ACS OMEGA

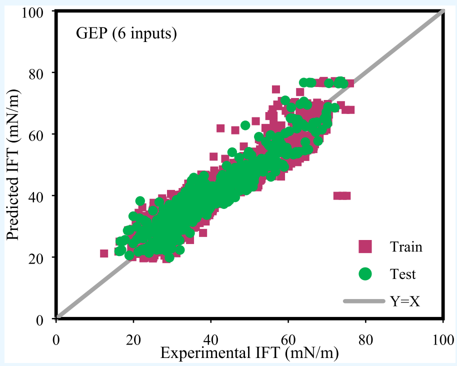
http://pubs.acs.org/journal/acsodf
Article

Toward Predicting Interfacial Tension of Impure and Pure CO_2-Brine Systems Using Robust Correlative Approaches

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

Chen Wei-yu, Lin Sun, Jiyong Zhou, Xuguang Li,* Liping Huang, Guang Xia, Xiangli Meng, and Kui Wang

Cite This: ACS Omega 2024, 9, 7937−7957



ABSTRACT: In the context of global climate change, significant attention is being directed toward renewable energy and the pivotal role of carbon capture and storage (CCS) technologies. These innovations involve secure CO_2 storage in deep saline aquifers through structural and capillary processes, with the interfacial tension (IFT) of the CO_2-brine system influencing the storage capacity of formations. In this study, an extensive data set of 2811 experimental data points was compiled to model the IFT of impure and pure CO_2-brine systems. Three white-box machine learning (ML) methods, namely, genetic programming (GP), gene expression programming (GEP), and group method of data handling (GMDH) were employed to establish accurate mathematical correlations. Notably, the study utilized two distinct modeling approaches: one focused on impurity compositions and the other incorporating a pseudocritical temperature variable (Tcm) offering a versatile predictive tool suitable for various gas mixtures. Among the correlation methods explored, GMDH, employing five inputs, exhibited exceptional accuracy and reliability across all metrics. Its mean absolute percentage error (MAPE) values for testing, training, and complete data sets stood at 7.63, 7.31, and 7.38%, respectively. In the case of six-input models, the GEP correlation displayed the highest precision, with MAPE values of 9.30, 8.06, and 8.31% for the testing, training, and total data sets, respectively. The sensitivity and trend analyses revealed that pressure exerted the most significant impact on the IFT of CO_2-brine, showcasing an adverse effect. Moreover, an impurity possessing a critical temperature below that of CO_2 resulted in an elevated IFT. Consequently, this relationship leads to higher impurity concentrations aligning with lower Tcm values and subsequently elevated IFT. Also, monovalent and divalent cation molalities exhibited a growing influence on the IFT, with divalent cations exerting approximately double the influence of monovalent cations. Finally, the Leverage approach confirmed both the reliability of the experimental data and the robust statistical validity of the best correlations established in this study.

1. INTRODUCTION

The escalation of carbon dioxide (CO_2) levels within the Earth's atmospheric composition stands as a pressing concern, given its profound impact on the dynamics of global climate alteration.[1] To this end, diverse strategies have been initiated to counteract the swift increase in CO_2 concentrations. Among these strategies, carbon capture and storage (CCS) emerges as a propitious method for addressing CO_2 emissions.[2−4] Underground geological structures, encompassing saline aquifers, depleted oil and gas reservoirs, and coal seams have emerged as robust candidates for the secure containment of CO_2.[5−10] On a global scale, deep saline aquifers offer the most extensive capacity for CO_2 storage in comparison to the two primary alternatives: unmineable coal seams and depleted hydrocarbon reservoirs.[11] The interfacial tension (IFT), which emerges due to variations in intermolecular forces between the surfaces of the two phases in contact, holds a pivotal influence over the behavior of multiphase flow within reservoirs.[12] In the context of two-phase displacement, the IFT plays a central role in dictating the dynamic characteristics of phase transportation. The way phases displace each other in porous media, including phenomena such as viscous and capillary fingering, is greatly influenced by the interplay of IFT between these phases. Ensuring CCS efficiency requires retaining injected CO_2 in the subsurface to prevent any escape to the surface. This goal is accomplished by using various trapping methods, such as dissolution, residual or capillary, structural, and mineral trapping.[13−17] Structural trapping curbs the ascent of buoyant

Received: October 11, 2023
Revised: January 18, 2024
Accepted: January 23, 2024
Published: February 6, 2024

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

CO_2 gas through the barrier of the cap rock. At a microscopic level, when the saturation of confined CO_2 remains low and drops below a certain threshold, the IFT between CO_2 and brine keeps residual CO_2 in place, leading to residual trapping.[18] When saturation levels are elevated, capillary trapping takes precedence.[19] Consequently, the efficacy of capillary trapping hinges on the IFT governing the equilibrium between vertical gravity and capillarity within a transitional capillary zone.[20,21] Furthermore, CO_2 can be crucially sequestered as small clusters within the gaps of rock pores, which is extremely important for the effective implementation of CCS projects. This containment is accomplished through the interaction between CO_2-brine IFT.[22]

A significant hurdle in implementing CCS is the considerable expense of isolating pure CO_2 from mixed anthropogenic sources. Direct utilization and subsequent underground injection of this flue gas present an appealing alternative. However, this approach is also susceptible to impurities such as nitrogen (N_2) and methane (CH_4). Although the ultimate aspiration is to sequester pure CO_2, practical economic considerations steer the preference toward utilizing the accessible impure CO_2.[23] As a result, it is customary for the injected CO_2 stream to encompass or be amalgamated with diverse impurities originating from various sources.[24,25] CH_4 as a greenhouse gas contributes to both indirect and direct mechanisms for retaining infrared radiation via the process of oxidation.[26] Co-injecting CO_2 with other substances has the potential to alter different properties, such as viscosity, diffusion coefficient, and density. The alteration of IFT in the resulting fluid systems, especially in gas-brine systems within saline formations, is substantial. These changes collectively influence the migration and entrapment of CO_2 within subsurface reservoirs.[26,27]

Accurate and reliable IFT values between CO_2 and brine under in situ conditions are of utmost importance. These values are essential for accurately assessing various aspects, such as the spread of CO_2 in displacement scenarios, its storage processes, and the final immobilization of CO_2 below the Earth's surface. Recent years have witnessed a significant amount of research dedicated to estimating the IFT between gas containing pure or impure CO_2 and brine. A multitude of experimental investigations have been conducted to furnish IFT data for these systems across varying temperature ranges, pressure conditions, salt types, salinity levels, and compositions of CO_2-containing gas.[28−36] These inquiries mainly depend on the pendant drop method because of its inherent advantages. This method allows for accurate measurement of the IFT even at high pressures and temperatures. However, it is crucial to acknowledge that this experimental method can be time-consuming and financially taxing, requiring subsequent interpretive methods. In contrast, only a small number of modeling investigations have been aimed at generating practical correlations. Furthermore, theoretical underpinnings for comprehending the effects of factors like temperature and pressure on IFT and rock wettability have also been established through the execution of molecular dynamics simulations. Besides, these models may overestimate predictions, particularly when dealing with high pressures.[37−39] Several researchers have put forward different correlations for predicting the pure CO_2-water IFT, specifically in scenarios where salinity is absent.[30,35,40] In addition, scientists have developed several empirical correlations to explain the IFT in pure CO_2-brine systems.[32,41,42] Conversely, employing the linear gradient theory, Yan et al.[33] proposed correlations for impure CO_2-brine mixtures to assess the IFT in gas−water combinations encompassing CO_2, N_2, and CH_4. However, the model exhibited limited precision, particularly when applied to gas mixtures containing CO_2. Utilizing the alternating condition expectation algorithm, IFT predictions spanning a broad pressure and temperature range of 0.1−60 MPa and 5.25−175 °C were conducted by Li et al.[43] However, despite the model's intricacy, it shows a lack of accuracy, coupled with an error exceeding 10%, emphasizing the critical necessity for the development of a more dependable predictive approach. Therefore, researchers explored artificial intelligence modeling due to its capacity to effectively represent intricate systems encompassing diverse included parameters.[44,45] Zhang et al.[46] employed a neural network to model CO_2-brine IFT, utilizing a database containing a total of 1716 data points. Kamari et al.[47] employed the identical database of 1716 data points, utilizing multiple machine learning (ML) models to establish a predictive model based on four input variables, including temperature, pressure, monovalent cation molality, and divalent cation molality for CO_2-brine IFT. Niroomand-Toomaj et al.[48] introduced a Radial Basis Function model, for the prediction of CO_2/aquifer brine IFT. This model is developed using 378 data points, encompassing various temperatures, pressures, and salinities. Partovi et al.[49] harnessed 1716 data points of CO_2-brine IFT and leveraged computer-based models to formulate hybrid models. These models yielded notably more accurate results compared with empirical correlations. Rashid et al.[50] employed 1019 experimentally measured IFT values within their intelligent modeling approach. Amooie et al.[51] employed a novel database containing 2517 experimental data points encompassing both pure and impure CO_2-brine systems. They applied several white-box and black-box ML techniques to predict the IFT in impure CO_2-brine systems. Nait Amar[52] employed the genetic programming (GP) approach to estimate IFT in both pure and impure CO_2-brine systems across diverse operational scenarios. Their study incorporated 2346 validated IFT measurements. Safaei-Farouji et al.[53] utilized a data set comprising 2184 experimental data points and employed robust ML methods to predict IFT within the CO_2-brine system. Drawing from a data set of 2517 experimental data points, Zhang et al.[54] conducted a study involving the modeling of CO_2-brine IFT. The findings indicated that increased pressure and/or decreased geothermal gradients result in a significant increase in the maximum structural trapping capacity. Evidently, in recent times, researchers have consistently dedicated efforts toward refining IFT prediction conditions, striving for broader and more inclusive modeling approaches. This entails the incorporation of a growing body of authentic experimental data spanning a diverse array of operational conditions. However, recognizing the significance of carbon sequestration in saline reservoirs, the utilization of expanded data sets, and the inclusion of impurity variables within the system hold the promise of constructing a comprehensive model endowed with advanced capabilities.

The objective of this research is to formulate enhanced, explicit mathematical correlations to estimate the IFT in CO_2-brine systems. This project is built on the most extensive data set gathered to date, covering 2811 data points. To achieve this goal, three renowned white-box ML methodologies, namely, genetic programming (GP), gene expression programming (GEP), and group method of data handling (GMDH), are employed. A variety of statistical and graphical assessment

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

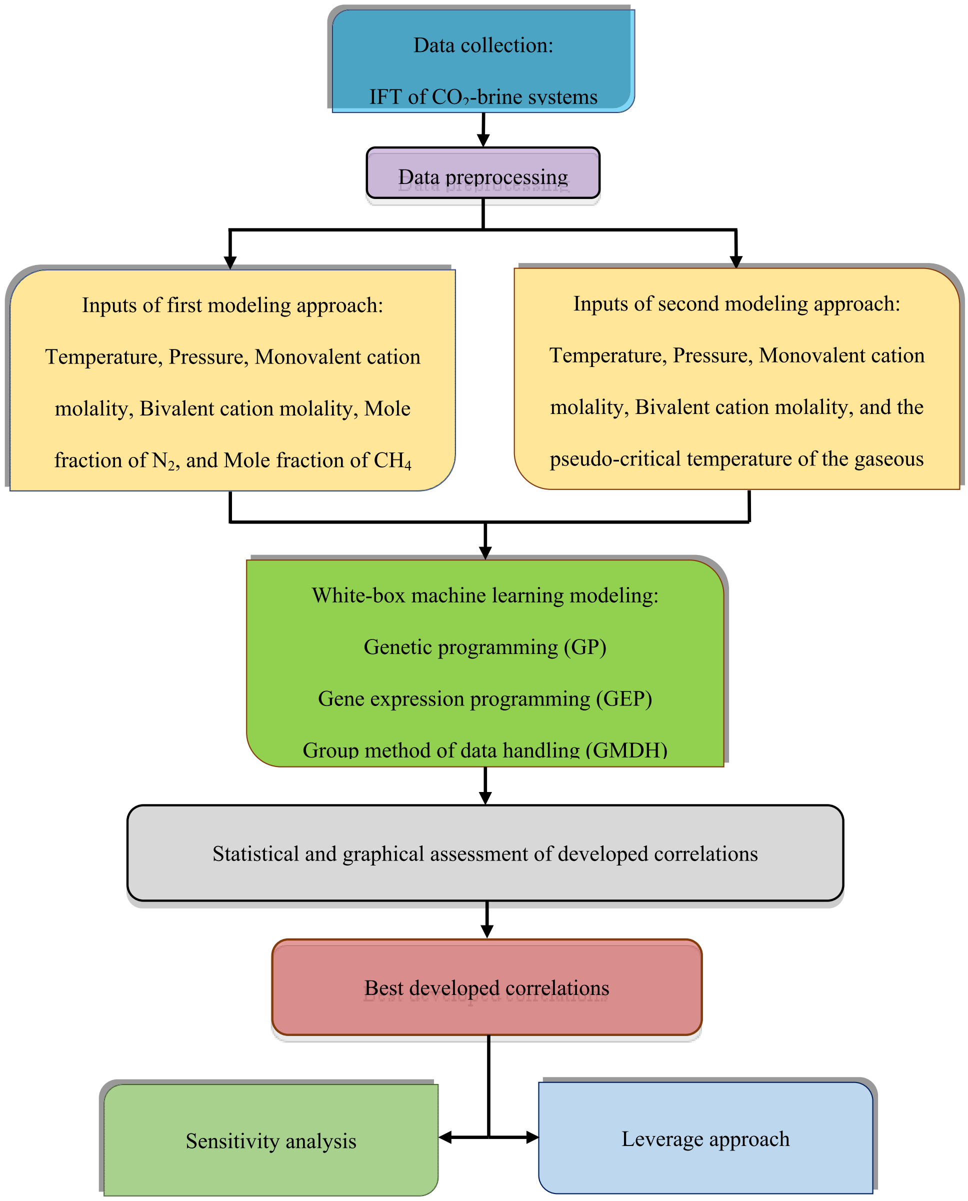


Figure 1. Research steps for modeling the CO_2-brine IFT.

techniques are used to validate these correlations. Furthermore, sensitivity analysis is conducted using Spearman rank correlation to investigate the influence of input variables on the output. Lastly, the leverage method is employed to determine the practical application range of the established correlations. Figure 1 illustrates the research process undertaken in this study to model the IFT between CO_2 and brine.

2. DATA COLLECTION

In order to model the IFT of both impure and pure CO_2-brine systems, 2811 experimental data points were gathered from reliable literature sources.[28−36,41,42,55−71] This database is the most comprehensive collected for the IFT of CO_2-brine to date, containing about 290 or more data from similar studies. Typically, in these investigations, impure brine is synthesized through the introduction of salts such as KCl, NaCl, $CaCl_2$, Na_2SO_4, and $MgCl_2$. Yet, it has been demonstrated that the resultant IFT of CO_2-brine primarily hinges on the valence of the cations present. Hence, the classification of brine impurities is undertaken through the bifurcation of input variables into two distinct categories: the molalities of monovalent and divalent cations. It is important to note that in the context of pure water, the molality values for both monovalent and divalent cations are designated as zero. In the context of the available experimental data, the impurity within the CO_2-rich gas phase is attributed solely to differing concentrations of the CH_4 and N_2 components. However, a fresh perspective is adopted, wherein individual impurities are not isolated based on their specific identities. Instead, the impure mixture is approached as a unified whole, discerned by the introduction of a novel critical parameter tailored for this distinction. This study employed two distinct modeling methodologies, one involving 6 inputs and the other involving 5 inputs. The initial approach encompassed 6 inputs: temperature, pressure, monovalent cation molality, bivalent cation molality, mole fraction of N_2, and mole fraction of CH_4 within the injected gas. Conversely, the second approach encompassed 5 inputs, specifically temperature, pressure, monovalent cation molality,

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

Table 1. Statistical Details of the Gathered Databank in This Research

	pressure (MPa)	temperature (K)	monovalent cation molality (mol/kg)	bivalent cation molality (mol/kg)	CH_4 (mole fraction)	N_2 (mole fraction)	Tcm (K)	IFT (mN/m)
mean	14.02	333.46	1.02	0.45	0.02	0.02	298.96	39.91
median	9.06	323.40	0.05	0.00	0.00	0.00	304.20	37.19
mode	2.00	373.15	0.00	0.00	0.00	0.00	304.20	31.90
SD	12.89	38.18	1.38	1.02	0.11	0.09	20.73	11.51
Kurtosis	3.37	−0.29	0.91	9.26	32.13	42.87	19.67	0.02
Skewness	1.82	0.71	1.35	2.99	5.64	6.48	−4.39	0.74
minimum	0.08	278.15	0.00	0.00	0.00	0.00	170.65	12.40
maximum	69.51	448.15	5.13	5.00	0.80	0.75	304.20	76.10
count	2811	2811	2811	2811	2811	2811	2811	2811
status	input	input	input	input	input	input	input	target

bivalent cation molality, and the pseudocritical temperature of the gaseous mixture (Tcm). To elucidate, in the second approach, rather than integrating the concentrations of CH_4 and N_2 as separate input variables, we introduce a solitary input variable (i.e., Tcm) that encapsulates the entirety of the mixture's composition. This consolidated approach effectively captures the essence of its constituents and is computed as follows:[72]

$$\mathrm{Tcm} = \sum_{i}^{nc,g} y_i Tc_i \tag{1}$$

where y_i represents the mole fraction of every component, with a corresponding critical temperature of Tc_i. Additionally, nc,g denotes the number of components present in the gas mixture.

Table 1 presents the statistical summary detailing the input and target parameters of the data set employed for the modeling process. A visualization known as a box plot portrays the five-number summary derived from a data set. This summary includes the minimum, first quartile, median, third quartile, and maximum values. In the construction of a box plot, a box spans from the first quartile to the third quartile, while a vertical line intersects the box at the median value. Figure 2 presents a collection of box plots representing all variables present within the database utilized in this study. The data presented in Table 1 illustrate the experimental IFT measurements, spanning a broad spectrum of operating pressures, ranging from 0.08 to 69.51 (MPa), and temperatures spanning from 278.15 to 448.15 (K). The extensive range of model input parameters and their varied span offer a robust foundation for constructing an inclusive predictive correlation for the IFT of CO_2-brine. To partition the data, a random separation of the databank into two subsets was employed, resulting in a training set comprising 80% of the entire data set, and a test set encompassing the remaining 20%.

3. MODEL DEVELOPMENT

3.1. Genetic Programming (GP). GP constitutes a specialized area within artificial intelligence and evolutionary computation.[73] It operates by drawing on concepts derived from natural selection and evolution to autonomously produce computer programs or models with the capacity to address distinct tasks or challenges. In scenarios characterized by nonlinearity and significant levels of imprecision, the employment of the GP method is widespread for the derivation of correlations.[74,75] GP systems exhibit structures reminiscent of those of trees. Nodes within these structures are categorized into internal or external groups based on their positions.[76] The process of the GP approach is illustrated in Figure 3 using a flowchart configuration. Employing an iterative methodology, this dependable, bioinspired technique ultimately yields a precise mathematical representation. Upon input of the data, the initial phase involves establishing a random community. Subsequently, an objective function is employed to assess the performance of the model. Following this, parent candidates with elevated correlation values are selected for modification using genetic operators, giving rise to offspring.[77] Successive iterations are produced following the same procedure, and the quality of subsequent generations is enhanced as desired knowledge from proficient individuals is transmitted to their progeny. The process concludes when the desired level of accuracy is achieved or the predetermined maximum number of iterations is attained.[78]

3.2. Gene Expression Programming (GEP). The GEP, pioneered by Ferreira in 2001,[79] stands as a sophisticated approach within the realm of soft computing. Operating within the evolutionary algorithm framework, it harnesses the tenets of evolution to achieve its goals. The distinctive benefit of GEP lies in its capacity to produce precise mathematical representations of the systems under examination. From a conceptual standpoint, GEP is considered an enhanced iteration of GP, which was first introduced by Koza.[73] GEP effectively addresses the challenges inherent in GP, notably the constraints posed by regression strategies.[79] Like any other evolutionary algorithm, GEP undertakes exploration for the optimal expression model by utilizing chromosomes that encode and represent potential solutions. Moreover, GEP introduces a pivotal component known as the expression tree into its framework. The process of obtaining the expression tree involves the conversion of chromosomes into tangible candidates. Within the structure of GEP, genes are utilized, encompassing both terminals and a head that encompasses functions. Each gene is characterized by a predetermined sequence of symbols, which correspond to various operators like $\{+, -, \times, /, \sqrt{}, \log\}$, alongside a terminal set encompassing$\{x, y, z\}$.[80] The GEP process involves several key stages. To begin, GEP parameters are established, encompassing critical aspects, such as population size, termination conditions, and gene length. Subsequently, an initial population of chromosomes is created, each representing a distinct mathematical expression and selected at random.[81] A fitness evaluation is then conducted, gauging the chromosomes' suitability based on a designated fitness function. The most promising individuals are identified and retained for the subsequent generation, while tournament selection is employed to detect candidates for recombination, yielding new offspring. Within this framework, options for recombination

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

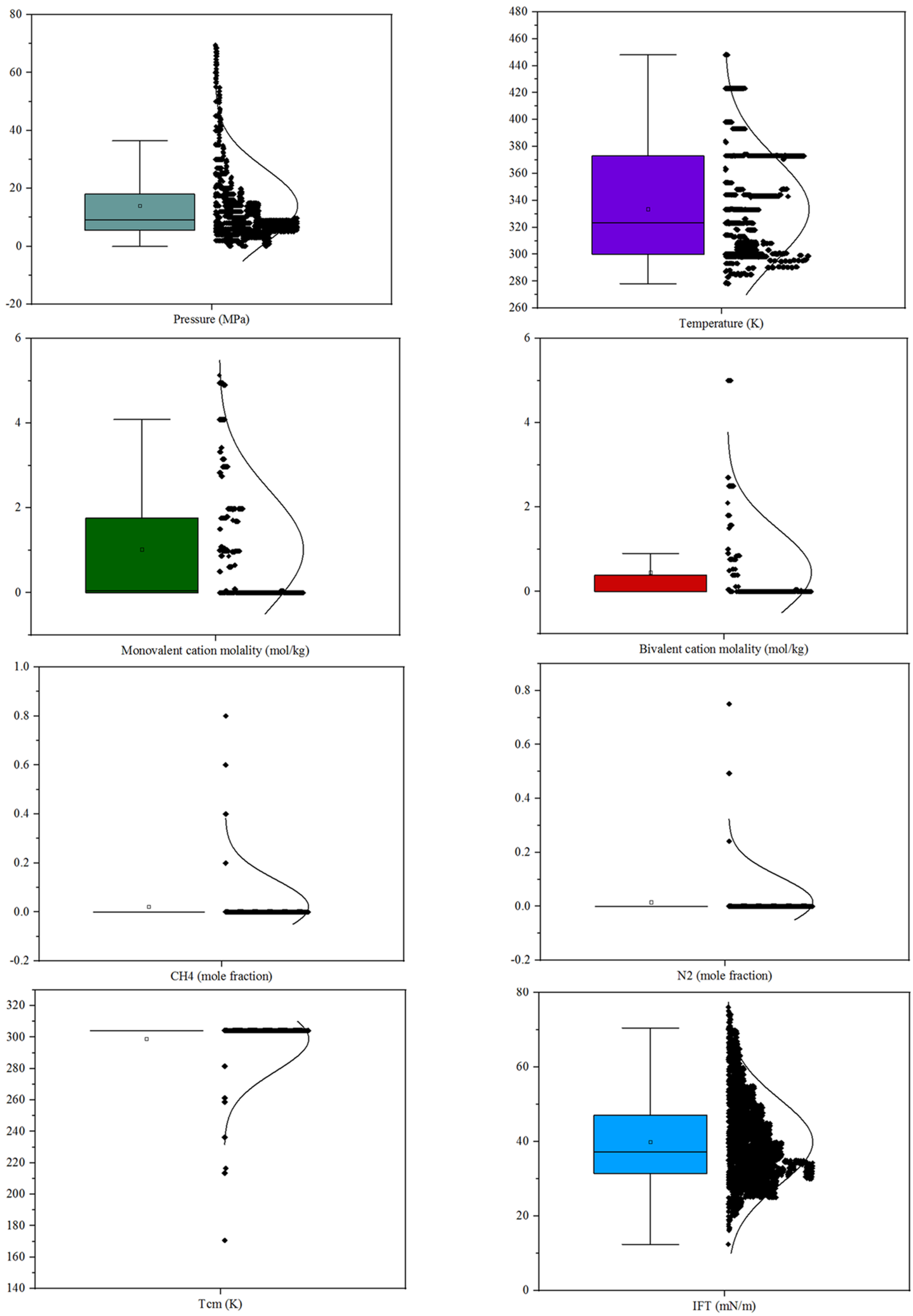


Figure 2. Box plots of all variables present within the database utilized in this study.

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

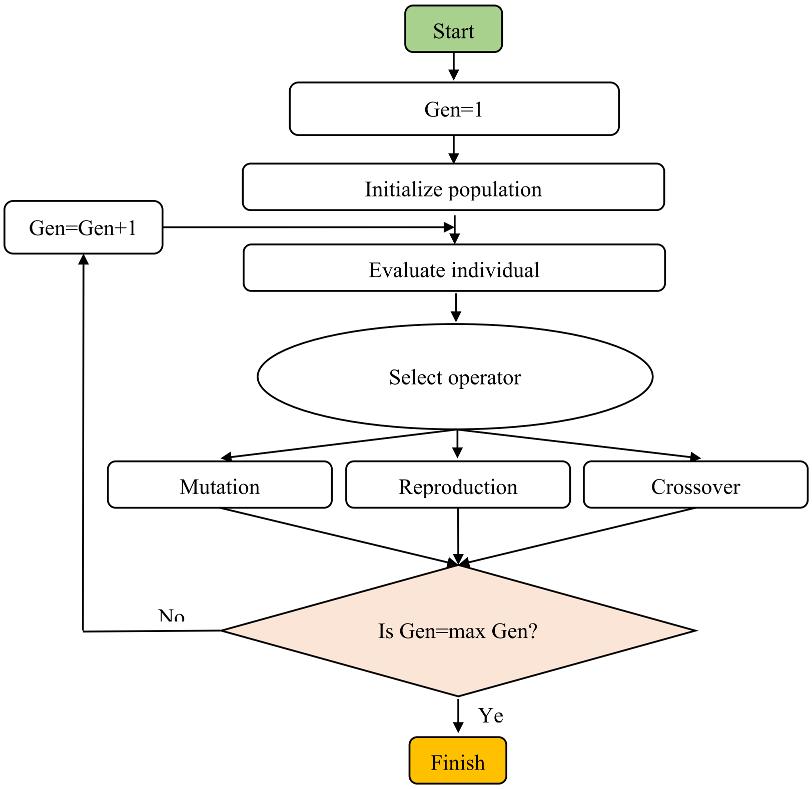


Figure 3. Flowchart of the GP approach.

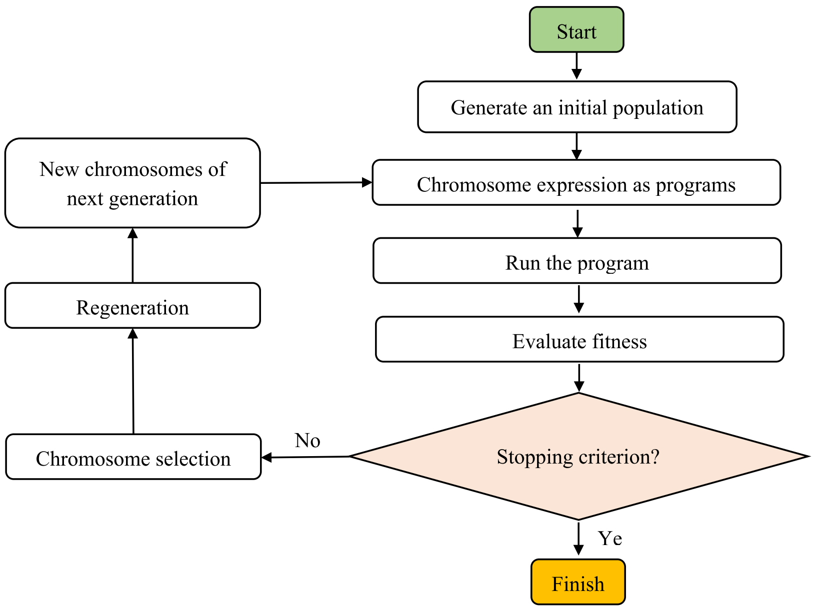


Figure 4. Flowchart of the GEP approach.

include both one- and two-point recombination techniques. The process is enriched by a mutation operation, which substantially impacts GEP's genomic makeup by substituting one element for another. Additionally, the concept of transposition and insertion is integrated, allowing specific sections of a chromosome's genome to be activated and repositioned.[79] These steps, encompassing fitness evaluation, selection, recombination, mutation, and genomic manipulation, are iteratively undertaken until a predefined termination criterion is satisfied. Figure 4 shows the flowchart of the GEP approach.

3.3. Group Method of Data Handling (GMDH). Ivakhnenko[82] pioneered the GMDH technique, an approach designed to model intricate systems characterized by multifaceted input data converging into a singular output. The fundamental objective of the GMDH approach involves constructing a feed-forward system function through the utilization of a second-order transfer function. This technique establishes the number of hidden layers, the number of

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

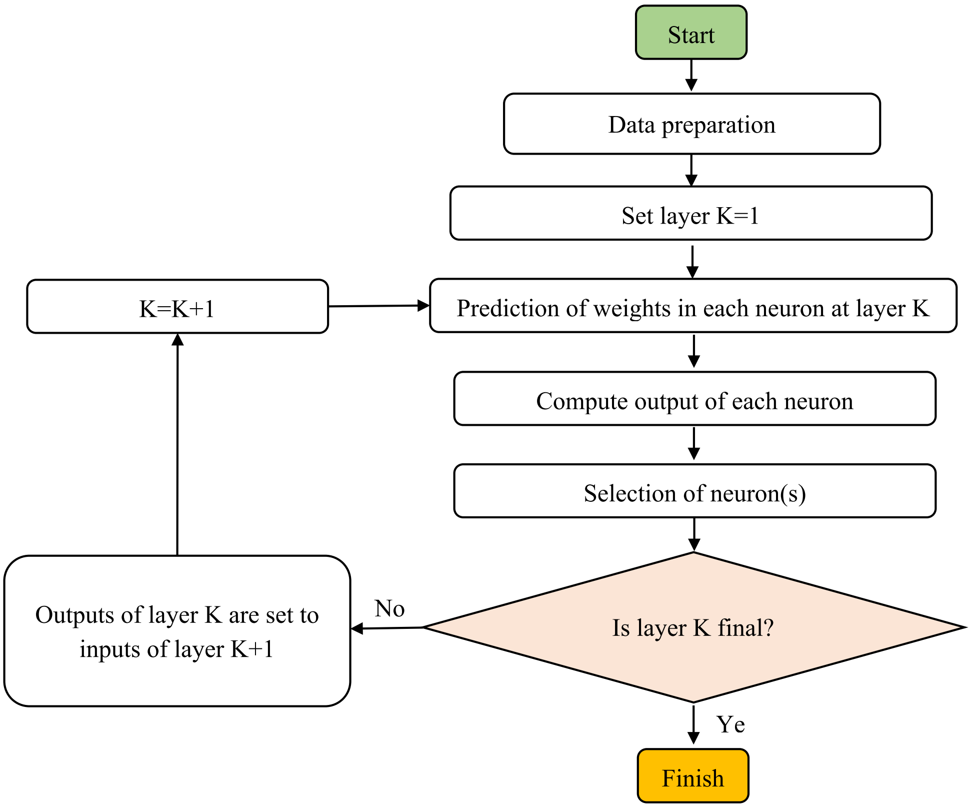


Figure 5. Flowchart of the GMDH approach.

neurons within those layers, and the most suitable model configuration. The interrelation between the dependent and independent variables in GMDH is represented by a nonlinear function termed the Volterra series, formulated as eq 2. To evaluate the Volterra series as a two-variable second-order polynomial, eq 3 is employed:[83]

$$\hat{y} = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} x_j a_{ijk} x_i x_j x_k ... \quad (2)$$

$$G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \quad (3)$$

The objective of the GMDH approach is to ascertain the unspecified parameters, denoted as a_i, within the Volterra series. These a_i parameters are computed for every pairing of input variables x_i and x_j inputs using regression methodologies.[84,85] Building upon this principle and accounting for the notion of least-squares error, the G function can be formulated as follows for a given set of M observations involving multi-input and single-output data pairs:[86,87]

$$E = \frac{\sum_{i=1}^{M} (y_i = G_i O)^2}{M} \quad (4)$$

$$y_i = f(x_{i1}, x_{i2}, ..., x_{im}), \; i = 1, 2, ..., m \quad (5)$$

A simple flowchart of the GMDH algorithm is shown in Figure 5.

4. RESULTS AND DISCUSSION

4.1. Developed Correlations. As previously indicated, the utilization of GP, GEP, and GMDH led to the creation of numerous equations by considering five and six input parameters to estimate impure and pure CO_2 and brine IFT. The formulated equations are outlined as follows:

GP correlation (6 inputs):

$$\mathrm{IFT} = \left(\left(c_0{}^*\mathrm{BCM} + c_1{}^*P + c_2{}^*\mathrm{MCM} + \exp(c_3{}^*\mathrm{CH_4})^* c_4 + \frac{c_5{}^*T}{(c_6{}^*P + c_7)} + \frac{c_8}{(c_9{}^*N_2 + c_{10})} + c_{11}\right)^* c_{12} + c_{13}\right) \quad (6)$$

$c_0 = 1.7933$
$c_1 = 0.02595$
$c_2 = 1.0789$
$c_9 = 0.375$
$c_4 = 25.226$
$c_0 = 1.227$
$c_6 = 3.603$
$c_7 = 14.895$
$c_1 = 1.0291$
$c_9 = -0.007072$
$c_{10} = -0.02436$
$c_1 = 24.592$
$c_{12} = 2.0911$
$c_{13} = 0.014857$

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

GP correlation (5 inputs):

$$\mathrm{IFT} = \left(\left(\left(c_0{}^{*}\mathrm{Tcm} + c_1{}^{*}\mathrm{MCM} + c_2{}^{*}\mathrm{BCM} + \frac{c_3{}^{*}\mathrm{Tcm}}{(c_4{}^{*}P + c_5)} + \frac{c_6{}^{*}T}{(c_7{}^{*}P + c_8)} + c_9\right) + c_{10}\right)\right) \quad (7)$$

$c_0 = -0.14747$
$c_1 = 2.5728$
$c_2 = 4.5562$
$c_3 = -0.29192$
$c_4 = -0.9349$
$c_5 = -3.3606$
$c_6 = -0.29668$
$c_7 = -0.86023$
$c_8 = -3.3606$
$c_9 = 62.314$
$c_{10} = 4.7827 \times 10^{-9}$

GEP correlation (6 inputs):

$$\mathrm{IFT} = \left(\frac{1}{T^{*}\left(c_1{}^{*}T+\frac{(c_2{}^{*}\mathrm{MCM}+c_3{}^{*}\mathrm{BCM}+c_4)}{c_5{}^{*}P} + c_6\right)^{*}(c_7{}^{*}\mathrm{N_2} + c_8)^{*}(c_9{}^{*}\mathrm{BCM}+(c_{10}{}^{*}\mathrm{MCM} + c_{11})^{*}(c_{12}{}^{*}\mathrm{CH_4} + c_{13}) + c_{14})^{*}c_{15}} + c_{16}\right) \quad (8)$$

$c_1 = -0.017632$
$c_2 = -1.4104$
$c_3 = -2.4452$
$c_4 = -16.302$
$c_5 = -0.55608$
$c_6 = 14.208$
$c_7 = 1.648$
$c_8 = 1.9531$
$c_9 = -2.5993$
$c_{10} = -0.37481$
$c_{11} = -13.088$
$c_{12} = 2.057$
$c_{13} = 3.8507$
$c_{14} = 18.404$
$c_{15} = 9.3247\times 10^{-8}$
$c_{16} = 78.36$

GEP correlation (5 inputs):

$$\mathrm{IFT} = \left(\frac{\left(c_0{}^{*}\mathrm{MCM}+c_1{}^{*}\mathrm{BCM}+\frac{\mathrm{Tcm}^{*}P^{*}c_2}{\left(c_3{}^{*}P+\frac{c_4{}^{*}T}{c_5{}^{*}\mathrm{Tcm}}\right)} + c_6\right)}{c_7{}^{*}T} + c_8\right) \quad (9)$$

$c_0 = 1446.8$
$c_1 = 2461.0$
$c_2 = 356.77$
$c_3 = -3.0301$
$c_4 = 2.4591$
$c_5 = -0.21942$
$c_6 = 16147.0$
$c_7 = 1.8543$
$c_8 = 48.195$

GMDH correlation (6 inputs):

$$\begin{aligned}
Y1 &= -237.329 + N12^{*}8.59043 - N12^{*}N3^{*}0.029114 - N12^{2*}0.0683583 + N3^{*}2.32834 - N3^{2*}0.00170334\\
N3 &= -5.70709 + N4^{*}0.945527 + N4^{*}N7^{*}0.0422513 - N4^{2*}0.0241664 + N7^{*}0.242468 - N7^{2*}0.0187927\\
N4 &= -42.6855 - N8^{*}1.61476 - N8^{*}N10^{*}0.00503938 + N8^{2*}0.0359466 + N10^{*}3.40845 - N10^{2*}0.0254462\\
N7 &= 60.3427 + \mathrm{MCM}^{*}N9^{*}0.0398817 + \mathrm{MCM}^{2*}0.129818 - N9^{*}2.41246 + N9^{2*}0.0444492\\
N8 &= -131.961 - P^{*}1.61189 + P^{2*}0.0194535 + T^{*}1.01047 - T^{2*}0.00132743\\
N9 &= 52.8677 - P^{*}1.50791 + P^{*}\mathrm{BCM}^{*}0.0254271 + P^{2*}0.0188808 + \mathrm{BCM}^{*}1.96934 + \mathrm{BCM}^{2*}0.20259\\
N10 &= 38.635 - \mathrm{MCM}^{*}1.5212 + \mathrm{MCM}^{*}\mathrm{BCM}^{*}3.24198 + MCM^{2*}0.517975 + \mathrm{BCM}^{2*}0.516763\\
N12 &= 39.3559 + (\mathrm{CH_4})^{2*}25.6457 + (N_2)^{2*}24.5578
\end{aligned} \quad (10)$$

GMDH correlation (5 inputs):

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

$$
\begin{aligned}
Y1 &= -7.1885 + N7^{*}0.689503 - N7^{*}N2^{*}0.0131058 + N2^{*}0.626504 + (N2)^{2*}0.00974851\\
N2 &= -14.6423 + P^{*}0.706114 - P^{*}N3^{*}0.0138623 - P^{2*}0.00531772 + N3^{*}1.53122 - (N3)^{2*}0.00456879\\
N3 &= -5.28129 + \mathrm{MCM}^{*}5.23378 - \mathrm{MCM}^{*}N4^{*}0.0590986 - \mathrm{MCM}^{2*}0.738552 + N4^{*}1.11658\\
N4 &= 18.5756 - N13^{*}2.47684 - N13^{*}N6^{*}0.0738438 + (N13)^{2*}0.0658759 + N6^{*}2.66314 + (N6)^{2*}0.0157403\\
N6 &= -227.611 + T^{*}1.05442 - T^{*}N9^{*}0.00392383 - T^{2*}0.00122593 + N9^{*}2.90394 - (N9)^{2*}0.00581509\\
N7 &= -42.6855 - N12^{*}1.61476 - N12^{*}N14^{*}0.00503938 + (N12)^{2*}0.0359466 + N14^{*}3.40845 - (N14)^{2*}0.0254462\\
N9 &= -135.413 - N11^{*}2.43339 + N11^{*}N15^{*}0.00798595 + (N11)^{2*}0.041285 + N15^{*}8.00916 - (N15)^{2*}0.0802109\\
N11 &= 52.8677 - P^{*}1.50791 + P^{*}\mathrm{BCM}^{*}0.0254271 + P^{2*}0.0188808 + \mathrm{BCM}^{*}1.96934 + \mathrm{BCM}^{2*}0.20259\\
N12 &= -131.961 - P^{*}1.61189 + P^{2*}0.0194535 + T^{*}1.01047 - T^{2*}0.00132743\\
N13 &= 61.9661 + \mathrm{BCM}^{*}\mathrm{Tcm}^{*}0.0122519 - \mathrm{BCM}^{2*}0.268553 - \mathrm{Tcm}^{2*}0.000260694\\
N14 &= 38.635 - \mathrm{MCM}^{*}1.5212 + \mathrm{MCM}^{*}\mathrm{BCM}^{*}3.24198 + \mathrm{MCM}^{2*}0.517975 + \mathrm{BCM}^{2*}0.516763\\
N15 &= 61.3114 - \mathrm{MCM}^{*}5.51071 + \mathrm{MCM}^{*}\mathrm{Tcm}^{*}0.015966 + \mathrm{MCM}^{2*}0.5061 - \mathrm{Tcm}^{2*}0.000247395
\end{aligned}
\tag{11}
$$

where T is the temperature (K), P is the pressure (MPa), MCM stands for monovalent cation molality (mol/kg), BCM stands for bivalent cation molality (mol/kg), Tcm is the pseudocritical temperature of the gaseous mixture, and N_2 and CH_4 show mole fractions of N_2 and CH_4 within the injected gas, respectively.

4.2. Assessment of the Correlations. Using seven statistical indicators, we assessed the precision of the proposed models. The study incorporated the following metrics for evaluation:[88]

1. Mean absolute percentage error (MAPE, %): MAPE measures the average percentage difference between predicted and actual values, indicating the overall accuracy of the model.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{abs}\left(\left[\frac{(y)_{\mathrm{exp}} - (y)_{\mathrm{pred}}}{y_{\mathrm{exp}}}\right]\right) \times 100 \tag{12}$$

2. Standard deviation (SD): SD quantifies the dispersion or spread of predicted values around the actual values, giving insight into the variability of the model's performance.

$$\mathrm{SD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_{\mathrm{exp}} - y_{\mathrm{pred}}}{y_{\mathrm{exp}}}\right)^2} \tag{13}$$

3. Root mean square error (RMSE): RMSE calculates the average magnitude of the differences between predicted and actual values, reflecting the model's overall prediction error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{exp}} - y_{\mathrm{pred}})^2} \tag{14}$$

4. Determination coefficient (R^2): R^2 assesses the proportion of variance in the dependent variable that is explained by the independent variables in the model. It indicates how well the model fits the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{\mathrm{exp}} - y_{\mathrm{pred}})^2}{\sum_{i=1}^{N}(y_{\mathrm{exp}} - \overline{y_{\mathrm{exp}}})^2} \tag{15}$$

in this context, N represents the number of data, y_{exp} denotes the experimental data, and y_{pred} signifies the data predicted by the correlations proposed in the study.

5. Mean absolute error (MAE): MAE computes the average absolute differences between predicted and actual values, providing a measure of the average prediction error. This assessment corresponds to a risk evaluation that mirrors the expected outcome of the absolute error loss or $l1$-norm loss. If $\hat{y}_i$ represents the expected outcome of the ith instance, and y_i denotes the corresponding actual value, the computed MAE for a total of "n_data" can be expressed as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{16}$$

6. Mean bias error (MBE): MBE calculates the average difference between predicted and actual values, indicating the overall bias or tendency of the model to overestimate or underestimate.

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i) \tag{17}$$

7. Nash−Sutcliffe efficiency (NSE): NSE evaluates the model's performance by comparing the predicted values to the mean observed value. It assesses how well the model captures the variation in the observed data.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_{\mathrm{o}}^{t} - y_{\mathrm{m}}^{t})^2}{\sum_{i=1}^{n}(y_{\mathrm{o}}^{t} - \bar{y}_{\mathrm{o}})^2} \tag{18}$$

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

Table 2. Results of Statistical Error Analysis for All Models

statistical criteria		RMSE	SD	R^2	MAPE (%)	MBE	MAE	NSE
GP (6 inputs)	test	4.8733	0.1389	0.8379	9.8978	−0.1404	3.6904	0.8151
	train	4.4064	0.1169	0.8495	8.4081	0.0352	3.2446	0.8204
	all	4.5038	0.1217	0.8470	8.7064	4.1×10^{-13}	3.3339	0.8193
GEP (6 inputs)	test	4.2900	0.1296	0.8744	9.2988	−0.0469	3.3839	0.8567
	train	4.1060	0.1106	0.8693	8.0647	0.0631	3.0523	0.8489
	all	4.1435	0.1147	0.8705	8.3119	0.0410	3.1187	0.8507
GMDH (6 inputs)	test	4.9537	0.1497	0.8326	10.3991	−0.1281	3.7652	0.8055
	train	4.6396	0.1296	0.8331	9.5159	0.0321	3.5744	0.7977
	all	4.7042	0.1339	0.8330	9.6928	-2.3×10^{-11}	3.6126	0.7996
GP (5 inputs)	test	4.7648	0.1424	0.8410	10.3019	−0.0889	3.7533	0.8203
	train	4.4857	0.1241	0.8451	9.0998	0.0223	3.4299	0.8141
	all	4.5430	0.1280	0.8443	9.3405	-2.9×10^{-11}	3.4947	0.8156
GEP (5 inputs)	test	4.2197	0.1305	0.8785	9.1968	−0.0516	3.3016	0.8628
	train	3.9132	0.1124	0.8813	8.0600	0.0129	2.9911	0.8650
	all	3.9764	0.1163	0.8807	8.2877	1.7×10^{-13}	3.0533	0.8645
GMDH (5 inputs)	test	3.7919	0.1095	0.8943	7.6343	0.1059	2.8170	0.8803
	train	3.5825	0.1020	0.9025	7.3111	−0.0265	2.7166	0.8923
	all	3.6254	0.1035	0.9008	7.3759	1.9×10^{-14}	2.7367	0.8899

Here, $\overline{y}_o$ represents the mean of observed data, while y_m signifies the simulated data. Additionally, y_o^t denotes the data being released at time instant t.

These metrics collectively provide a comprehensive evaluation of the accuracy and reliability of the suggested correlations in predicting the desired outcomes. The computed values of the aforementioned statistical parameters are listed in Table 2. Based on the statistical assessment presented in Table 2, it is evident that the GMDH correlation utilizing five inputs has demonstrated the utmost precision and dependability across all indicators. Notably, its MAPE values for the testing, training, and complete data sets are recorded at 7.63, 7.31, and 7.38%, respectively. On the other hand, in the case of six-input models, the GEP correlation displayed the highest precision, with MAPE values of 9.30, 8.06, and 8.31 for the testing, training, and total data sets, respectively. Taking into account all of the statistical parameters, the sequence of correlations based on accuracy is as follows: GMDH (5 inputs), GEP (5 inputs), GEP (6 inputs), GP (6 inputs), GP (5 inputs), and GMDH (6 inputs). The consolidated approach (5 inputs), which utilizes Tcm of a gaseous mixture rather than treating CH_4 and N_2 as separate input variables, effectively captures the essence of its constituents, represents impurities in CO_2-containing gas, and reduces the complexity of mathematical models with higher accuracy compared to those involving six inputs.

Moving forward, a graphical examination can be employed to evaluate the precision of the suggested correlations. To initiate this investigation, graphical representations showcasing the cross-plots of IFT values predicted by the correlations against the corresponding experimental values were depicted in Figure 6. Within the cross-plots, a significant clustering of data points around the $X = Y$ reference line is observed across all models, indicative of the correlations' commendable accuracy. Notably, the GMDH model formulated with five inputs exhibits a notably enhanced concentration around the $X = Y$ line, underscoring its superior predictive capability for the CO_2-brine interfacial tension.

Furthermore, Figure 7 provides insight into the distribution of errors during both the testing and training phases of the models. A thorough examination of the error distribution graphs reveals the absence of any discernible error pattern with the majority of computed errors clustering around the zero-error line. Notably, the GMDH model constructed using five inputs exhibits a reduced dispersion of errors compared to the other models, underscoring its elevated accuracy in predicting the IFT of both impure and pure CO_2-brine systems.

After the analyses outlined in this study were conducted, a cumulative frequency plot depicting the entirety of the data against absolute error values has been generated for all predictive correlations, as depicted in Figure 8. This visual representation reveals that over 70% of the predicted IFT data using all correlations exhibit an absolute error of less than 10%. Furthermore, more than 90% of the predicted data demonstrate an absolute error of less than 20%. In contrast, the GMDH model employing five inputs displays absolute errors of approximately 8 and 15% for 70% and 90% of the IFT data, respectively. This highlights the exceptional accuracy of the GMDH correlation when compared to the other correlations.

4.3. Trend Analysis. During the following stage of graphical analysis, the accuracy of the most well-established correlations, specifically GMDH (employing 5 inputs) and GEP (employing 6 inputs), is examined for their ability to predict the expected physical trend of IFT in CO_2-brine systems. Initially, the impact of the pressure on IFT measurements is investigated for a pure CO_2 and brine system. This assessment is conducted under a consistent temperature of 333 K, encompassing a wide pressure spectrum ranging from 1 to 68 MPa. The system under consideration maintains a monovalent cation molality of 1.98 mol/kg, while the divalent cation molality remains at zero, aligning with experimental observations detailed in the literature.[68] The objective is to employ established correlations to project the experimental outcomes. As illustrated in Figure 9, the overall IFT of the CO_2-brine system experiences a decline as the pressure increases. The trend of IFT reduction with respect to pressure reveals two discernible segments: an initial sharp decline, succeeded by a considerably gradual decrease within a subsequent pseudo-steady plateau phase. CO_2 possesses critical conditions at a temperature of 304.2 K and a pressure of 7.38 MPa.[89] For the examined system at a constant

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

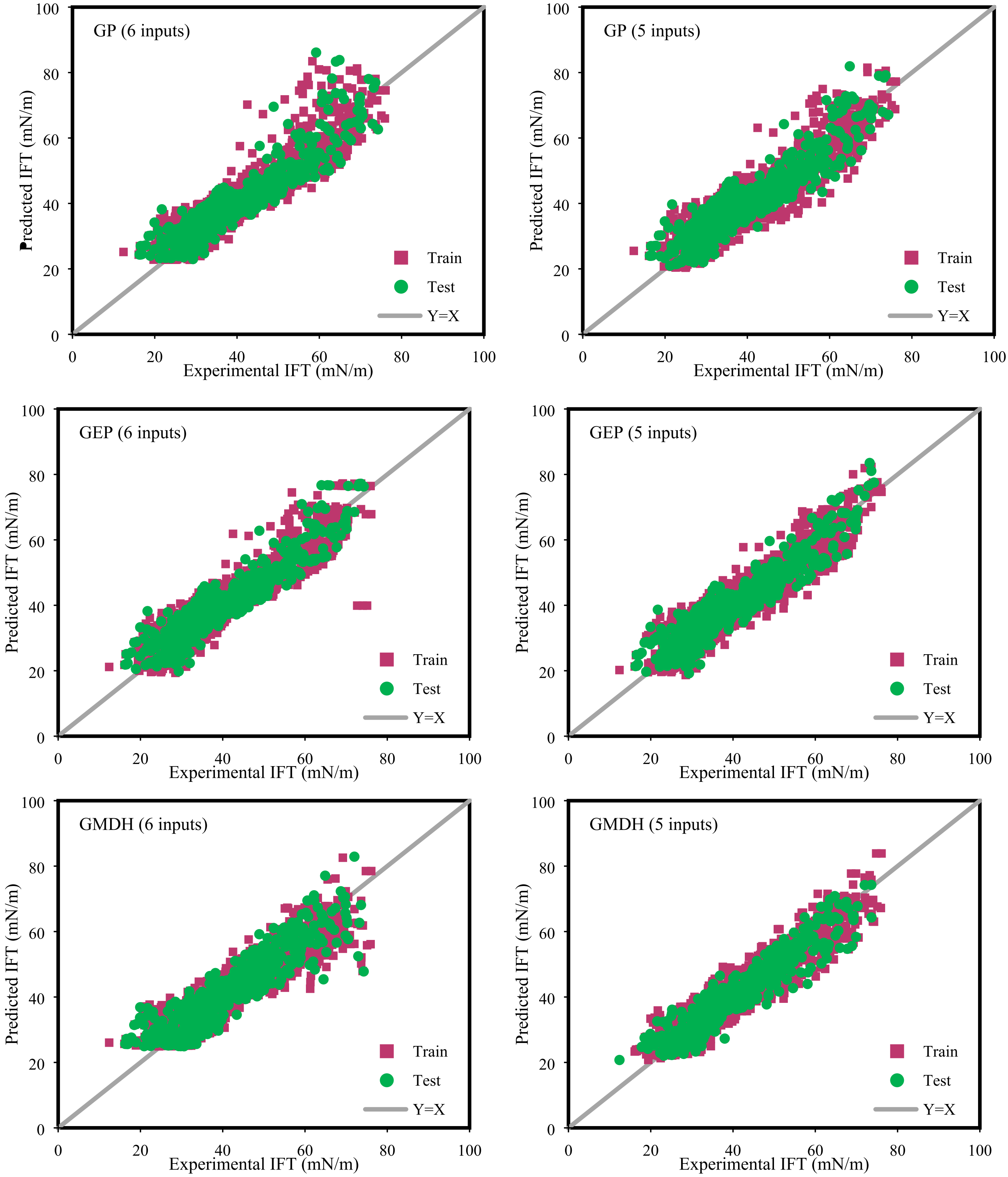


Figure 6. Cross-plots of the developed correlations.

temperature of 333 K, which surpasses the critical temperature of CO_2, the state of CO_2 is influenced by pressure. When the pressure remains below the critical pressure of CO_2, the CO_2 exists in a gaseous phase. However, as the pressure surpasses the critical point, CO_2 transitions into a supercritical state. In this context, the IFT displays a distinct behavior. At lower pressures, the IFT experiences a sharp decrease. This is primarily attributed to the enhancing (partial) dissolution of CO_2 gas within the aqueous phase coupled with its heightened affinity for adsorption at the interface with rising pressure. As pressure continues to rise, the CO_2-brine IFT reaches a pseudo-plateau, coinciding with the point at which CO_2 enters the supercritical state. Other researchers have also documented this phenomenon for the supercritical or liquid phase of

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

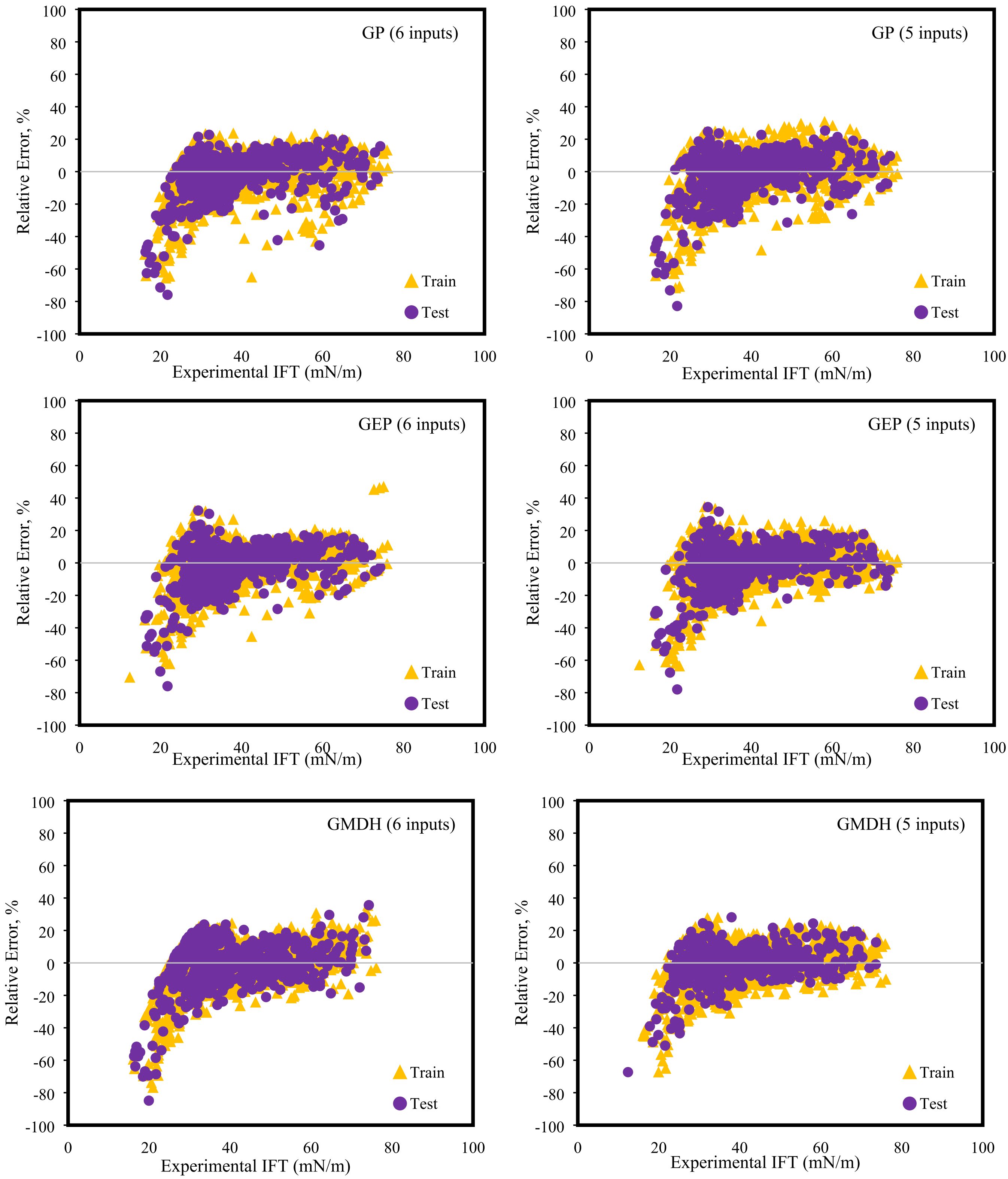


Figure 7. Error distribution plots of the developed correlations.

CO_2.[28,42,63,6928,42,63,69] During the supercritical or liquid phase, CO_2 transforms into a semifluid state with minimal or no compressibility, leading to nearly constant density differences within the CO_2-brine. Notably, the IFT of the CO_2-brine exhibits limited pressure dependency after achieving the pseudo-plateau state. Overall, the intricate temperature and pressure interactions impacting the IFT of CO_2-brine are found to be more intricate than was theoretically expected. This complexity can be attributed to the phase changes of CO_2 and its solubility effects within the brine, influenced by alterations in pressure and temperature.[63] As illustrated in Figure 9, both correlations effectively capture the trend of decreasing IFT with increasing pressure and provide accurate estimations.

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

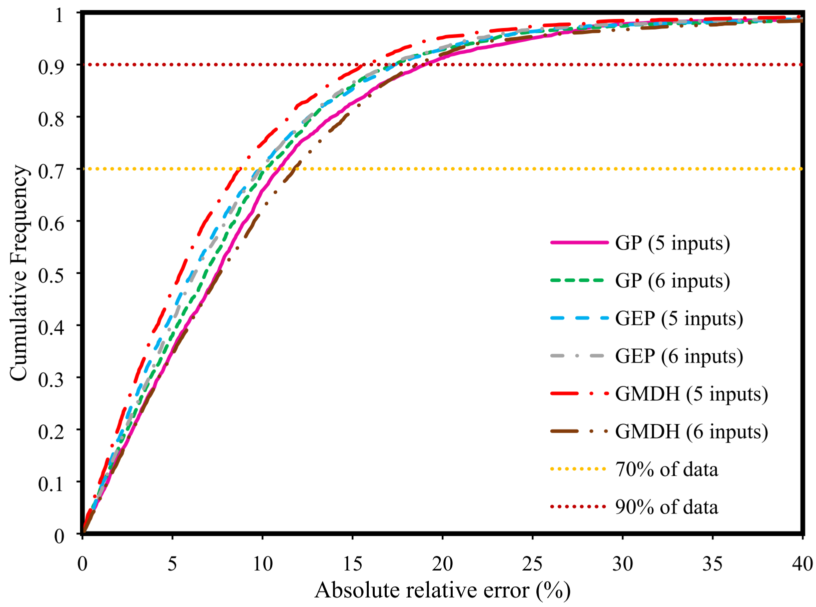


Figure 8. Cumulative frequency plot of the developed correlations.

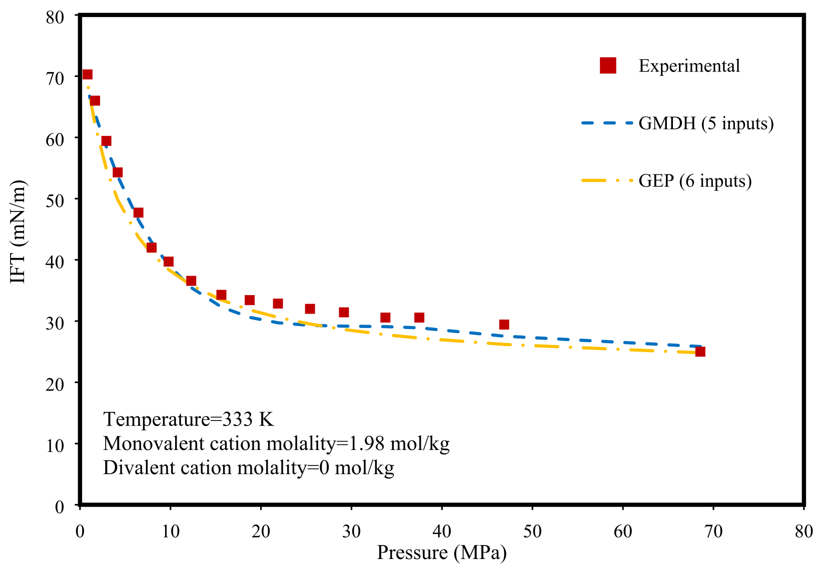


Figure 9. Experimental data[68] and prediction of correlations for the impact of pressure on IFT of a pure CO_2-brine system.

The subsequent analysis involves examining the influence of the temperature on the IFT of a pure CO_2-brine system composed of NaCl and KCl. This investigation maintains a monovalent cation molality of 1.98 mol/kg, while the divalent cation molality remains at zero. This experimental setup aligns with prior research detailed in the literature.[63] As depicted in Figure 10, there is a general trend of increasing the IFT for the CO_2-brine system as the temperature rises. This phenomenon is accurately captured by correlations, demonstrating low error in their predictions. While the presented example indicates an increasing trend, it is crucial to acknowledge that the impact of temperature on the CO_2-brine IFT is not always like this. In fact, the relationship between temperature and the CO_2-brine IFT is nonmonotonic, meaning it does not follow a consistent upward or downward trajectory. This behavior has been observed in various studies. The effect of temperature on IFT can vary based on the specific temperature and pressure ranges. In some cases, increasing temperature may lead to an increase in IFT, while in others, it could result in a decrease or no significant change in IFT.[42,51,55,63] This complex behavior is attributed to factors such as solubility changes and phase alterations induced by temperature and pressure variations. The interplay of these factors contributes to the intricate behavior of CO_2-brine IFT under different conditions.[63]

Moving forward, the investigation delves into the impact of salinity on the CO_2-brine IFT under constant temperature and pressure conditions. The objective is to evaluate the predictive capabilities of correlations in capturing IFT variations with changing salinity. To accomplish this, multiple CO_2-brine systems with different salinities are considered, featuring equivalent molalities of monovalent and divalent cations, as illustrated in Figure 11. This experimental setup aligns with prior research detailed in the literature.[55] An established notion suggests that as water salinity increases, the solubility of CO_2 in brine decreases.[90] Consequently, the IFT between CO_2 and brine exhibits a consistent upward trend in response to salinity

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

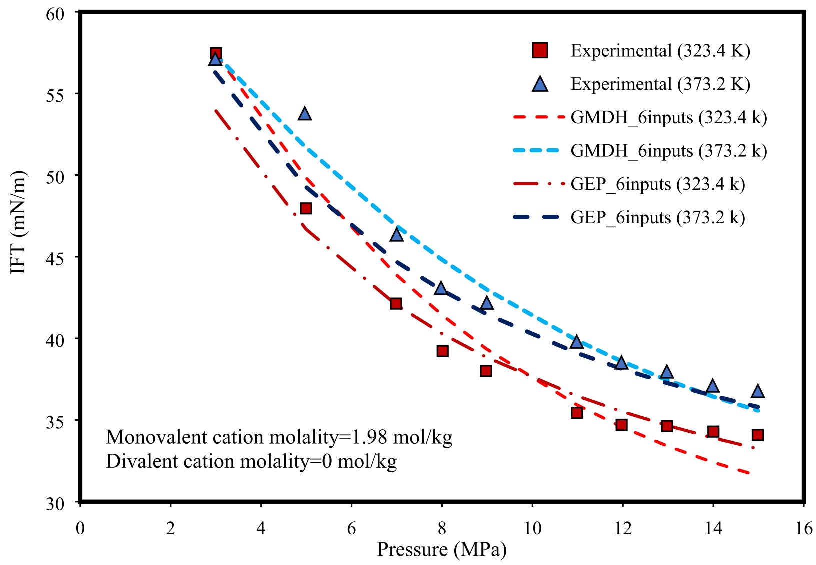


Figure 10. Experimental data[63] and prediction of correlations for the impact of temperature on IFT of a pure CO_2-brine system.

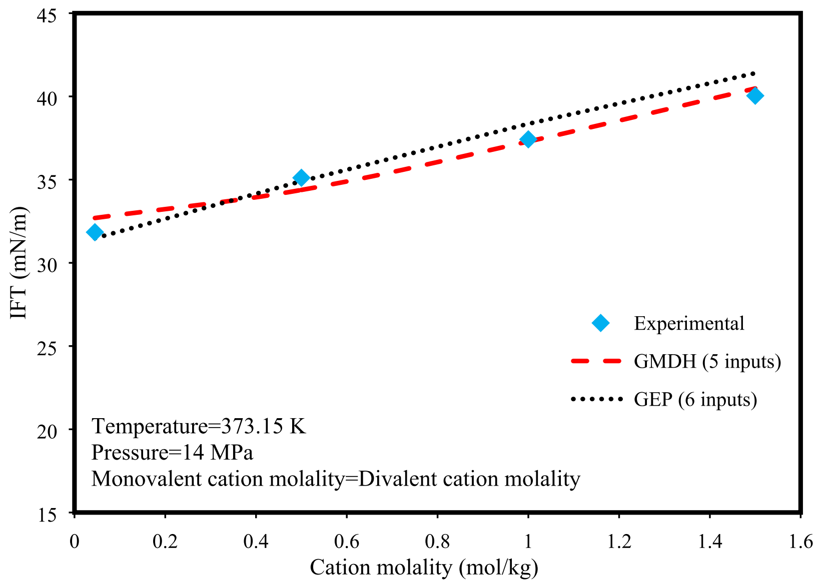


Figure 11. Experimental data[55] and prediction of correlations for the impact of salinity on IFT of a pure CO_2-brine system.

changes, regardless of the specific salt type. The mechanism underlying this behavior is intertwined with the distribution of cations within the aqueous phase. Typically, cations exhibit a minimal affinity for accumulation at the interface, resulting in their concentration in the bulk aqueous phase. This leads to a negative adsorption phenomenon at the interface, causing a depletion of cations. Consequently, water molecules tend to migrate from the interface into the solution to compensate for this effect.[91] As a consequence of this interaction, the mutual solubility between CO_2 and water is diminished, ultimately contributing to an increase in the IFT between them. This intricate interplay between cation distribution, water-CO_2 mutual solubility, and resultant IFT behavior emphasizes the complexity of these phenomena.[51] As shown in Figure 11, again, the developed correlations exhibit a promising level of accuracy in capturing the aforementioned trends.

Subsequently, we turn our attention to investigating the influence of impurities present in a gas stream on the IFT within the systems of impure CO_2 and water. This analysis is conducted under conditions of constant temperature and pressure. The objective is to gain insight into how the presence of impurities impacts the IFT between CO_2 and water. Taking into account the critical temperatures of CO_2, CH_4, and N_2 as 304.2, 190.85, and 126.2 K, respectively, the determination of Tcm can be achieved using eq 1. This calculation involves considering different compositions of CH_4, N_2, and CO_2. Given the substantial variance between the critical temperatures of these impurities and those of pure CO_2, it follows that the Tcm for an impure mixture is notably less than that of pure CO_2 gas. Consequently, a reduced Tcm indicates a diminished CO_2 composition while signifying an increased prevalence of impurities within the gas stream.[31,33,43] Figure 12 visually represents the influence of impurities on the IFT of CO_2-water under a constant pressure of 15 MPa and a temperature of 313.15 K, as experimentally investigated in the literature.[31,33] Additionally, the predictions generated by GMDH (5 inputs) are presented in the same context. The depicted figure underscores the accurate representation by the GMDH

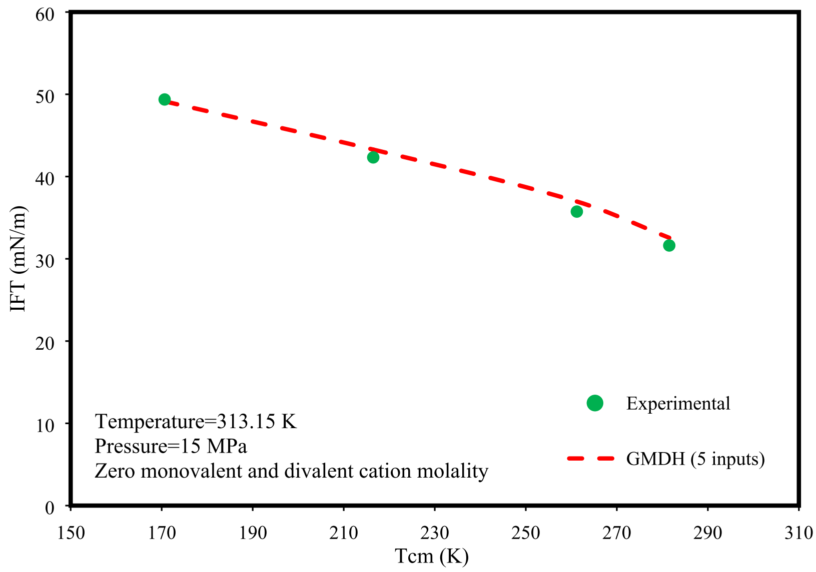


Figure 12. Experimental data[31,33] and prediction of GMDH correlation for the impact of impurities on IFT of impure CO_2-water systems.

correlation of the escalating pattern in the IFT corresponding to the rise in impurities (or, inversely, the decline in Tcm). The fundamental explanation for the observed IFT augmentation in the presence of non-CO_2 constituents lies in their diminished solubility, such as N_2, in comparison to the notable solubility of CO_2 in water.[27]

4.4. Sensitivity Analysis. In the context of parametric investigations, it is of great significance to ascertain the effects of various inputs on the resulting outcome. As a point to bear in mind, this study employed two distinct modeling methodologies, one involving 5 inputs and the other involving 6 inputs. The initial approach encompassed 6 inputs: temperature, pressure, monovalent cation molality, bivalent cation molality, mole fraction of N_2, and mole fraction of CH_4 within the injected gas. Conversely, the second approach encompassed 5 inputs, specifically temperature, pressure, monovalent cation molality, bivalent cation molality, and Tcm of the injected gas. In this context, the Pearson correlation coefficient[92] was calculated to determine the influence exerted by individual input variables on the response of both the GEP model (utilizing 6 inputs) and the GMDH model (utilizing 5 inputs), representing the most robust correlations established within this study. A greater value of the relevancy factor (r) associated with an input parameter signifies a more pronounced significance and impact on the IFT of CO_2-brine. The ensuing equation was employed to quantify the relevance factor:[93,94]

$$r(\mathrm{inp}_i, \mathrm{IFT}) = \frac{\sum_{j=1}^{n} (\mathrm{inp}_{i,j} - \mathrm{inp}_{a,i})(\mathrm{IFT}_j - \mathrm{IFT}_a)}{\sqrt{\sum_{j=1}^{n} (\mathrm{inp}_{i,j} - \mathrm{inp}_{a,i})^2 \sum_{j=1}^{n} (\mathrm{IFT}_j - \mathrm{IFT}_a)^2}} \quad (19)$$

where $\mathrm{inp}_{a,i}$ and $\mathrm{inp}_{i,j}$ denote the average value and the jth value of the ith input, respectively (where i can represent any of the input variables). Also, IFT_j represents the jth value of estimated IFT data and IFT_a stands for the average value of the IFT data. Figure 13 portrays the comparative influences of the input variables under consideration on the IFT between CO_2 and brine. As depicted in Figure 13, it is evident in both modeling methodologies that pressure exhibited the most substantial influence on the IFT of CO_2-brine, demonstrating a detrimental effect. Conversely, within the framework of both modeling strategies, temperature is observed to exert the least pronounced impact on the IFT. Nevertheless, as temperature exhibits a nonmonotonic nature, we hold the view that the acquired coefficient modestly underestimates its influence. In Figure 13a, it is evident that the influence on IFT ranges from the highest to the lowest in the following order: pressure, bivalent cation molality, mole fraction of CH_4, mole fraction of N_2, monovalent cation molality, and temperature. On the other hand, Figure 13a shows that pressure, Tcm, bivalent cation molality, monovalent cation molality, and temperature had the greatest to lowest influences on the IFT, respectively. Our analysis reveals a notable and adverse influence of the Tcm of CO_2-containing gas on the IFT between impure CO_2 and brine. This observation implies that any impurity possessing a Tc lower than that of CO_2 leads to an elevation in the IFT within the system and conversely. Consequently, heightened impurity concentrations align with lower Tcm values, consequently resulting in an elevated IFT. While this trend holds true for CH_4 and N_2, it is essential to verify the universality of this pattern for other non-CO_2 components that were not part of the current investigation. To conclude, the monovalent and divalent cation molalities demonstrate an escalating influence on the IFT, with divalent cations notably indicating approximately double the impact of monovalent cations.

4.5. Leverage Technique. The Leverage approach[95−97] is a potent technique in statistical analysis for identifying influential data points within regression models. Central to this method are standardized residuals (R) and Hat matrix leverage (H), which quantify deviations of predicted values from real data and measure the influence of individual observations, respectively. Critical leverage (H^*) value establishes a threshold for identifying high-leverage points.[98] The Leverage approach's elements form a comprehensive framework for assessing model reliability and data quality in regression analysis. Set within the mathematical realm of $-3 \leq R \leq 3$ and $H \leq H^*$, data points assume the distinguished role of "valid" representatives, resolutely navigating the intricate pathways of statistical thresholds. These valid data are in the

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

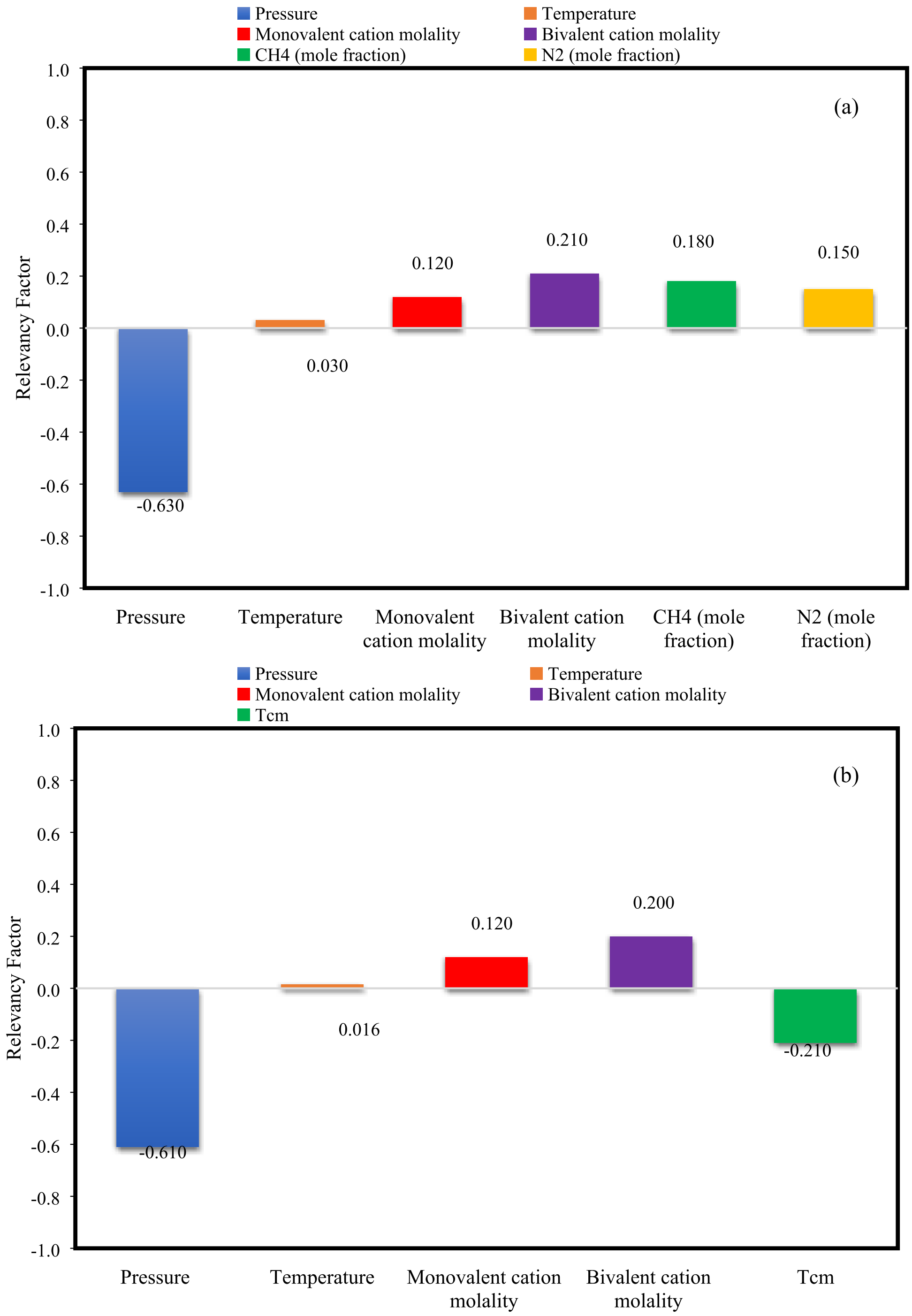


Figure 13. Sensitivity analysis using the results of (a) GEP (6 inputs) and (b) GMDH (5 inputs).

applicability domain of the model. Data points within $-3 \leq R \leq 3$ and $H^* \leq H$ are deemed "good high-leverage data", which are outside of the applicability domain of the model extending beyond statistical bounds yet well-predicted by the model. In contrast, data points exceeding R values of 3 or falling below -3 are categorized as "suspected" reflecting greater predictive uncertainty and are considered experimentally doubtful.[99] The utilization of William's plot becomes viable for delineating the range of applicability for the GEP model (6 inputs) and the GMDH model (5 inputs), representing the most robust correlation established within this study, as illustrated in Figure 14. Both of these correlations exhibit the majority of data falling within the realm of validity, and they are duly acknowledged for their credibility. The GEP model, featuring 6 inputs, showcased approximately 6.5% outliers alongside a mere 0.5% of suspected data. Similarly, the GMDH model, with 5 inputs, exhibited roughly 3.9% outliers and 1% suspected data. These findings collectively endorse the substantial application scope and reliability of both correlations, affirming the robust validity and reliability of the database employed in the modeling process.

5. SUMMARY AND CONCLUSIONS

This study employed three white-box machine learning (ML) models to forecast the IFT of the CO_2-brine systems. A total of

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

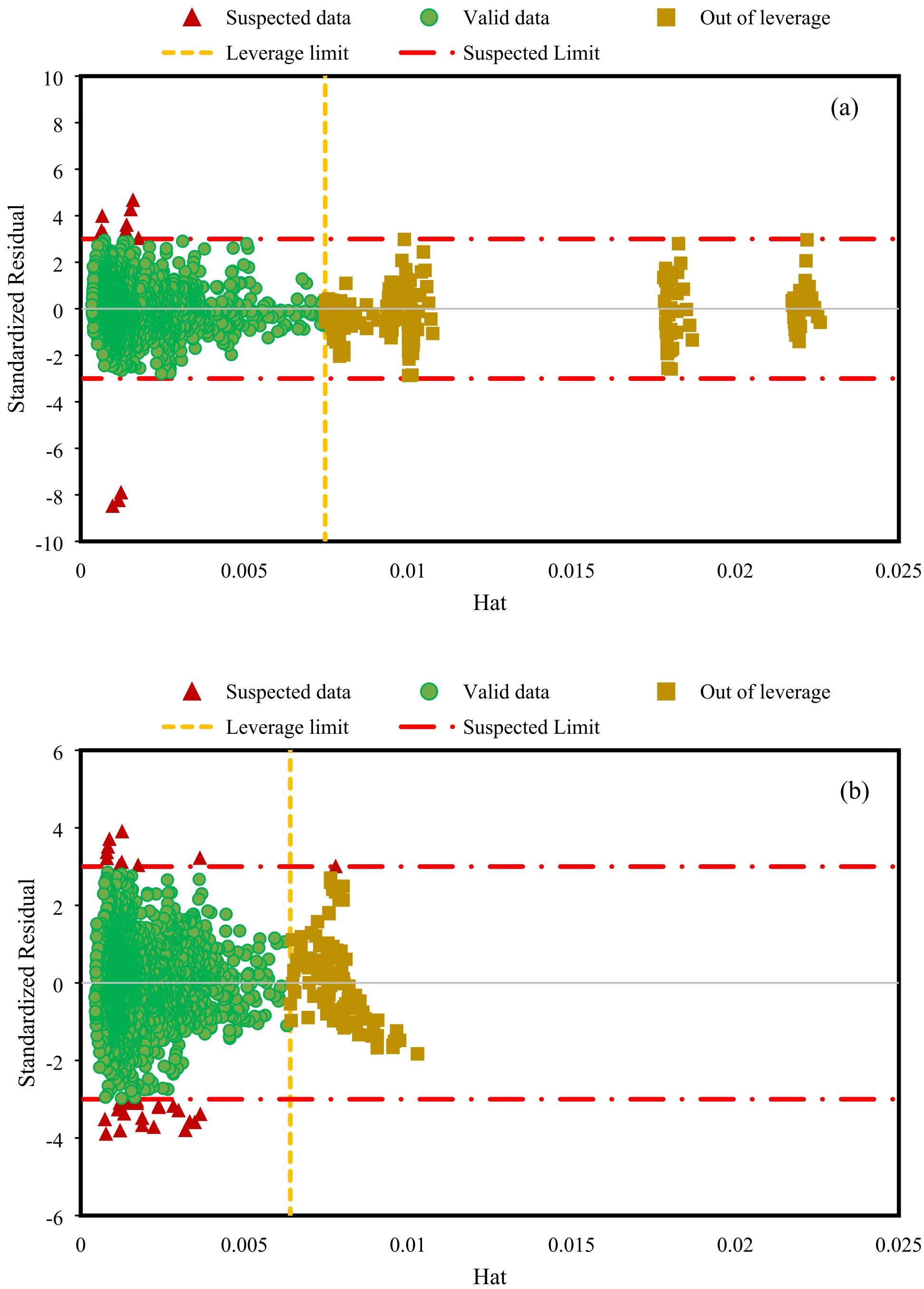


Figure 14. Identification of applicability scope of correlations: (a) GEP (6 inputs) and (b) GMDH (5 inputs).

2811 experimental IFT data points, encompassing diverse pressure and temperature operational conditions, were gathered from literature sources. These data incorporated impurities from both the gas (primarily N_2 and CH_4) and aqueous phases (including monovalent and divalent salt types). The following conclusions can be deduced from the acquired findings:

1. Among the correlations tested, the GMDH model, utilizing five inputs, showcased remarkable accuracy and reliability across all metrics. Its MAPE values for testing, training, and complete data sets were 7.63, 7.31, and 7.38%, respectively. Conversely, in the case of six-input models, the GEP correlation achieved superior precision, with MAPE values of 9.30, 8.06, and 8.31% for testing, training, and total data sets, respectively.
2. Taking into account all of the statistical parameters, the sequence of correlations based on accuracy is as follows: GMDH (5 inputs), GEP (5 inputs), GEP (6 inputs), GP (6 inputs), GP (5 inputs), and GMDH (6 inputs). The consolidated approach (5 inputs), which utilizes Tcm of a gaseous mixture rather than treating CH_4 and N_2 as separate input variables, effectively captures the essence of its constituents, represents impurities in CO_2-containing gas, and reduces the complexity of mathematical models with higher accuracy compared to those involving six inputs.

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

3. The sensitivity analysis indicated that pressure had the most significant impact on CO_2-brine IFT, resulting in a detrimental effect. Temperature, in contrast, exhibited the least pronounced influence on IFT, though its effect might be slightly underestimated due to its nonlinear behavior. In the context of the six-input modeling approach, pressure, bivalent cation molality, CH_4 mole fraction, N_2 mole fraction, monovalent cation molality, and temperature ranked from highest to lowest in terms of their impact on IFT. Similarly, within the five-input modeling approach, pressure, CO_2-containing gas pseudocritical temperature (Tcm), bivalent cation molality, monovalent cation molality, and temperature exerted the most to least influence on IFT, respectively.
4. There was an adverse influence of gas Tcm on the IFT between impure CO_2 and brine. This relationship leads to higher impurity concentrations aligning with lower Tcm values and subsequently elevated IFT. Moreover, monovalent and divalent cation molalities escalate the influence on IFT, with divalent cations exhibiting approximately double the influence of monovalent cations.
5. The Leverage approach confirmed both the strong reliability of the experimental data and the robust statistical validity of the best correlations established in this study. Specifically, the GEP model with 6 inputs revealed around 0.5% of potentially suspected data points, while the GMDH model with 5 inputs displayed approximately 1% of such instances.

AUTHOR INFORMATION

Corresponding Author

Xuguang Li − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*; orcid.org/0009-0003-0072-8309; Email: lixg41@cnooc.com.cn

Authors

Chen Wei-yu − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Lin Sun − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Jiyong Zhou − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Liping Huang − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Guang Xia − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Xiangli Meng − *CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300452, China*

Kui Wang − *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.3c07956

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to their colleagues in CNOOC for their support.

REFERENCES

(1) 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.

(2) Gibbins, J.; Chalmers, H. Carbon capture and storage. *Energy Policy* **2008**, *36* (12), 4317−4322.

(3) 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.

(4) 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.

(5) 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.

(6) 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.

(7) 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.

(8) OECD, I. *Energy and Air Pollution: World Energy Outlook Special Report 2016*; OECD, 2016.

(9) 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.

(10) 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.

(11) 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.

(12) Espinoza, D. N.; Santamarina, J. C. Water-CO2-mineral systems: Interfacial tension, contact angle, and diffusion—Implications to CO2 geological storage. *Water Resour. Res.* **2010**, *46* (7), No. 08634, DOI: 10.1029/2009WR008634.

(13) Amooie, M. A.; Soltanian, M. R.; Moortgat, J. Solutal convection in porous media: Comparison between boundary conditions of constant concentration and constant flux. *Phys. Rev. E* **2018**, *98* (3), No. 033118.

(14) Al Hameli, F.; Belhaj, H.; Al Dhuhoori, M. CO2 sequestration overview in geological formations: Trapping mechanisms matrix assessment. *Energies* **2022**, *15* (20), 7805.

(15) Holtzman, R.; Szulczewski, M. L.; Juanes, R. Capillary fracturing in granular media. *Phys. Rev. Lett.* **2012**, *108* (26), No. 264504.

(16) Weitz, D. A.; Stokes, J.; Ball, R.; Kushnick, A. Dynamic capillary pressure in porous media: Origin of the viscous-fingering length scale. *Phys. Rev. Lett.* **1987**, *59* (26), 2967.

(17) Gershenzon, N. I.; Ritzi, R. W., Jr; Dominic, D. F.; Soltanian, M.; Mehnert, E.; Okwen, R. T. Influence of small-scale fluvial architecture on CO2 trapping processes in deep brine reservoirs. *Water Resour. Res.* **2015**, *51* (10), 8240−8256.

(18) Taku Ide, S.; Jessen, K.; Orr, F. M., Jr Storage of CO2 in saline aquifers: Effects of gravity, viscous, and capillary forces on amount and timing of trapping. *Int. J. Greenhouse Gas Control* **2007**, *1* (4), 481−491.

(19) Doughty, C. Investigation of CO 2 plume behavior for a large-scale pilot test of geologic carbon storage in a saline formation. *Transp. Porous Media* **2010**, *82*, 49−76.

(20) Emami-Meybodi, H.; Hassanzadeh, H. Two-phase convective mixing under a buoyant plume of CO2 in deep saline aquifers. *Adv. Water Resour.* **2015**, *76*, 55−71.
(21) Martinez, M. J.; Hesse, M. A. Two-phase convective CO2 dissolution in saline aquifers. *Water Resour. Res.* **2016**, *52* (1), 585−599.
(22) Huppert, H. E.; Neufeld, J. A. The fluid mechanics of carbon dioxide sequestration. *Annu. Rev. Fluid Mech.* **2014**, *46*, 255−272.
(23) Bachu, S. Sequestration of CO2 in geological media in response to climate change: road map for site selection using the transform of the geological space into the CO2 phase space. *Energy Convers. Manage.* **2002**, *43* (1), 87−102.
(24) Blanco, S. T.; Rivas, C.; Fernandez, J.; Artal, M.; Velasco, I. Influence of methane in CO2 transport and storage for CCS technology. *Environ. Sci. Technol.* **2012**, *46* (23), 13016−13023.
(25) D'Alessandro, D. M.; Smit, B.; Long, J. R. Carbon dioxide capture: prospects for new materials. *Angew. Chem., Int. Ed.* **2010**, *49* (35), 6058−6082.
(26) Farrelly, D. J.; Everard, C. D.; Fagan, C. C.; McDonnell, K. P. Carbon sequestration and the role of biological carbon mitigation: a review. *Renewable sustainable Energy Rev.* **2013**, *21*, 712−727.
(27) Jafari Raad, S. M.; Hassanzadeh, H. Does impure CO2 impede or accelerate the onset of convective mixing in geological storage? *Int. J. Greenhouse Gas Control* **2016**, *54*, 250−257.
(28) Aggelopoulos, C. A.; Robin, M.; Perfetti, E.; Vizika, O. CO2/CaCl2 solution interfacial tensions under CO2 geological storage conditions: influence of cation valence on interfacial tension. *Adv. Water Resour.* **2010**, *33* (6), 691−697.
(29) Chiquet, P.; Daridon, J.-L.; Broseta, D.; Thibeau, S. CO2/water interfacial tensions under pressure and temperature conditions of CO2 geological storage. *Energy Convers. Manage.* **2007**, *48* (3), 736−744.
(30) Georgiadis, A.; Maitland, G.; Trusler, J. M.; Bismarck, A. Interfacial tension measurements of the (H2O+ CO2) system at elevated pressures and temperatures. *J. Chem. Eng. Data* **2010**, *55* (10), 4168−4175.
(31) Ren, Q.-Y.; Chen, G.-J.; Yan, W.; Guo, T.-M. Interfacial tension of (CO2+ CH4)+ water from 298 to 373 K and pressures up to 30 MPa. *J. Chem. Eng. Data* **2000**, *45* (4), 610−612.
(32) Li, X.; Boek, E.; Maitland, G. C.; Trusler, J. M. Interfacial Tension of (Brines+ CO2):(0.864 NaCl+ 0.136 KCl) at Temperatures between (298 and 448) K, Pressures between (2 and 50) MPa, and Total Molalities of (1 to 5) mol·kg−1. *J. Chem. Eng. Data* **2012**, *57* (4), 1078−1088.
(33) Yan, W.; Zhao, G.-Y.; Chen, G.-J.; Guo, T.-M. Interfacial tension of (methane + nitrogen) + water and (carbon dioxide + nitrogen)+ water systems. *J. Chem. Eng. Data* **2001**, *46* (6), 1544−1548.
(34) Bachu, S.; Bennion, D. B. Interfacial tension between CO2, freshwater, and brine in the range of pressure from (2 to 27) MPa, temperature from (20 to 125) C, and water salinity from (0 to 334 000) mg· L−1. *J. Chem. Eng. Data* **2009**, *54* (3), 765−775.
(35) Hebach, A.; Oberhof, A.; Dahmen, N.; Kögel, A.; Ederer, H.; Dinjus, E. Interfacial tension at elevated pressures measurements and correlations in the water+ carbon dioxide system. *J. Chem. Eng. Data* **2002**, *47* (6), 1540−1546.
(36) Kvamme, B.; Kuznetsova, T.; Hebach, A.; Oberhof, A.; Lunde, E. Measurements and modelling of interfacial tension for water+ carbon dioxide systems at elevated pressures. *Comput. Mater. Sci.* **2007**, *38* (3), 506−513.
(37) Chen, C.; Chai, Z.; Shen, W.; Li, W. Effects of impurities on CO2 sequestration in saline aquifers: Perspective of interfacial tension and wettability. *Ind. Eng. Chem. Res.* **2018**, *57* (1), 371−379.
(38) Nielsen, L. C.; Bourg, I. C.; Sposito, G. Predicting CO2−water interfacial tension under pressure and temperature conditions of geologic CO2 storage. *Geochim. Cosmochim. Acta* **2012**, *81*, 28−38.
(39) Iglauer, S.; Mathew, M.; Bresme, F. Molecular dynamics computations of brine−CO2 interfacial tensions and brine−CO2−quartz contact angles and their effects on structural and residual trapping mechanisms in carbon geo-sequestration. *J. Colloid Interface Sci.* **2012**, *386* (1), 405−414.
(40) Massoudi, R.; King, A., Jr Effect of pressure on the surface tension of water. Adsorption of low molecular weight gases on water at 25. deg. *J. Phys. Chem. A* **1974**, *78* (22), 2262−2266.
(41) Bennion, D. B.; Bachu, S. In *A Correlation of the Interfacial Tension between Supercritical Phase CO2 and Equilibrium Brines as a Function of Salinity, Temperature and Pressure*, SPE Annual Technical Conference and Exhibition? SPE-114479-MS; SPE, 2008.
(42) Chalbaud, C.; Robin, M.; Lombard, J.; Martin, F.; Egermann, P.; Bertin, H. Interfacial tension measurements and wettability evaluation for geological CO2 storage. *Adv. Water Resour.* **2009**, *32* (1), 98−109.
(43) Li, Z.; Wang, S.; Li, S.; Liu, W.; Li, B.; Lv, Q.-C. Accurate determination of the CO2−brine interfacial tension using graphical alternating conditional expectation. *Energy Fuels* **2014**, *28* (1), 624−635.
(44) 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.
(45) 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.
(46) Zhang, J.; Feng, Q.; Wang, S.; Zhang, X.; Wang, S. Estimation of CO2−brine interfacial tension using an artificial neural network. *J. Supercrit. Fluids* **2016**, *107*, 31−37.
(47) Kamari, A.; Pournik, M.; Rostami, A.; Amirlatifi, A.; Mohammadi, A. H. Characterizing the CO2-brine interfacial tension (IFT) using robust modeling approaches: A comparative study. *J. Mol. Liq.* **2017**, *246*, 32−38.
(48) Niroomand-Toomaj, E.; Etemadi, A.; Shokrollahi, A. Radial basis function modeling approach to prognosticate the interfacial tension CO2/Aquifer Brine. *J. Mol. Liq.* **2017**, *238*, 540−544.
(49) Partovi, M.; Mosalanezhad, M.; Lotfi, S.; Barati-Harooni, A.; Najafi-Marghmaleki, A.; Mohammadi, A. H. On the estimation of CO2-brine interfacial tension. *J. Mol. Liq.* **2017**, *243*, 265−272.
(50) Rashid, S.; Harimi, B.; Hamidpour, E. Prediction of CO2-Brine interfacial tension using a rigorous approach. *J. Nat. Gas Sci. Eng.* **2017**, *45*, 108−117.
(51) Amooie, M. A.; Hemmati-Sarapardeh, A.; Karan, K.; Husein, M. M.; Soltanian, M. R.; Dabir, B. Data-driven modeling of interfacial tension in impure CO2-brine systems with implications for geological carbon storage. *Int. J. Greenhouse Gas Control* **2019**, *90*, No. 102811.
(52) Nait Amar, M. Towards improved genetic programming based-correlations for predicting the interfacial tension of the systems pure/impure CO2-brine. *J. Taiwan Inst. Chem. Eng.* **2021**, *127*, 186−196.
(53) Safaei-Farouji, M.; Thanh, H. V.; Dashtgoli, D. S.; Yasin, Q.; Radwan, A. E.; Ashraf, U.; Lee, K.-K. Application of robust intelligent schemes for accurate modelling interfacial tension of CO2 brine systems: Implications for structural CO2 trapping. *Fuel* **2022**, *319*, No. 123821.
(54) Zhang, J.; Feng, Q.; Zhang, X.; Shu, C.; Wang, S.; Wu, K. A supervised learning approach for accurate modeling of CO2−brine interfacial tension with application in identifying the optimum sequestration depth in saline aquifers. *Energy Fuels* **2020**, *34* (6), 7353−7362.
(55) 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.
(56) Bachu, S.; Bennion, D. B. Dependence of CO2-brine interfacial tension on aquifer pressure, temperature and water salinity. *Energy Procedia* **2009**, *1* (1), 3157−3164.
(57) Bikkina, P. K.; Shoham, O.; Uppaluri, R. Equilibrated interfacial tension data of the CO2−water system at high pressures and

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

moderate temperatures. *J. Chem. Eng. Data* **2011**, *56* (10), 3725−3733.

(58) Chalbaud, C.; Robin, M.; Egermann, P. In *Interfacial Tension Data and Correlations of Brine/CO2 Systems under Reservoir Conditions*, SPE Annual Technical Conference and Exhibition? SPE-102918-MS; SPE, 2006.

(59) Li, X.; Boek, E. S.; Maitland, G. C.; Trusler, J. M. Interfacial Tension of (Brines+ CO2): CaCl2 (aq), MgCl2 (aq), and Na2SO4 (aq) at Temperatures between (343 and 423) K, Pressures between (2 and 50) MPa, and Molalities of (0.5 to 5) mol· kg−1. *J. Chem. Eng. Data* **2012**, *57* (5), 1369−1375.

(60) Arif, M.; Jones, F.; Barifcani, A.; Iglauer, S. Electrochemical investigation of the effect of temperature, salinity and salt type on brine/mineral interfacial properties. *Int. J. Greenhouse Gas Control* **2017**, *59*, 136−147.

(61) Arif, M.; Al-Yaseri, A. Z.; Barifcani, A.; Lebedev, M.; Iglauer, S. Impact of pressure and temperature on CO2−brine−mica contact angles and CO2−brine interfacial tension: Implications for carbon geo-sequestration. *J. Colloid Interface Sci.* **2016**, *462*, 208−215.

(62) Sarmadivaleh, M.; Al-Yaseri, A. Z.; Iglauer, S. Influence of temperature and pressure on quartz−water−CO2 contact angle and CO2−water interfacial tension. *J. Colloid Interface Sci.* **2015**, *441*, 59−64.

(63) Mutailipu, M.; Liu, Y.; Jiang, L.; Zhang, Y. Measurement and estimation of CO2−brine interfacial tension and rock wettability under CO2 sub-and super-critical conditions. *J. Colloid Interface Sci.* **2019**, *534*, 605−617.

(64) Mutailipu, M.; Jiang, L.; Fu, J.; Wang, Z.; Yu, T.; Lu, Z.; Liu, Y. Effects of Na+, K+, Ca2+, and Mg2+ cations on CO2−brine interfacial tension under offshore storage conditions. *Greenhouse Gases: Sci. Technol.* **2018**, *8* (4), 762−780.

(65) Chun, B.-S.; Wilkinson, G. T. Interfacial tension in high-pressure carbon dioxide mixtures. *Ind. Eng. Chem. Res.* **1995**, *34* (12), 4371−4377.

(66) Al-Anssari, S.; Barifcani, A.; Keshavarz, A.; Iglauer, S. Impact of nanoparticles on the CO2-brine interfacial tension at high pressure and temperature. *J. Colloid Interface Sci.* **2018**, *532*, 136−142.

(67) Akutsu, T.; Yamaji, Y.; Yamaguchi, H.; Watanabe, M.; Smith, R. L., Jr; Inomata, H. Interfacial tension between water and high pressure CO2 in the presence of hydrocarbon surfactants. *Fluid Phase Equilib.* **2007**, *257* (2), 163−168.

(68) Pereira, L. *Interfacial Tension of Reservoir Fluids: An Integrated Experimental and Modelling Investigation*; Heriot-Watt University, 2016.

(69) Pereira, L. M.; Chapoy, A.; Burgass, R.; Oliveira, M. B.; Coutinho, J. A.; Tohidi, B. Study of the impact of high temperatures and pressures on the equilibrium densities and interfacial tension of the carbon dioxide/water system. *J. Chem. Thermodyn.* **2016**, *93*, 404−415.

(70) Pereira, L. M.; Chapoy, A.; Burgass, R.; Tohidi, B. Interfacial tension of CO2+ brine systems: Experiments and predictive modelling. *Adv. Water Resour.* **2017**, *103*, 64−75.

(71) Wang, H.; Luo, M.; Lun, Z.; Lv, C.; Zhao, C.; Zhao, Q.; Lang, D. Determination and Comparison of Interfacial Interactions between CO 2, Crude Oil, and Brine at Reservoir Conditions. *Energy Fuels* **2018**, *32* (8), 8187−8192.

(72) Danesh, A. *PVT and Phase Behaviour of Petroleum Reservoir Fluids*; Elsevier, 1998.

(73) Koza, J. R. *Genetic Programming II: Automatic Discovery of Reusable Programs*; MIT Press, 1994.

(74) Okhovat, A.; Mousavi, S. M. Modeling of arsenic, chromium and cadmium removal by nanofiltration process using genetic programming. *Appl. Soft Comput.* **2012**, *12* (2), 793−799.

(75) 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.

(76) Suh, C.; Choi, B.; Lee, S.; Kim, D.; Cho, J. Application of genetic programming to develop the model for estimating membrane damage in the membrane integrity test using fluorescent nanoparticle. *Desalination* **2011**, *281*, 80−87.

(77) Naghizadeh, A.; Larestani, A.; Amar, M. N.; Hemmati-Sarapardeh, A. Predicting viscosity of CO2−N2 gaseous mixtures using advanced intelligent schemes. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109359.

(78) Rostami, A.; Arabloo, M.; Ebadi, H. Genetic programming (GP) approach for prediction of supercritical CO2 thermal conductivity. *Chem. Eng. Res. Des.* **2017**, *122*, 164−175.

(79) Ferreira, C. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. 2001, arXiv:cs/0102027. arXiv.org e-Print archive. https://arxiv.org/abs/cs/0102027.

(80) Teodorescu, L.; Sherwood, D. High energy physics event selection with gene expression programming. *Comput. Phys. Commun.* **2008**, *178* (6), 409−419.

(81) 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.

(82) Ivakhnenko, A. G. Polynomial theory of complex systems. *IEEE Trans. Syst. Man Cybern.* **1971**, 364−378.

(83) Ebtehaj, I.; Bonakdari, H.; Zaji, A. H.; Azimi, H.; Khoshbin, F. GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. *Eng. Sci. Technol.* **2015**, *18* (4), 746−757.

(84) Farlow, S. Self-Organizing Method in Modeling: GMDH. In *Type Algorithm*; GMDH, 1984.

(85) Iba, H.; deGaris, H.; Sato, T. A numerical approach to genetic programming for system identification. *Evol. Comput.* **1995**, *3* (4), 417−452.

(86) Jamali, A. Pareto Robust Design of Controllers with Probabilistic Uncertainties using Multi Objective Evolutionary Algorithms. Ph.D. Thesis, Guilan University: Guilan, Iran, 2009.

(87) Nariman-Zadeh, N.; Darvizeh, A.; Jamali, A.; Moeini, A. Evolutionary design of generalized polynomial neural networks for modelling and prediction of explosive forming process. *J. Mater. Process. Technol.* **2005**, *164*−*165*, 1561−1571.

(88) Mohammadi, M.-R.; Hadavimoghaddam, F.; Atashrouz, S.; Hemmati-Sarapardeh, A.; Abedi, A.; Mohaddespour, A. Application of robust machine learning methods to modeling hydrogen solubility in hydrocarbon fuels. *Int. J. Hydrogen Energy* **2022**, *47* (1), 320−338.

(89) Sapkale, G.; Patil, S.; Surwase, U.; Bhatbhage, P. Supercritical fluid extraction. *Int. J. Chem. Sci.* **2010**, *8* (2), 729−743.

(90) Enick, R. M.; Klara, S. M. CO2 solubility in water and brine under reservoir conditions. *Chem. Eng. Commun.* **1990**, *90* (1), 23−33.

(91) Moeini, F.; Hemmati-Sarapardeh, A.; Ghazanfari, M.-H.; Masihi, M.; Ayatollahi, S. Toward mechanistic understanding of heavy crude oil/brine interfacial tension: The roles of salinity, temperature and pressure. *Fluid Phase Equilib.* **2014**, *375*, 191−200.

(92) 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.

(93) 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.

(94) Hemmati-Sarapardeh, A.; Aminshahidy, B.; Pajouhandeh, A.; Yousefi, S. H.; Hosseini-Kaldozakh, S. A. A soft computing approach for the determination of crude oil viscosity: light and intermediate crude oil systems. *J. Taiwan Inst. Chem. Eng.* **2016**, *59*, 1−10.

(95) Leroy, A. M.; Rousseeuw, P. J. *Robust Regression and Outlier Detection*; John Wiley & Sons, 1987.

(96) Goodall, C. R. 13 Computation using the QR decomposition. In *Handbook of Statistics*; Elsevier, 1993.

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957

(97) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26* (5), 694−701.
(98) 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.
(99) Hadavimoghaddam, F.; Mohammadi, M.-R.; Atashrouz, S.; Bostani, A.; Hemmati-Sarapardeh, A.; Mohaddespour, A. Modeling hydrogen solubility in alcohols using group method of data handling and genetic programming. *Int. J. Hydrogen Energy* **2023**, *48* (7), 2689−2704.

https://doi.org/10.1021/acsomega.3c07956
ACS Omega 2024, 9, 7937−7957