model is relatively simple, consisting of 3 hidden layers with 1024 neurons per layer. This simpler structure is sufficient to capture the effects of temperature and pressure on solubility. In contrast, the saline water solubility model features a deeper network with 5 hidden layers, where the number of neurons decreases progressively (from 1024 to 64), better capturing the complex nonlinear relationships between temperature, pressure, and salt concentration. To improve the model's generalization ability, the saline water model also incorporated random noise for data augmentation.

Regarding optimization strategies, the saline water model used a higher initial learning rate (0.001) combined with a learning rate decay mechanism to ensure the model converged quickly and underwent fine-tuning in the later stages. The pure water model, on the other hand, employed a lower initial learning rate (0.0001) without using learning rate decay, as its solubility variations were relatively simple, allowing the model to converge in fewer iterations. Additionally, the saline water model was set with a higher patience parameter (50) to ensure sufficient time for learning complex relationships, whereas the pure water model used a lower patience parameter to accelerate convergence.

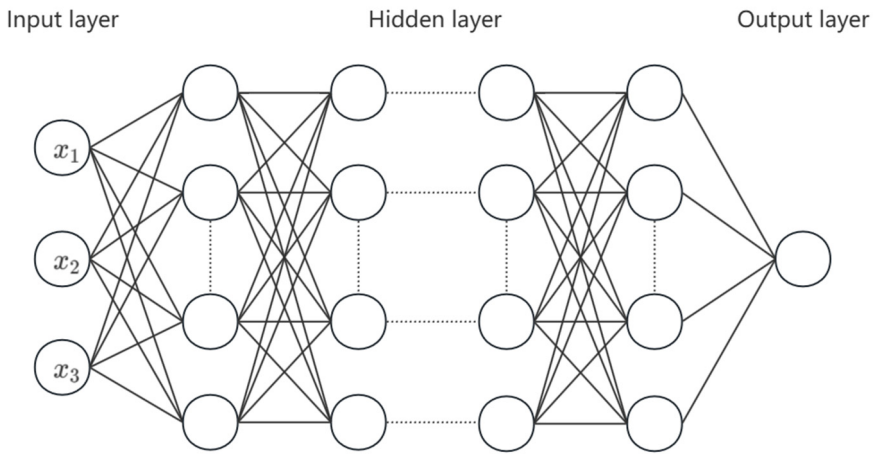



**Figure 1.** Illustration of the architecture of the MLP model.

The design of the MLP (multilayer perceptron) is based on a large number of interconnected perceptrons or neurons, as shown in Figure 1. These neurons are distributed among the following three types of layers:

Input Layer: The number of neurons depends on the number of input features ($x_1$, $x_2$, $x_3$). For example, if the input data includes temperature, pressure, and their interaction term, the input layer will have three neurons.

Hidden Layer: This is the core part of the model, consisting of one or more fully connected layers. The number of neurons and layers can be adjusted according to the complexity of the problem. Common activation functions include the ReLU (rectified linear unit), leaky ReLU, sigmoid, and tanh. The ReLU is the most commonly used activation function as it effectively introduces nonlinearity and avoids the vanishing gradient problem. However, it can cause the "dying ReLU" issue in the negative value region. Leaky ReLU addresses this limitation by allowing a small gradient in the negative region. Sigmoid and tanh are often used in scenarios where the output needs to be constrained within a specific range, but they may lead to the vanishing gradient problem.

Output Layer: The number of neurons depends on the specific task. For instance, regression tasks usually have one output neuron, while classification tasks have neurons corresponding to the number of classes. Optimization strategies for the MLP model include adjusting the number of hidden layers and neurons, learning rate adjustment, and using dropout regularization to prevent overfitting. Normalizing the input data and using adaptive learning rate optimizers like Adam can improve training efficiency and model performance.

It is worth noting that each neural connection between different layers is associated with a set of weights. Additionally, neurons in the output and hidden layers can include a bias term. The key to improving MLP model performance during the learning phase is to minimize the error between the target and the model prediction. The learning phase is considered a process of finding the optimal weights and biases. Many algorithms have been proposed to optimize the weights and biases of the MLP [48].

In this study, in terms of data processing and enhancement, the pure water model first normalizes the input features (such as temperature and pressure) using MinMaxScaler, compressing them into the [0,1] interval to ensure that the scale of each feature is consistent. Meanwhile, the target variable (solubility) is standardized using a StandardScaler and adjusted to a distribution with a mean of 0 and a standard deviation of 1. In feature engineering, the pure water model introduces a temperature pressure interaction term to capture the nonlinear relationship between the two and improve the accuracy of the model. On the contrary, the saltwater model adopts data-augmentation techniques to expand the diversity of training data by adding random noise (noise-factor) to input features, aiming to improve the robustness of the model, avoid overfitting, and enhance the generalization ability. The saltwater model also uses the MinMaxScaler for data normalization to ensure that the scale of input features (temperature, pressure, salinity) is consistent, which ensure efficient training of the neural network.

In terms of model structure, the neural network of the pure water model consists of three hidden layers, each layer containing 1024 neurons. The network structure is relatively large and aims to capture the complex nonlinear relationships between temperature, pressure, and their interaction terms. Dropout (0.2) layers were used after each hidden layer to reduce overfitting. The ReLU activation function is used for the hidden layer to accelerate training and avoid gradient vanishing problems, while the output layer uses a linear activation function because solubility is a continuous value. The pure water model has a relatively simple network structure, sufficient data volume, few features, and low nonlinear complexity, so the risk of overfitting is low. Therefore, the early stopping mechanism of overfitting was not used. However, the model still effectively reduces overfitting and enhances its generalization ability through the dropout (0.2 dropout rate). The neural network of the saltwater model consists of five hidden layers, with the number of neurons decreasing layer by layer (1024, 512, 256, 128, 64). This deeper structure helps to capture the more complex relationships between salinity, temperature, and pressure. The model adopts an Adam optimizer with attenuation to gradually reduce the learning rate during training, thereby helping the model converge more stably. In addition, the saltwater model uses an early stopping strategy to monitor the loss of the validation set. When the loss does not improve in the specified patch round, training will stop, effectively avoiding overfitting and reducing training time.

In terms of model training and evaluation, the pure water model had 2000 training rounds, a batch size of 32, and used validation_stplit = 0.2. 20% of the training data was used as the validation set to monitor the performance during the training process. The evaluation indicators include the root mean square error (RMSE), coefficient of determination ($R^2$), and mean absolute percentage relative error (AAPRE), which help to comprehensively evaluate the fitting effect of the model. The training rounds of the saltwater model are set to 1000 and an early stopping mechanism is used, so the actual number of training rounds may be less than 1000, depending on changes in the validation set loss. The saltwater model also uses RMSE, $R^2$, and AAPRE as evaluation criteria.

After training, both models were tested on the dataset. To evaluate the performance of the constructed neural network models in predicting $CO_2$ solubility, four common regression evaluation metrics were employed: root mean squared error (RMSE), coefficient of determination ($R^2$ score), average absolute percentage relative error (AAPRE)，and mean absolute error (MAE).

Average absolute percent relative error (AAPRE):

$$AAPRE = \frac{1}{N}\sum_{i=1}^{N}|((S_{iEXP} - S_{iPRED})/S_{iEXP}) \times 100| \quad (1)$$

Coefficient of determination ($R^2$):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(S_{iEXP}-S_{iPRED})^2}{\sum_{i=1}^{N}(S_{iEXP}-\overline{S_{iEXP}})^2} \quad (2)$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_{iEXP} - S_{iPRED})^2} \quad (3)$$

Mean absolute error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|S_{iEXP} - S_{iPRED}| \quad (4)$$

In the above equations, $S_{iEXP}$, $S_{iPRED}$,$\overline{S_{iEXP}}$,and $N$ refer to experimental solubility, predicted solubility, the mean of experimental solubilities, and the total number of data points, respectively.

The RMSE of the models for predicting $CO_2$ solubility was X, indicating a certain degree of error between predicted and actual values. The $R^2$ value was Y, suggesting that the model could explain Y% of the data variance, indicating that the model performed well in capturing the relationship between the input and target variables. The AAPRE was Z%, representing a relatively low average relative prediction error, further confirming the model's accuracy. By using RMSE, $R^2$, and AAPRE in combination, this study was able to assess the model's performance from multiple perspectives. RMSE provides an absolute measure of prediction error, $R^2$ evaluates the goodness of fit, and AAPRE offers a measure of prediction accuracy in relative error form. The combined use of these metrics allows a comprehensive evaluation of the model's predictive capabilities and provides a reference for further model optimization.

## 3. Results and disussions

### *3.1. ANN modeling results for $CO_2$ solubility in water*

#### 3.1.1. Model training results

Figure 2 shows the trend of $CO_2$ solubility with pressure at different temperatures (318.23 K, 333.15 K, and 353.15 K). As the pressure increases, solubility rises significantly, especially in the 0–30 MPa range where the solubility increases most rapidly. Above 30 MPa, the rate of solubility increase begins to slow down. The effect of temperature on solubility shows some dependence: at lower pressures, the solubility is slightly lower at higher temperatures (353.15 K), but as pressure increases, the solubility at all temperatures tends to converge. The experimental data fits well with the model predictions, indicating that the model accurately captures the trend of solubility changes with pressure.

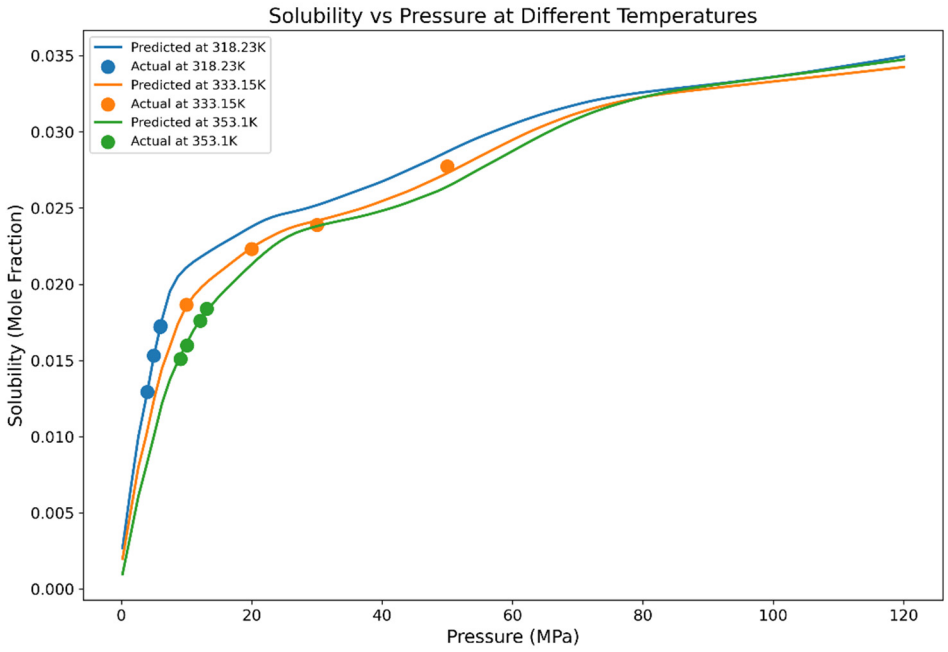



**Figure 2.** MLP predicted solubility vs. pressure curves at 318.23 K, 333.15 K, and 353.15 K.

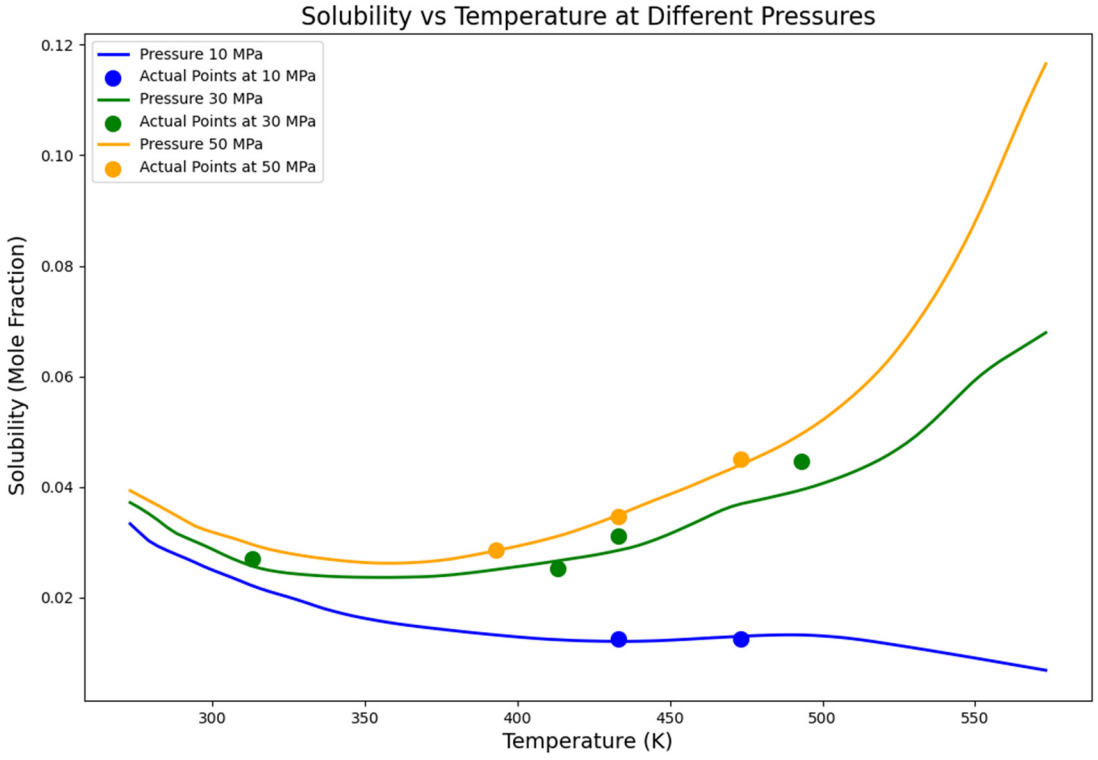



**Figure 3.** MLP predicted solubility vs. temperature curves at 10 MPa, 30 MPa, and 50 Mpa.

Figure 3 presents the relationship between $CO_2$ solubility and temperature at different pressures (10 MPa, 30 MPa, and 50 MPa). Under low-pressure conditions (10 MPa), $CO_2$ solubility decreases significantly with increasing temperature, especially between 300 K and 400 K, where temperature has a notable negative effect on solubility. However, under medium- and high-pressure conditions (30 MPa and 50 MPa), the solubility initially decreases with rising temperature but then begins to increase, especially when the temperature exceeds 450 K. This suggests that under high pressure, increasing temperature can significantly enhance solubility, particularly at very high temperatures and pressures, where the positive impact of temperature on solubility becomes most pronounced.

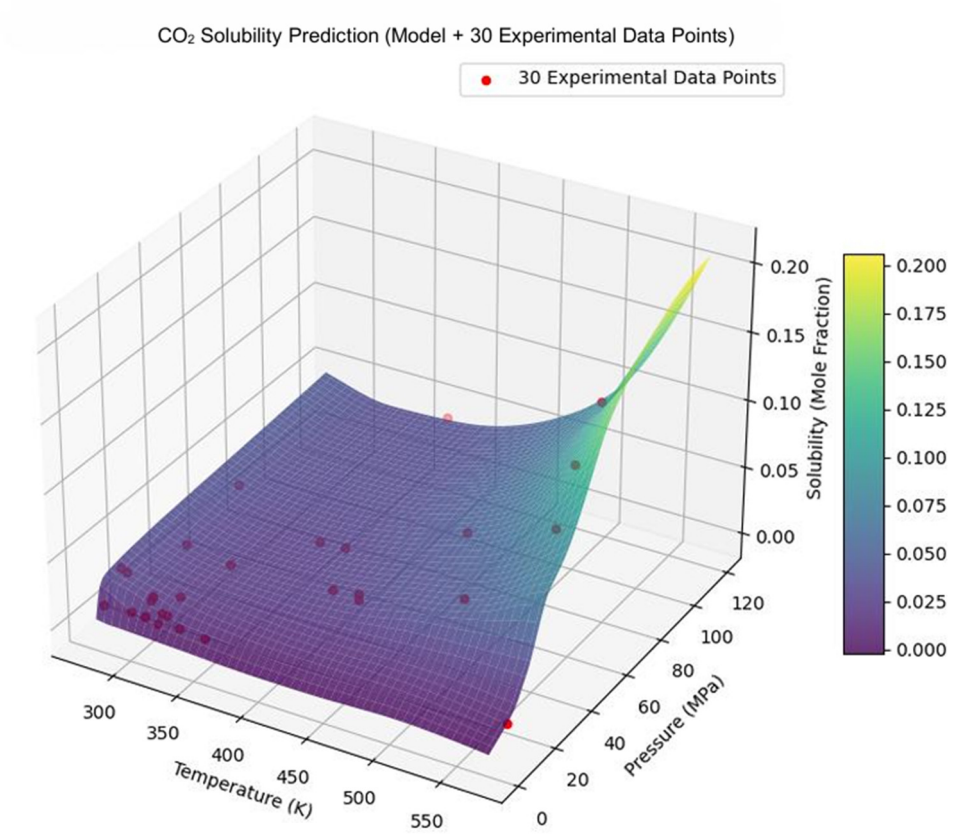



**Figure 4.** MLP predicted solubility in a 3D plot of pressure and temperature.

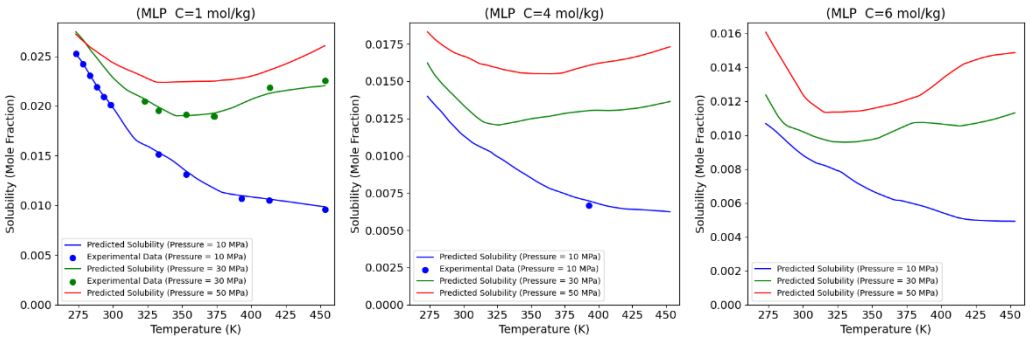
Figures 2–3 illustrate the nonlinear relationship between $CO_2$ solubility, pressure, and temperature. As pressure increases, especially under high-pressure conditions near 120 MPa, solubility increases significantly, whereas at lower pressures (below 20 MPa), the solubility increase is relatively gradual. Additionally, the impact of temperature on solubility depends on the pressure level. At low pressures (below 20 MPa), increasing temperature slightly decreases solubility, while at high pressures (above 80 MPa), increasing temperature leads to a significant rise in solubility. This indicates that the positive effect of temperature becomes more pronounced under high-pressure conditions, particularly when temperatures exceed 500 K, where solubility increases rapidly. The combination of these three figures further reveals that $CO_2$ solubility increases with rising pressure and that under high-pressure and high-

temperature conditions, solubility increases significantly. In contrast, under low-pressure conditions, increasing temperature generally causes a decrease in solubility. The trends depicted in both the 3D and 2D graphs are consistent in showing the changes in $CO_2$ solubility.

Figure 4 presents a three-dimensional visualization illustrating the relationship between $CO_2$ solubility in pure water, temperature, and pressure, as predicted by the MLP model. The figure clearly depicts the notable nonlinear effects of pressure and temperature on $CO_2$ solubility. Specifically, as pressure increases, $CO_2$ solubility significantly rises, particularly under higher-pressure conditions approaching 120 MPa, whereas the increase in solubility is relatively gradual at lower pressures (below 20 MPa). Additionally, the influence of temperature on $CO_2$ solubility varies depending on the pressure levels. At lower pressures (below 20 MPa), increasing temperature generally leads to a slight reduction in $CO_2$ solubility. In contrast, at higher pressures (above 80 MPa), increasing temperature significantly enhances solubility, especially when the temperature exceeds approximately 500 K. Overall, the plot demonstrates that $CO_2$ solubility increases markedly with pressure, and the impact of temperature is complex, reducing solubility at lower pressures but substantially increasing it under high-pressure conditions.

### *3.2. ANN modeling results for $CO_2$ solubility in brine*

#### 3.2.1. Model training results



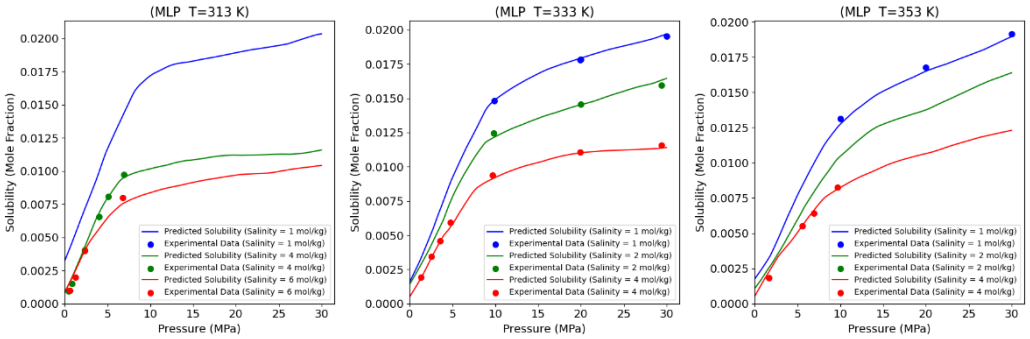

(a) 1 mol/kg (b) 4 mol/kg (c) 6 mol/kg

**Figure 5.** Temperature dependent curves for predicting solubility of the MLP at different salinities.

Figure 5 illustrate the variation of $CO_2$ solubility with temperature across different salinity levels. Under all salinity conditions, $CO_2$ solubility increases with rising pressure. At lower pressures (10 MPa), an increase in temperature generally leads to a decrease in solubility. Conversely, at higher pressures (50 MPa), solubility first decreases and then increases with the rising temperature. This trend is consistent across all three salinity conditions. Under low-pressure conditions of 10 MPa, solubility significantly decreases with increasing temperature, with the most pronounced decrease observed at a salinity of 1 mol/kg. At 30 MPa, the declining trend in solubility due to the rising temperature gradually weakens and may even level off or slightly increase at higher temperatures. At 50 MPa, the effect of

temperature on solubility becomes more complex, with solubility gradually increasing at elevated temperatures.

As salinity (NaCl) increases from 1 mol/kg to 6 mol/kg, the overall level of $CO_2$ solubility decreases significantly. Even under the same temperature and pressure conditions, solubility is notably lower at higher salinity levels compared to lower ones. Under low-pressure conditions (10 MPa), an increase in temperature leads to a significant decrease in solubility. In contrast, under high-pressure conditions (50 MPa), rising temperatures may instead lead to an increase in solubility. This phenomenon is observed across different salinity conditions, indicating that the effects of temperature and pressure on $CO_2$ solubility are interdependent.

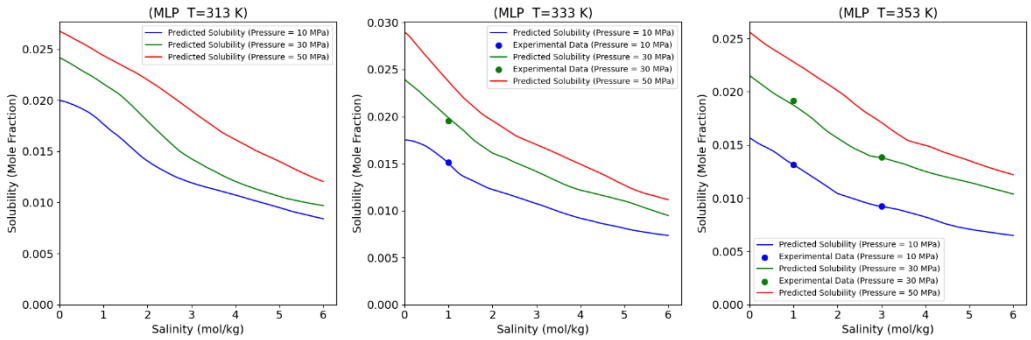



(a) 313 K (b) 333 K (c) 353 K

**Figure 6.** Prediction of the $CO_2$-NaCl brine solubility variation curve of the MLP with pressure at different temperatures.

Figure 6 depict the relationship between $CO_2$ solubility and pressure at temperatures of 313 K, 333 K, and 353 K, respectively. Across all these temperatures, $CO_2$ solubility increases with rising pressure. However, the rate at which solubility increases diminishes as pressure continues to rise. Notably, in the low-pressure range (0–10 MPa), solubility increases most rapidly. In the 10–20 MPa range, the growth rate of solubility slows down, and in the 20–30 MPa range, the increment becomes even smaller. This suggests that the impact of pressure on solubility is more pronounced at lower pressures, with the effect approaching saturation as pressure increases.

For instance, Figure 6(b) shows that as salinity increases, the solubility curve rises gradually at various pressures, but the rate of increase slows down progressively. This indicates that while higher pressures continue to enhance $CO_2$ solubility, the incremental benefits decrease as pressure further escalates, reflecting a diminishing return effect.




(a) 313 K (b) 333 K (c) 353 K

**Figure 7.** Curve of MLP-predicted $CO_2$-NaCl brine solubility with salinity variation at different temperatures.

Figure 7 illustrate the trend of $CO_2$ solubility as salinity increases. Overall, $CO_2$ solubility decreases significantly with rising salinity, and the magnitude of this reduction diminishes as salinity continues to increase. Under low salinity conditions (1–2 mol/kg), the decrease in solubility is more pronounced, whereas at higher salinities (4 mol/kg and above), the decline in solubility tends to stabilize.

For example, Figure 7(c) shows that at pressures of 10 MPa, 30 MPa, and 50 MPa, solubility gradually decreases as salinity increases, but the rate of decline slows down. This indicates that the inhibitory effect of high salinity on $CO_2$ solubility begins to stabilize at elevated salinity levels, suggesting that the negative impact of salinity on solubility reaches a point of diminishing returns at higher salinities.

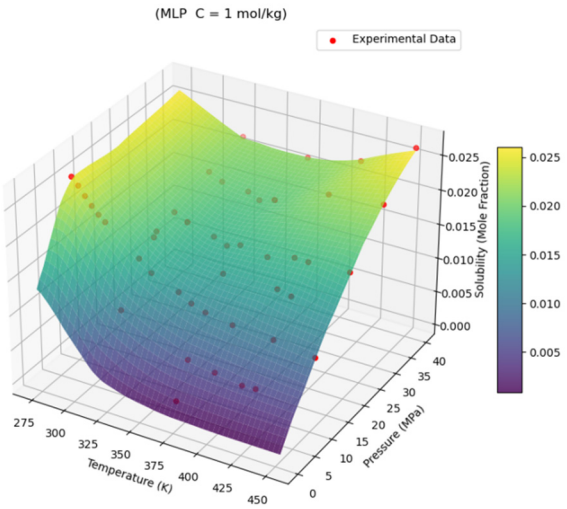



**Figure 8.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 1 mol/kg of NaCl brine.

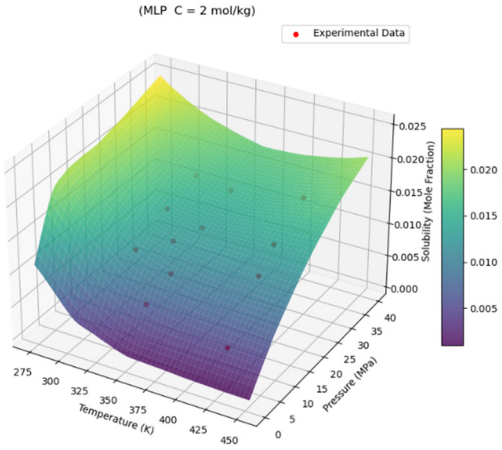



**Figure 9.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 2 mol/kg of NaCl brine.

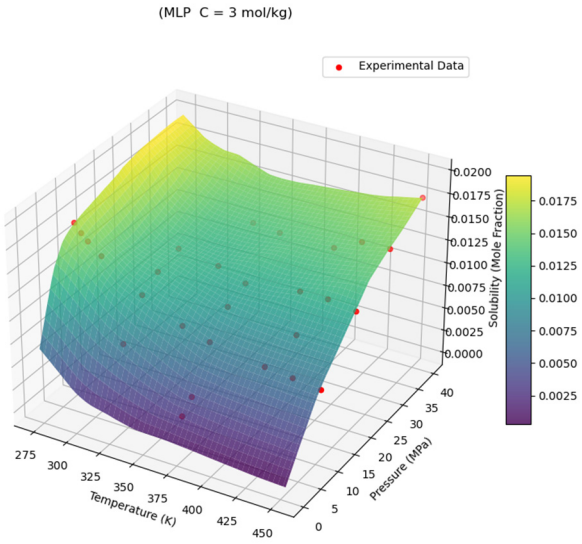



**Figure 10.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 3 mol/kg of NaCl brine.

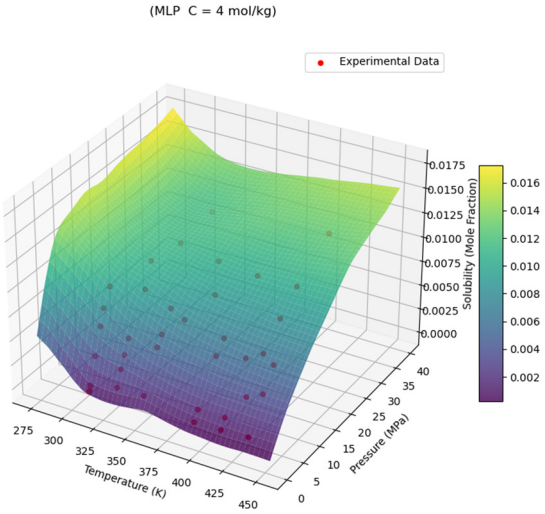



**Figure 11.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 4 mol/kg of NaCl brine.

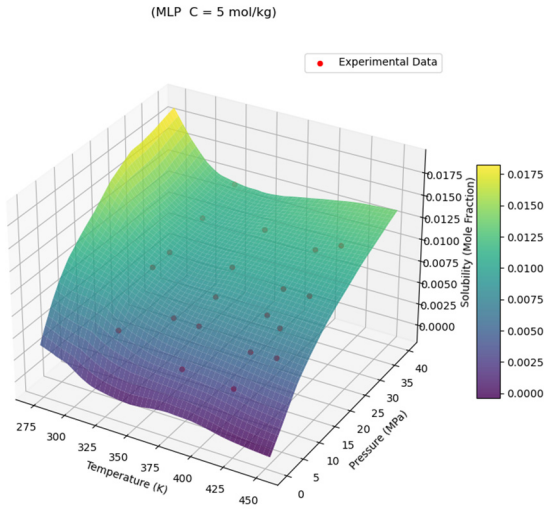



**Figure 12.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 5 mol/kg of NaCl brine.

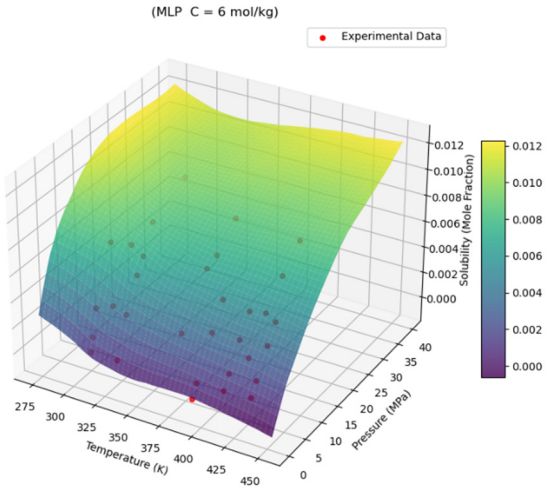



**Figure 13.** Three-dimensional relationship diagram of MLP-predicted solubility with temperature and pressure at a salinity of 6 mol/kg of NaCl brine.

Figures 8–13 depict the relationship between $CO_2$ solubility, temperature, and pressure under varying salinity conditions. Across all salinity levels, increased pressure consistently leads to higher $CO_2$ solubility. However, as salinity increases, the enhancing effect of pressure on solubility becomes less pronounced. At low pressures, rising temperatures typically result in a decrease in solubility, whereas at high pressures, increasing temperatures may cause solubility to increase. This effect is more noticeable at lower salinity levels, while at higher salinities, the positive impact of temperature on solubility is diminished.

As salinity increases, $CO_2$ solubility significantly declines. Higher salinity not only reduces the overall solubility levels but also attenuates the combined effects of temperature and pressure on solubility, especially under high-pressure and high-temperature conditions, where the impact of salinity becomes more pronounced. This consistency between the 3D and 2D plots confirms that the observed trends in $CO_2$ solubility changes are reliably depicted in both types of visualizations. The trends shown in the 3D plots align well with the phenomena observed in the 2D plots, with both collectively revealing the patterns of $CO_2$ solubility variation under different conditions.

### 3.2.2. Model predictive performance

The MLP nonlinear regression model was utilized to capture the complex relationships between $CO_2$ solubility and the independent variables, allowing for the prediction and visualization of these relationships through the charts. The presented figures clearly demonstrate the generalization capability of the trained model on unseen data and how temperature, pressure, and salinity influence $CO_2$ solubility.

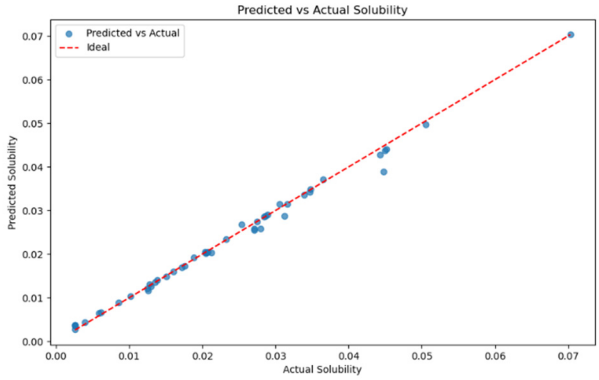



**Figure 14.** Scatter plot of MLP-predicted water solubility vs. experimental data in the validation set.

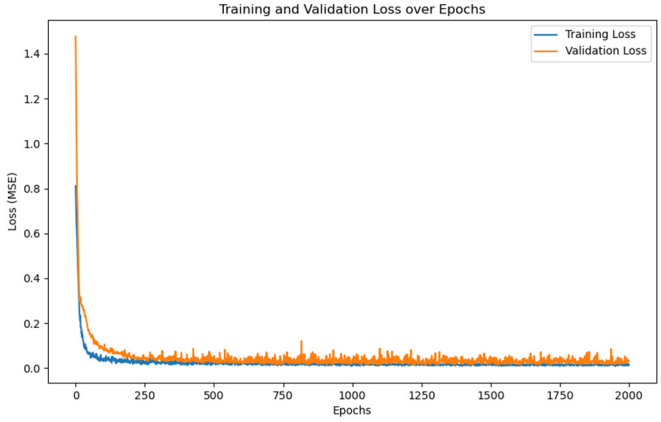



**Figure 15.** Changes in the loss function of the MLP model with training times.

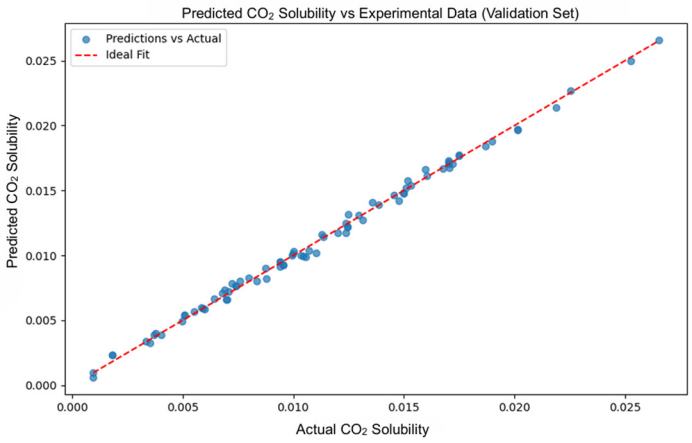



**Figure 16.** Scatter plot of MLP-predicted brine solubility vs. experimental data in the validation set.

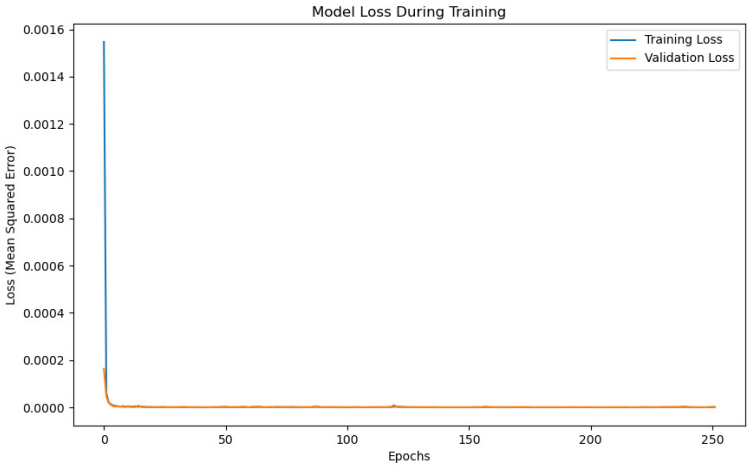



**Figure 17.** Changes in the loss function of the MLP model with training times.

**Table 3.** Performance evaluation of the models.

| | | AAPRE | RMSE | $R^2$ |
|---|---|---|---|---|
| MLP-Water | Training | 4.77416 | 0.00286 | 0.98587 |
| | Testing | 3.44168 | 0.00114 | 0.99362 |
| | Total | 4.11874 | 0.00108 | 0.99424 |

**Table 4.** Performance evaluation of the models.

| | | AAPRE | RMSE | $R^2$ |
|---|---|---|---|---|
| MLP-Brine | Training | 5.76440 | 0.00027 | 0.99751 |
| | Testing | 4.05568 | 0.00045 | 0.99472 |
| | Total | 4.91004 | 0.00036 | 0.99612 |

Figures 14 and 16, respectively, show the comparison between the predicted solubility and experimental values for the pure water and brine models in the validation set based on the MLP. These predicted validation set data can be obtained from Table A3 and Table A4. It can be observed that the predicted values of the water MLP model are closely distributed along the diagonal line, with most prediction points closely adhering to the ideal fit line, indicating small discrepancies between the predicted and experimental values. This demonstrates the model's strong predictive performance. Similarly, the predicted values of the brine MLP model are also closely distributed along the diagonal, with most points tightly following the ideal fit line, showing minimal differences between the predictions and experimental values.

Tables 3 and 4 report the performance evaluation of both models, indicating that the proposed machine learning models have achieved excellent predictive performance, as reflected by high R-squared values and low root mean square errors throughout the paradigm.

Figures 15 and 17 display the convergence processes of training and validation losses for $CO_2$ solubility models in water and brine, respectively. In Figure 6, which shows the water model, the loss

decreases at a relatively slower rate, especially within the first 200 epochs, with both validation and training losses gradually converging. However, the validation loss exhibits significant fluctuations, possibly due to the model containing many neurons, making it sensitive to noise during training, which causes these variations. The training loss is slightly lower than the validation loss, suggesting mild overfitting, although the predictive performance remains good.

In Figure 17, which shows the brine model, both training and validation losses rapidly decrease within the initial few epochs, reaching very low values in just a few training rounds, indicating quick convergence of the model. This phenomenon suggests that the brine model quickly learns the features within the data, and the consistency between training and validation performance indicates strong generalization ability without evident overfitting or underfitting issues.

In summary, the brine model exhibits rapid convergence, with stable loss curves and validation loss closely matching the training loss, suggesting that the model effectively captures the complex nonlinear relationships in brine. In contrast, the water model, with its relatively simpler input features (only temperature and pressure), despite including temperature-pressure interaction terms, shows more fluctuations when faced with a deeper neural network structure. The stability during the training process is somewhat lower, but the predictive performance still reflects a relatively low average error percentage, indicating accurate predictions.

### *3.3. Discussion*

#### 3.3.1. The improper expansion of the $CO_2$-brine solubility prediction model to the $CO_2$-water case

This section is used to compare the difference between predicting 0 salinity and pure water, and verify whether the 0 salinity model can be used to predict pure water.

The salinity variable (set to zero) was introduced into the pure water data to meet the three-variable requirement of the brine model. These newly generated data were then applied to the brine prediction model, and the prediction results are shown in Figures 18 and 19.

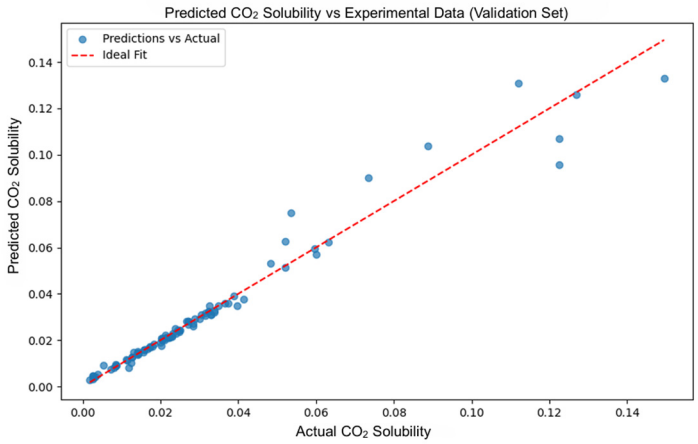



**Figure 18.** Scatter plot of MLP-predicted solubility and experimental data distribution in the validation set.

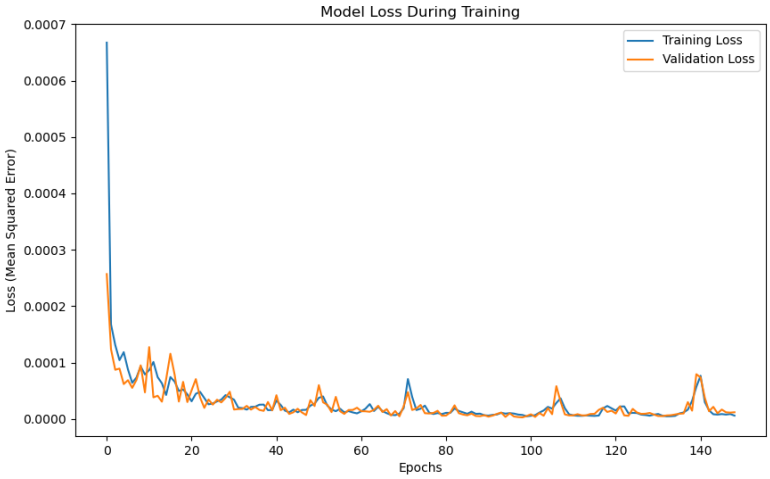



**Figure 19.** Changes in the loss function of the MLP model with training times.

**Table 5.** Performance evaluation of pure water data in brine modeling.

| | AAPRE | RMSE | $R^2$ |
|---|---|---|---|
| Training | 6.64132 | 0.00353 | 0.97179 |
| Testing | 7.66653 | 0.00448 | 0.97400 |
| Total | 7.15392 | 0.00401 | 0.97289 |

In Figure 18, most data points are concentrated near the red dashed line, especially in the lower $CO_2$ solubility range, where the model's predictions closely match the actual values, demonstrating good performance. However, as the solubility values increase, particularly in the higher $CO_2$ solubility range (greater than 0.06), some prediction points deviate from the ideal line, indicating that the model's predictive performance is poor at high solubility levels, with noticeable bias. Table 5 shows that the $R^2$ value is small relative to the brine model, and the large AAPRE leads to large errors and poor prediction performance, which leads to the preliminary determination that the brine model is not applicable to the prediction of solubility at salinity 0.

In Figure 19, the training loss and validation loss initially decrease rapidly, showing that the model can quickly fit some features of the training data. However, as training progresses, the loss values level off at a relatively low level but still exhibit some fluctuations, particularly in later epochs, where losses increase, indicating the model's limited generalization capability in aqueous solutions. This suggests that while the model is well-trained in saline environments, it fails to effectively capture the variation of $CO_2$ solubility in pure water, leading to persistent errors. This further confirms the model's inadequate adaptability to pure water environments.

Although the brine model performs well within its training environment (brine), in aqueous solutions, it lacks proper modeling of the physical and chemical mechanisms governing $CO_2$ solubility in pure water, resulting in losses that do not effectively converge, indicating insufficient generalization ability. On the other hand, models specifically trained for aqueous environments show lower losses and better generalization capabilities. Physically, the solubility of $CO_2$ in water and brine is influenced by different mechanisms, with the presence of salt ions significantly altering $CO_2$ solubility behavior.

Therefore, a unified prediction using the same machine learning model for both environments is not feasible.

From the detailed descriptions provided earlier, we have clearly understood that structural differences not only affect the training complexity of the models but also have a significant impact on their generalization ability and prediction performance. In the following section, the focus will be on comparing the performance differences between the two models, especially in assessing whether the saline water model can accurately predict the solubility in pure water when the salt concentration is zero. This will further explore why different model designs are needed for different environments.

By measuring errors (such as RMSE and $R^2$), the differences in model adaptability to various solution environments can be quantified. Analysis based on these metrics shows that the brine model performs exceptionally well in saline environments, with RMSE and $R^2$ values indicating that its predictions closely match the actual conditions in brine. However, when this model is used to predict $CO_2$ solubility in water (i.e., at zero salinity), its performance deteriorates significantly, with increased prediction errors, especially in regions of high solubility. In contrast, models specifically designed for aqueous solutions exhibit better predictive accuracy in pure water, with lower errors and $R^2$ values close to 1, indicating a high degree of agreement between predicted and actual values. These results further confirm that the brine and water models have different adaptabilities and cannot be used interchangeably.

The predictive models suggest that pure water can dissolve a substantial amount of $CO_2$, with solubility increasing rapidly as pressure rises and being more pronounced at lower temperatures. This is attributed to the absence of ionic interferences in pure water, which facilitates the dissolution of $CO_2$ molecules. Conversely, the solubility of $CO_2$ in brine is considerably lower than in pure water due to the salting-out effect, where higher salinity levels hinder the solubility of $CO_2$. As salinity escalates, ions present in the water, such as $Na^+$ and $Cl^-$, compete for positions that $CO_2$ would otherwise occupy to interact with water molecules, thereby diminishing $CO_2$ solubility. This effect becomes especially noticeable at elevated salinity levels, and with increasing pressure, the enhancement of $CO_2$ solubility in brine is less significant compared to that in pure water.

### 3.3.2. Differences supported by physical and chemical principles

Even when the salinity is nominally zero in the brine model, the predicted $CO_2$ solubility may still deviate slightly from the pure water model. This discrepancy can be attributed to two factors: (1) Parameter discrepancies: The parameters of the brine model are calibrated to account for the influence of salinity on solubility. Even at zero salinity, these parameters may differ from those of the pure water model, resulting in minor prediction variations. (2) Residual ionic influences: In computational models, properties related to the presence of ions in brine might persist even when salinity is set to zero, leading to minor deviations from the pure water model. This could encompass factors such as the activity coefficient of water and underlying effects in ion interaction models. These subtle differences underscore the complexity of accurately modeling $CO_2$ solubility in various aqueous environments.

From a physico-chemical perspective, on the other hand, the mechanisms governing $CO_2$ solubility in water and brine are fundamentally different. In water, $CO_2$ molecules primarily form hydrogen bonds with water molecules and are influenced by van der Waals forces, resulting in a

relatively linear relationship between solubility and changes in temperature and pressure [49,50]. In brine, ions such as $Na^+$ and $Cl^-$ alter the interactions between water and $CO_2$ through solvation and shielding effects, leading to the salting-out effect [49], where $CO_2$ solubility decreases as salinity increases. The salting-out effect refers to the phenomenon where the solubility of a solute, particularly a gas, decreases when salts are added to a solution. This effect typically occurs in solutions containing dissolved gases, nonpolar solutes, or organic solutes. When salt is added to a solution, the ions from the salt form strong interactions with water molecules, reducing the freedom of movement of the water molecules. As a result, the interaction between water molecules and gas molecules is diminished, leading to a decrease in the solubility of the gas.

When salts dissolve in water, they dissociate into ions, creating an ionic environment that affects interactions between solvent and solute molecules. In the salting-out effect, the presence of ions weakens interactions between the solute and water molecules, reducing $CO_2$ solubility by altering the solvation free energy of the solute, making it less favorable [51]. Consequently, $CO_2$ solubility in brine exhibits more complex nonlinear variations with temperature, pressure, and salinity. Due to these differing solubility mechanisms and significant differences in physical and chemical processes, features learned by the model in brine cannot effectively generalize to aqueous solutions, and vice versa. Combining experimental and theoretical studies, these differences support the conclusion that brine and water models cannot share the same MLP model.

#### 3.3.3. Comparative analysis of MLP and CatBoost

We found that Yang [52] no longer focused on studying black box models, so we introduce a new model, called CatBoost, and compare it with those models. The prediction results are shown in the Table 6:

**Table 6.** Performance evaluation of CatBoost and MLP models in pure water and brine modeling.

| Model | Data point | Testing set | | | |
|---|---|---|---|---|---|
| | | $R^2$ | AAPRE | RMSE | MAE |
| CatBoost-Water | 265 | 0.93524 | 11.88348 | 0.00635 | 0.00286 |
| MLP-Water | 265 | 0.99424 | 4.11874 | 0.00108 | 0.00159 |
| CatBoost-Brine | 240 | 0.97867 | 16.94977 | 0.00072 | 0.00057 |
| MLP-Brine | 240 | 0.99612 | 4.91004 | 0.00036 | 0.00031 |

As can be seen from the table above, the predictions of the CatBoost model do not achieve the accuracy of the MLP, and the MLP's prediction results are more reliable.

## 4. Conclusions

This study investigated MLP artificial neural networks for the $CO_2 + H_2O$ and $CO_2 + H_2O + NaCl$ systems. Model accuracy was validated through comparison with literature data. The $CO_2$ solubility prediction model in water achieved an RMSE of 0.00108, $R^2$ of 0.99424, and AAPRE of 4.11874%. The

$CO_2$ solubility prediction model in brine achieved an RMSE of 0.00041, $R^2$ of 0.99473, and AAPRE of 4.7243%. Both models demonstrated commendable predictive abilities. Testing whether the brine model could predict $CO_2$ solubility in pure water (salinity of 0) revealed an RMSE of 0.00548, $R^2$ of 0.96118, and AAPRE of 9.59813% for the brine model, with a lower $R^2$ and an average error approaching 10%, indicating higher errors and reduced model performance. Therefore, the brine model is unsuitable for predicting $CO_2$ solubility in pure water, as there are notable differences between the two.

Several key differences exist between the solubility models of $CO_2$ in water and brine. Regarding neural network architecture, the water model uses a three-input layer configuration, including temperature, pressure, and their interaction terms, focusing on optimizing data preprocessing (e.g., feature engineering). In contrast, the brine model adds "salinity" as a variable and adopts a more complex structure and regularization strategies (e.g., data augmentation and additional hidden layers) to better fit the solubility changes in brine.

Experimental data further confirmed that when predicting solubility in water by setting salinity to zero in the brine model, the results were inferior to those of the pure water model. This is because the physico-chemical mechanisms of the two are different; the solubility mechanism of $CO_2$ in pure water is simpler, while the brine model considers complex factors such as the salting-out effect. Even with zero salinity, residual effects within the brine model can influence predictions.

## Author contributions

Shuo Yang: Investigation, Formal analysis, Writing–Original draft; Dong Wang: Investigation, Formal analysis, Writing–Reviewing and editing; Zeguang Dong: Formal analysis; Yingge Li: Conceptualization; Dongxing Du: Conceptualization, Methodology, Writing–Reviewing and editing.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

The paper is supported by the Graduate Tutor Foundation of Qingdao University of Science and Technology (120202190414).

## Conflict of interest

The authors declare there are no conflicts of interest regarding the publication of this article.

## References

1. Rogelj J, Den Elzen M, Höhne N, et al. (2016) Paris Agreement climate proposals need a boost to keep warming well below 2°C. *Nature*: 631–639. https://doi.org/10.1038/nature18307

2. Du D, Sun S, Zhang N, et al. (2015) Pressure distribution measurements for $CO_2$ foam flow in porous media. *J Porous Media*: 1119–1126. https://doi.org/10.1615/JPorMedia.2015012151
3. Du D, Zhang N, Li Y, et al. (2017) Parametric studies on foam displacement behavior in a layered heterogeneous porous media based on the stochastic population balance model. *J Nat Gas Sci Eng* 48: 1–12. https://doi.org/10.1016/j.jngse.2017.08.035
4. Du D, Zhang D, Li Y, et al. (2019) Numerical investigations on the inlet and outlet behavior of foam flow process in porous media using stochastic bubble population balance model. *J Pet Sci Eng* 176:537–553. https://doi.org/10.1016/j.petrol.2019.01.073
5. Du D, Zhang X, Yu K, et al. (2020) Parameter Screening Study for Optimizing the Static Properties of Nanoparticle-Stabilized $CO_2$ Foam Based on Orthogonal Experimental Design. *ACS Omega* 5: 4014–4023. https://doi.org/10.1021/acsomega.9b03543
6. Du D, Zhao D, Li Y, et al. (2021) Parameter calibration of the stochastic bubble population balance model for predicting NP-stabilized foam flow characteristics in porous media. *Colloids Surf A* 614: 126180. https://doi.org/10.1016/j.colsurfa.2021.126180
7. Li Y, Zhao D, Du D (2022) Computational study on the three phase displacement characteristics of foam fluids in porous media. *J Pet Sci Eng* 215: 110732. https://doi.org/10.1016/j.petrol.2022.110732
8. Mutailipu M, Song Y, Yao Q, et al. (2024) Solubility and interfacial tension models for $CO_2$–brine systems under $CO_2$ geological storage conditions. *Fuel* 357: 129712. https://doi.org/10.1016/j.fuel.2023.129712
9. Du D, Zhang X, Wan C, et al. (2021) Determination of the effective thermal conductivity of the porous media based on digital rock physics. *Geothermics* 97: 102267. https://doi.org/10.1016/j.geothermics.2021.102267
10. Eyitayo S, Arbad N, Okere C, et al. (2025) Advancing geological storage of carbon dioxide ($CO_2$) with emerging technologies for climate change mitigation. *Int J Environ Sci Technol* 22: 5023–5056. https://doi.org/10.1007/s13762-024-06074-w
11. Liu Y, Hu T, Rui Z, et al. (2023). An Integrated Framework for Geothermal Energy Storage with $CO_2$ Sequestration and Utilization. *Engineering*: 121–130. https://doi.org/10.1016/j.eng.2022.12.010
12. Wu Y, Li P (2020) The potential of coupled carbon storage and geothermal extraction in a $CO_2$-enhanced geothermal system: a review. *Geotherm Energy* 8. https://doi.org/10.1186/s40517-020-00173-w
13. Xu Z, Zhao H, Fan L, et al. (2024) A literature review of using supercritical $CO_2$ for geothermal energy extraction: Potential, methods, challenges, and perspectives. *Renew Energ Focus* 51: 100637. https://doi.org/10.1016/j.ref.2024.100637
14. Cui P, Liu Z, Cui X, et al. (2023) Impact of water on miscibility characteristics of the $CO_2$/n-hexadecane system using the pendant drop shape analysis method. *Arabian J Chem* 16: 105038. https://doi.org/10.1016/j.arabjc.2023.105038
15. Liu Z, Yan S, Zang H, et al. (2023) Quantization of the water presence effect on the diffusion coefficients of the $CO_2$/oil system with the dynamic pendant drop volume analysis technique. *Chem Eng Sci* 281: 119142. https://doi.org/10.1016/j.ces.2023.119142
16. Hu J, Duan Z, Zhu C, et al. (2007) PVTx properties of the $CO_2$-$H_2O$ and $CO_2$-$H_2O$-NaCl systems below 647 K: Assessment of experimental data and thermodynamic models. *Chemical Geology*: 249–267. https://doi.org/10.1016/j.chemgeo.2006.11.011
17. Ruth LA, Squires AM, Graff RA, et al. (1991) Desulfurization of Fuels With Calcined Dolomite. 1. Introduction and First Kinetic Results. *Chem Soc Diu Fuel Chem* 30: 67.

18. Schmidt C, Bodnar RJ (2000) Synthetic fluid inclusions: XVI. PVTX properties in the system $H_2O$-NaCl-$CO_2$ at elevated temperatures, pressures, and salinites. *Geochim Cosmochim Acta* 64: 3853–3869. https://doi.org/10.1016/S0016-7037(00)00471-3
19. Wang X, Cui X, Wang F, et al. (2021) Miscibility characteristics of the $CO_2$/n-hexadecane system with presence of water component based on the phase equilibrium calculation on the interface region. *Colloids Surf A* 629: 127463. https://doi.org/10.1016/j.colsurfa.2021.127463
20. Valtz A, Chapoy A, Coquelet C, et al. (2004) Vapour-liquid equilibria in the carbon dioxide-water system, measurement and modelling from 278.2 to 318.2 K. *Fluid Phase Equilib* 226: 333–344. https://doi.org/10.1016/j.fluid.2004.10.013
21. Carvalho PJ, Pereira LMC, Gonçalves NPF, et al. (2015) Carbon dioxide solubility in aqueous solutions of NaCl: Measurements and modeling with electrolyte equations of state. *Fluid Phase Equilib* 388: 100–106. https://doi.org/10.1016/j.fluid.2014.12.043
22. Guo H, Chen Y, Hu Q, et al. (2014) Quantitative Raman spectroscopic investigation of geo-fluids high-pressure phase equilibria: Part I. Accurate calibration and determination of $CO_2$ solubility in water from 273.15 to 573.15K and from 10 to 120MPa. *Fluid Phase Equilib* 382: 70–79. https://doi.org/10.1016/j.fluid.2014.08.032
23. Talebian SH, Masoudi R, Tan IM, et al. (2014) Foam assisted $CO_2$-EOR: A review of concept, challenges, and future prospects. *J Pet Sci Eng* 120: 202–215. https://doi.org/10.1016/j.petrol.2014.05.013
24. Zhao H, Lvov SN (2016) Phase behavior of the $CO_2$-$H_2O$ system at temperatures of 273-623 K and pressures of 0.1-200 MPa using Peng-Robinson-Stryjek-Vera equation of state with a modified Wong-Sandler mixing rule: An extension to the $CO_2$-$CH_4$-$H_2O$ system. *Fluid Phase Equilib* 417: 96–108. https://doi.org/10.1016/j.fluid.2016.02.027
25. Cui X, Zheng L, Liu Z, et al. (2022). Determination of the Minimum Miscibility Pressure of the $CO_2$/oil system based on quantification of the oil droplet volume reduction behavior. *Colloids Surf A* 653: 130058. https://doi.org/10.1016/j.colsurfa.2022.130058
26. Peridas G, Mordick Schmidt B (2021) The role of carbon capture and storage in the race to carbon neutrality. *Electr J* 34: 106996. https://doi.org/10.1016/j.tej.2021.106996
27. Zhaoa H, Fedkin MV, Dilmore RM, et al. (2015) Carbon dioxide solubility in aqueous solutions of sodium chloride at geological conditions: Experimental results at 323.15, 373.15, and 423.15K and 150bar and modeling up to 573.15K and 2000bar. *Geochim Cosmochim Acta* 149: 165–189. https://doi.org/10.1016/j.gca.2014.11.004
28. Duan Z, Sun R, Zhu C, et al. (2006) An improved model for the calculation of $CO_2$ solubility in aqueous solutions containing $Na^+$, $K^+$, $Ca^{2+}$, $Mg^{2+}$, $Cl^-$, and $SO_4^{2-}$. *Mar Chem* 98: 131–139. https://doi.org/10.1016/j.marchem.2005.09.001
29. Spycher N, Pruess K, Ennis-King J (2003) $CO_2$-$H_2O$ mixtures in the geological sequestration of $CO_2$. I. Assessment and calculation of mutual solubilities from 12 to 100 °C and up to 600 bar. *Geochim Cosmochim Acta* 67: 3015–3031. https://doi.org/10.1016/S0016-7037(03)00273-4
30. Liu Z, Cui P, Cui X, et al. (2022) Prediction of $CO_2$ solubility in NaCl brine under geological conditions with an improved binary interaction parameter in the Søreide-Whitson model. *Geothermics* 105: 102544. https://doi.org/10.1016/j.geothermics.2022.102544
31. Schmidhuber J (2015) Deep Learning in neural networks: An overview. *Neural Networks*: 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
32. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521: 436–444. https://doi.org/10.1038/nature14539

33. Vanneschi L, Castelli M (2018) Multilayer perceptrons. *Encycl Bioinf Comput Biol* 1: 612–620. https://doi.org/10.1016/B978-0-12-809633-8.20339-7
34. Rumelhart DE, Hinton GE, Williams RJ (2019) Learning Representations by Back-Propagating Errors. *Cognitive Modeling*: 213–222. https://doi.org/10.7551/mitpress/1888.003.0013
35. Goel A, Goel AK, Kumar A (2023) The role of artificial neural network and machine learning in utilizing spatial information. *Spat Inf Res* 31: 275–285. https://doi.org/10.1007/s41324-022-00494-x
36. Kim H (2022) Deep Learning. In: *Artificial Intelligence for 6G*, 22: 247–303. https://doi.org/10.1007/978-3-030-95041-5_6
37. Madani SA, Mohammadi MR, Atashrouz S, et al. (2021) Modeling of nitrogen solubility in normal alkanes using machine learning methods compared with cubic and PC-SAFT equations of state. *Sci Rep* 11: 24403. https://doi.org/10.1038/s41598-021-03643-8
38. Mohammadi MR, Hadavimoghaddam F, Pourmahdi M, et al. (2021) Modeling hydrogen solubility in hydrocarbons using extreme gradient boosting and equations of state. *Sci Rep* 11: 17911. https://doi.org/10.1038/s41598-021-97131-8
39. Nakhaei-Kohani R, Taslimi-Renani E, Hadavimoghaddam F, et al. (2022) Modeling solubility of $CO_2$–$N_2$ gas mixtures in aqueous electrolyte systems using artificial intelligence techniques and equations of state. *Sci Rep* 12*:* 3625. https://doi.org/10.1038/s41598-022-07393-z
40. Mahmoudzadeh A, Amiri-Ramsheh B, Atashrouz S, et al. (2024). Modeling $CO_2$ solubility in water using gradient boosting and light gradient boosting machine. *Sci Rep* 14: 13511. https://doi.org/10.1038/s41598-024-63159-9
41. Bamberger A, Sieder G, Maurer G (2004) High-pressure phase equilibrium of the ternary system carbon dioxide + water + acetic acid at temperatures from 313 to 353 K. *J Supercrit Fluids* 32: 15–25. https://doi.org/10.1016/j.supflu.2003.12.014
42. Chapoy A, Mohammadi AH, Chareton A, et al. (2004) Measurement and Modeling of Gas Solubility and Literature Review of the Properties for the Carbon Dioxide-Water System. *Ind Eng Chem Res* 43: 1794–1802. https://doi.org/10.1021/ie034232t
43. Liu Y, Hou M, Yang G, et al. (2011) Solubility of $CO_2$ in aqueous solutions of NaCl, KCl, CaCl 2 and their mixed salts at different temperatures and pressures. *J Supercrit Fluids* 56: 125–129. https://doi.org/10.1016/j.supflu.2010.12.003
44. Chabab S, Théveneau P, Corvisier J, et al. (2019) Thermodynamic study of the $CO_2$ – $H_2O$ – NaCl system: Measurements of $CO_2$ solubility and modeling of phase equilibria using Soreide and Whitson, electrolyte CPA and SIT models. *Int J Greenhouse Gas Control* 91: 102825. https://doi.org/10.1016/j.ijggc.2019.102825
45. Guo H, Huang Y, Chen Y, et al. (2016) Quantitative Raman Spectroscopic Measurements of $CO_2$ 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* 61: 466–474. https://doi.org/10.1021/acs.jced.5b00651
46. Yan W, Huang S, Stenby EH (2011) Measurement and modeling of $CO_2$ solubility in NaCl brine and $CO_2$-saturated NaCl brine density. *Int J Greenhouse Gas Control* 5: 1460–1477. https://doi.org/10.1016/j.ijggc.2011.08.004
47. Longe PO, Danso DK, Gyamfi G, et al. (2024). Predicting $CO_2$ and $H_2$ Solubility in Pure Water and Various Aqueous Systems: Implication for CO2–EOR, Carbon Capture and Sequestration, Natural Hydrogen Production and Underground Hydrogen Storage. *Energies* 17: 5723. https://doi.org/10.3390/en17225723

48. Ng CSW, Djema H, Nait Amar M, et al. (2022) Modeling interfacial tension of the hydrogen-brine system using robust machine learning techniques: Implication for underground hydrogen storage. *Int J Hydrogen Energy* 47: 39595–39605. https://doi.org/10.1016/j.ijhydene.2022.09.120
49. Di M, Sun R, Geng L, et al. (2021) An accurate model to calculate $CO_2$ solubility in pure water and in seawater at hydrate–liquid water two-phase equilibrium. *Minerals* 11: 393. https://doi.org/10.3390/min11040393
50. Zhang XQ, Yao CJ, Zhou YR, et al. (2024) Research on the Law and Influencing Factors of $CO_2$ Reinjection and Storage in Saline Aquifer. In Springer Series in Geomechanics and Geoengineering (Vol. 2). *Springer Nature Singapore*. https://doi.org/10.1007/978-981-97-0268-8_55
51. Blazquez S, Conde MM, Vega C (2024) Solubility of $CO_2$ in salty water: adsorption, interfacial tension and salting out effect. *Mol. Phys* 122. https://doi.org/10.1080/00268976.2024.2306242
52. Yang A, Sun S, Su Y, et al. (2024) Insight to the prediction of $CO_2$ solubility in ionic liquids based on the interpretable machine learning model. *Chem Eng Sci* 297: 120266. https://doi.org/10.1016/j.ces.2024.120266


AIMS Press