ACS OMEGA

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

Article

Modeling Viscosity of CO_2−N_2 Gaseous Mixtures Using Robust Tree-Based Techniques: Extra Tree, Random Forest, GBoost, and LightGBM

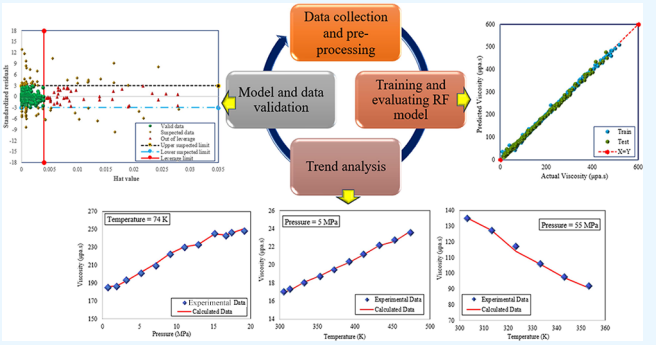
Haimin Zheng,*[,⊥] Atena Mahmoudzadeh,[⊥] Behnam Amiri-Ramsheh, and Abdolhossein Hemmati-Sarapardeh*

Cite This: *ACS Omega* 2023, 8, 13863−13875

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Carbon dioxide (CO_2) has an essential role in most enhanced oil recovery (EOR) methods in the oil industry. Oil swelling and viscosity reduction are the dominant mechanisms in an immiscible CO_2-EOR process. Besides numerous CO_2 applications in EOR, most oil reservoirs do not have access to natural CO_2, and capturing it from flue gas and other sources is costly. Flue gases are available in huge quantities at a significantly lower price and can be considered economically viable agents for EOR operations. In this work, four powerful machine learning algorithms, namely, extra tree (ET), random forest (RF), gradient boosting (GBoost), and light gradient boosted machine (LightGBM) were utilized to accurately estimate the viscosity of CO_2−N_2 mixtures. To this aim, a databank was employed, containing 3036 data points over wide ranges of pressures and temperatures. Temperature, pressure, and CO_2 mole fraction were applied as input parameters, and the viscosity of the CO_2−N_2 mixture was the output. The RF smart model had the highest precision with the lowest average absolute percent relative error (AAPRE) of 1.58%, root mean square error (RMSE) of 2.221, and determination coefficient (R^2) of 0.9993. The trend analysis showed that the RF model could precisely predict the real physical behavior of the CO_2−N_2 viscosity variation. Finally, the outlier detection was performed using the leverage approach to demonstrate the validity of the utilized databank and the applicability area of the developed RF model. Accordingly, nearly 96% of the data points seemed to be dependable and valid, and the rest of them were located in the suspected and out-of-leverage data zones.



1. INTRODUCTION

Carbon dioxide (CO_2) has an essential role in some enhanced oil recovery (EOR) methods in the oil industry. CO_2 can be utilized to improve oil recovery in a broad range of oil reservoirs.[1] Oil swelling and viscosity reduction are the dominant mechanisms in an immiscible CO_2-EOR process.[2] Some developments have combined CO_2 injection with other injection methods to minimize problems, such as fingering and early breakthroughs in case of continuous CO_2 flooding.[3] These advancements include CO_2 water-alternating-gas (WAG) injection, which benefits from both advantages of water flooding and gas injection by improving the macroscopic and microscopic sweep efficiency[4,5] and mobility control through relative permeability reduction,[6] and surfactant-assisted CO_2 injection that balances the low CO_2 viscosity and diverts flow to low permeable regions by making foams.[7,8] Furthermore, two relatively new methods, polymer-assisted CO_2 injection,[9−11] and nanoparticle-assisted CO_2 flooding[12−14] claim to overcome some drawbacks of previous methods such as oil trapping in WAG[15] and instabilities in foam.[16] Also, polymer decreases gas solubility due to molecular weight increment.[17]

Besides numerous CO_2 applications in EOR, most oil reservoirs do not have access to natural CO_2, and capturing it from flue gas and other sources is costly.[18] As a result of the fact that flue gases are available in huge quantities at a significantly lower price, the oil industry has therefore given flue gas a great amount of attention.[19,20] Also, it is considered as an environmental solution for reducing greenhouse gas emissions.[21] The composition of flue gas is dependent on the fuel used in the combustion process but consists mainly of N_2, CO_2, and a small fraction of water vapor, O_2, and SO_2.[22]

Received: January 12, 2023
Accepted: March 23, 2023
Published: April 6, 2023

Table 1. Existing Empirical Correlations along with Their Operational Conditions

study	correlation	operational conditions	ref
	$A = \left(0.0038539 + \frac{T + 459.6}{1.8 \times 0.0000356}\right) - \left(\left(\frac{0.0004131 + (T + 459.6)}{1.8 \times 0.0000016}\right) \times \sqrt{M_w}\right)$		
Chen and Ruth	$\mu = \left((-0.488439 + P_r \times (-0.0943952) + P_r^2 \times 0.01199591) \times \frac{1}{T_r} + (0.8269923 + P_r \times 1.71241 + P_r^2 \times (-0.0700968)) \times \frac{1}{T_r^4} + (1.20700968 + P_r \times 0.0301188 + P_r^2 \times (-0.0048318))\right) \times A$	not reported	38
	$A = -0.141645 + 0.018076 \times P_r + 0.00214 \times P_r^2 - 0.004192 \times \ln(P_r) - 0.000386 \times \ln^2(P_r)$		
	$B = \frac{0.187138}{T_r} + (0.569211 \times \ln^2(T_r))$		
Sanjari et al.	$C = 1 + 0.000387 \times P_r^2$	P_r between 0.01 and 21 and T_r between 1 and 3	45
	$D = \frac{-2.857176}{T_r} + \frac{2.925776}{T_r^2} + \frac{-1.062425}{T_r^3}$		
	$\mu = \frac{A(P_r) + B(T_r)}{C(P_r) + D(T_r)}$		
Standing	$\mu = (1.79 \times 10^{-5} - 2.062 \times 10^{-6} \times \gamma_g) \times (T + 459.6) + 0.008188 - 0.00615 \times \log \gamma_g$	high temperatures and pressures	36
Heidaryan et al.	$A = 1.022872 - 1.651432 \times \frac{M_w}{T} + 5.757386 \times \left(\frac{M_w}{T}\right)^2 - 7.389282 \times 0.01 \times \rho + 8.389065 \times 0.01 \times \rho^2 + 2.977476 \times 0.1 \times \rho^3$	temperatures (up to 400 K) and pressures (up to 14,000 KPa)	46
	$B = 1 - 1.451318 \times \frac{M_w}{T} + 4.682506 \times \left(\frac{M_w}{T}\right)^2 + 1.918239 \times \left(\frac{M_w}{T}\right)^3 - 9.8449 \times 0.01 \times \rho$		
	$\mu = \ln \frac{A}{B}$		

In recent years, several studies have been conducted comparing the effectiveness of flue gas and pure CO_2 in EOR methods. Johnson et al.[23] demonstrated that the flue gas huff-n-puff technique can likely be economically performed in some shallow reservoirs. Zhang et al.[24] investigated huff-n-puff cyclic gas injection to improve light oil recovery by CO_2/flue gas using core flood tests and adjusting the Peng−Robinson (PR) equation of state (EoS) to match laboratory data. They indicated that the injection of flue gas was the most efficient approach, whereas CO_2 injection was the least efficient. In the petroleum industry, making accurate predictions of gas viscosity is truly necessary based on its effect on reservoir recovery, fluid flow, and storage. In the absence of experimental measurement, empirical correlations are used to determine the viscosity of gases.[25]

The first attempt to calculate the viscosity of CO_2−N_2 mixtures was made by Leidenfrost and co-workers.[26] They conducted experiments using an oscillating-disk viscometer. At 20 °C, the pressure range was between 1 and 20 atmospheres (2.13 MPa) with mole fraction variations. They developed their work on three other binary mixtures, including He/Kr, He/N_2, and Ar/CO_2 by investigating the temperature effect in the range of 20−30 °C. Based on the measured data points, a second-order polynomial empirical density correlation depends on the mole fraction, and a temperature-free term was generated by Kestin et al.[27] In 1974, Kestin and Ro[28] measured tertiary gas mixtures in addition to binary mixtures at low densities and temperatures varying from 25 to 700 °C. Gururaja et al.[29] measured the binary system viscosity including CO_2−N_2 with the aid of an oscillating disk using the capillary viscometer method to measure the viscosity of a wide range of pure gases and mixtures. Depending on low or high pressures, slip with gas expansion and kinetic correction must be considered in the capillary method.[30] In a recent study, the viscosity of CO_2−N_2 was tested by two independent rotating body viscometers at a low pressure of 0.1 MPa and high pressures between 3 and 8 MPa in the temperature range of 273−473 K.[31]

Although experimental measurements are the most accurate approaches to determining the viscosity of a gas, practical difficulties and measurement limits caused the development of a significant number of correlations. In 1954, Carr et al.[32] proposed a two-step graphical correlation method. First, natural gas viscosity was determined using the apparent molecular weight at atmospheric pressure and then corrected by a second plate on desired pressure and temperature.

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

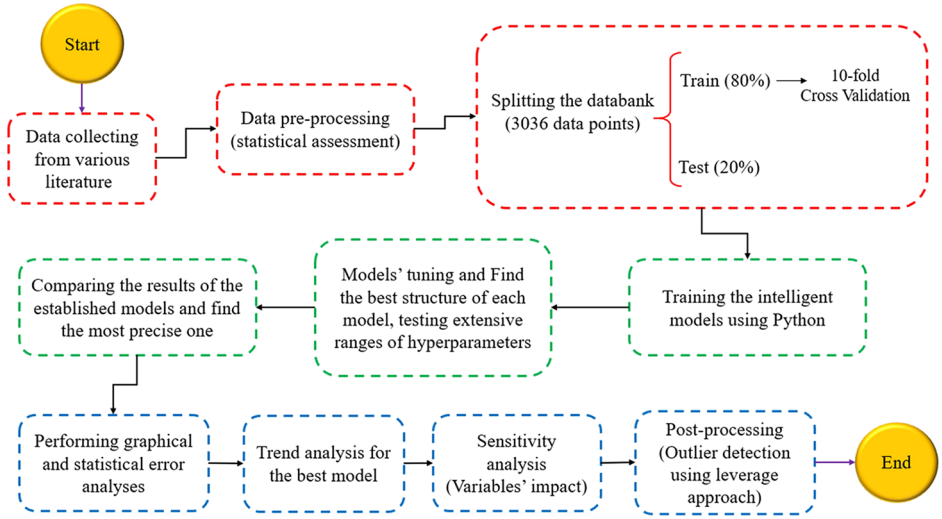


Figure 1. Overview of the present research.

Lohrenz et al.[33] calculated the viscosities of in situ reservoir gases and liquids and their results showed an average absolute percent relative error (AAPRE) of 16%. Whitson and Brule[34] mentioned that some changes in the correlation developed by Dean and Stiel[35] in combination with Standing's correlation,[36] where applicable in high pressure, give acceptable results to gas viscosity estimation. Vesovic and Wakeham[37] introduced a gas viscosity correlation considering thermodynamic characteristics at the molecular level. Chen and Ruth[38] performed a comparative evaluation of the well-known natural gas correlations that had been established prior to 1993 and claimed that Dranchuk et al.[39] correlation provided the most accurate value of the viscosity ratio. In 2014, Jarrahian and Heidaryan[40] developed a correlation on 29 multi-component mixtures containing 3231 data points in the pressure range of 0.1−137.8 MPa and the temperature range of 0.1−137.8 °C. Their research described gas viscosity as a function of diluted viscosity, pseudo-reduced pressure, pseudo-reduced temperature,[41] and pseudo-critical parameters computed using the Standing technique.[36] Sanjari et al.[42] suggested a model of natural gas viscosity employing molecular weight, density, and temperature. Their correlation's AAPRE was less than 1%. Yang et al.[43] established a semi-theoretical model based on a theory that relates natural gas viscosity to temperature and density, known as the kinetic theory of gas.[44] They reported an AAPRE of less than 1.9% in the temperature and pressure ranges of 250−450 K and 0.1−140.0 MPa, respectively. Table 1 represents the mathematical correlations of Chen and Ruth,[38] Sanjari et al.,[45] Standing,[36] and Heidaryan et al.[46] along with their operational conditions.

Experimental methods are typically not cost-effective and time-consuming. The emergence of artificial intelligence attracted several researchers since this approach is capable of dealing with prior challenges in gas viscosity determination.[47,48] Abooali and Khamehchi[49] developed a method that was a function of pseudo-reduced temperature, pseudo-reduced pressure, apparent molecular weight, and gas density to predict natural gas dynamics viscosity by operating the genetic program on a database including 1938 data points. Deumah et al.[50] examined the efficiency of four different models, namely, multi-linear regression (MLR), decision tree (DT), random forest (RF), and K-nearest neighbors (KNN), to estimate the gas viscosity of a specific gas field. They found that the best accuracy belonged to the DT model by root mean square error (RMSE) of 0.000832 in pressure and temperature ranges of 14.7−3500 Psia and 70−221 °F, respectively. Baniasadi and Khamehchi[25] utilized artificial neural networks (ANNs) on 2083 sets of data of hydrocarbon gas compositions including methane and heavier components with varying mole fractions to predict viscosity using reduced temperature, reduced pressure, and gas density as the main input parameters. AlQuraishi and Shokir[51] implemented generalized regression neural networks (GRNNs) on 4445 experimental measurements containing pure gases and gas mixtures to develop a prediction model for viscosity by an AAPRE of 3.65%. In a recent study, Naghizadeh et al.[52] developed a predictive model to estimate CO_2−N_2 viscosities using multilayer perceptron (MLP), boosted regression tree (BRT) coupled with evolutionary algorithms, cascade forward neural network (CFNN), and GRNN smart paradigms. The results of their work yielded RMSE and R^2 values of 3.95 and 0.9975, respectively, for the BRT network coupled with an artificial bee colony (ABC) optimizer in the testing data set.

The aim of the present study is to provide intelligent approaches for accurately predicting the viscosity of a CO_2−N_2 gas mixture. To achieve this, an extensive data bank including 3036 data points on wide ranges of pressure and temperature is employed. The gaseous mixture viscosity is determined as a function of pressure, temperature, and CO_2 mole fraction by four developed models, namely, extra tree (ET), random forest (RF), gradient boosting (GBoost), and light gradient boosting machine (LightGBM). The robustness of the developed models is evaluated by graphical and statistical assessments. In addition, the models' performance is examined under physical behavior through trend analyses. To validate the dataset and certify the applicability domain of the models, outlier detection utilizing the leverage approach is performed. Figure 1 shows an overview of the present work step-by-step.

2. DEVELOPMENT OF INTELLIGENT MODELS

2.1. Data Preparation. In order to tune the smart techniques, a databank containing 3036 data points was utilized, with input parameters including pressure, temperature, and mole fraction of CO_2 from previous studies.[27,33,53−78] This

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

Table 2. Statistical Description of the Databank Utilized in This Study

parameter	mean	standard deviation	kurtosis	skewness	minimum	maximum	status
temperature (K)	296.59	104.02	3.32	0.22	66.50	973.15	input
pressure (MPa)	19.55	47.36	33.64	5.36	0.0013	453.20	input
CO_2 mole fraction	0.46	0.48	−1.92	0.16	0	1	input
viscosity (μPa·s)	61.07	80.99	6.85	2.57	6.68	514.70	output

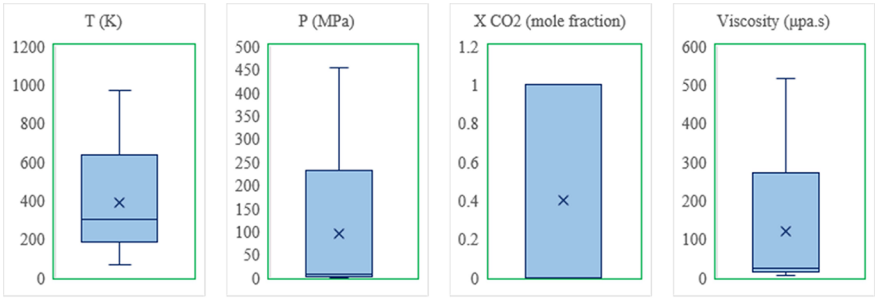


Figure 2. Box charts of the inputs and target values.

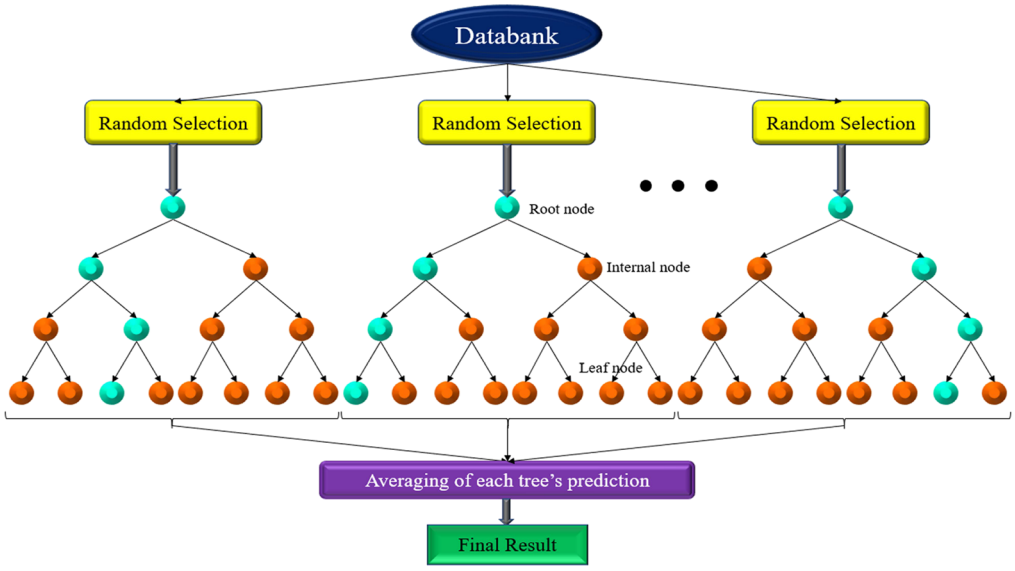


Figure 3. Schematic pattern of an ET model.

data bank was also used by Naghizadeh et al.[52] Table 2 summarizes the statistical properties of the dataset. Throughout the development process, the entire dataset was divided randomly into two subsets, namely, train and test parts carrying 80 and 20% of all data points, respectively. The major part was used to train models, while 20% was used to evaluate the model's efficacy and reliability. Furthermore, to allocate the chance of appearing in the training and validation to each observation from the databank, K-fold cross-validation was applied to the training dataset. In order to prevent overfitting during the models' training process, K-fold cross-validation was used. As a consequence, a K-fold of 10 gave the optimum result according to the size of the viscosity databank for the models. It means that the training dataset is randomly divided into tenfolds and then fits the model by applying K-1 (9) folds and validating the model using the remained fold. After that, the testing set is used to ensure more reliability and accuracy of the developed models.

The box plots of the inputs and output parameters are shown in Figure 2. This figure is a trustable preprocessing approach to verify the validation of the used data points based on five statistical features, including minimum, maximum, median (the middle value of the database), Q_1 (the median of the lower half of the database), and the third quarter or Q_3 (the median of the upper half of the database). The box is sketched from values Q_1 to Q_3 with a horizontal line depicted in the middle to demonstrate the median value. Furthermore, the lowest point is the minimum and the top point shows the maximum of the databank. According to this figure, if no data point is located in the range of more than 1.5 times the box length, it can be concluded that the databank follows a normal distribution and there is no outlier data point. As a result, Figure 2 illustrates that all data points used in this work are statistically valid and suitable.

2.2. Intelligent Techniques. *2.2.1. Extra Tree (ET).* The ET is a tree-based ensemble learning technique that is widely applicable in machine learning tasks such as regression and classification.[79,80] The ET algorithm combines a number of decision trees and applies the averaging approach on each decision tree's prediction value.[81] Each ET model utilizes all

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

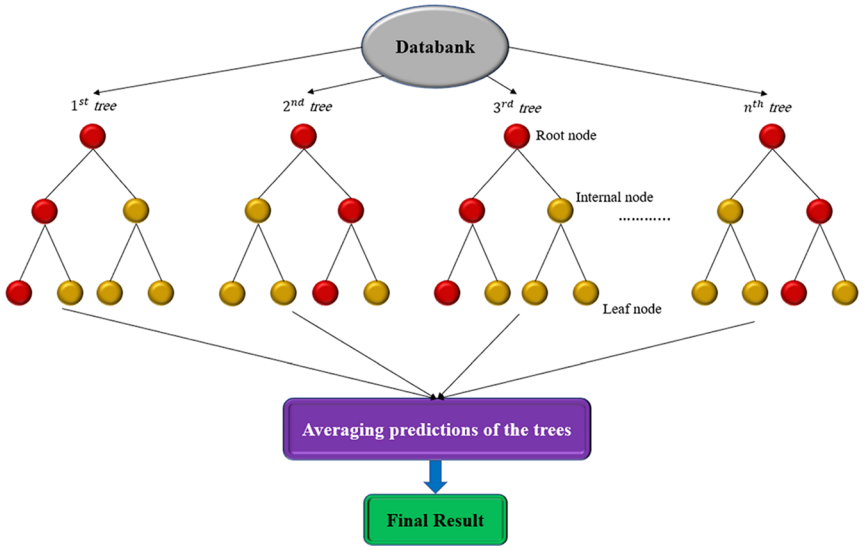


Figure 4. Schematic image of an RF model.

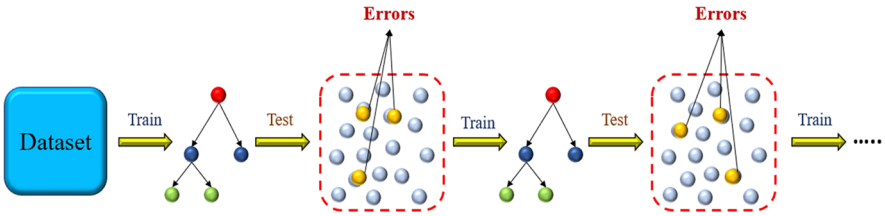


Figure 5. Schematic structure of a GBoost algorithm.

the cutting points and divides nodes at these points, haphazardly. Employing the whole learning data points to grow the trees in order to minimize the bias values is a key feature of each ET algorithm.[82] Two important parameters of the ET algorithm include N_{min} which denotes the minimum sample size needed to separate the neurons and K which presents the number of haphazard splits picked up in each neuron. The K value determines the training procedure in every tree's structure. Reducing the variance and controlling the overfitting of the model are two significant benefits of an ET model in comparison to a single decision tree.[83] The maximum depth (the longest path of nodes from the root node to the last leaf node) and an optimum number of trees are needed to have a trustworthy efficiency in the ET learning process.[84] Figure 3 illustrates a schematic pattern of an ET model.

2.2.2. Random Forest (RF). The RF, as depicted in Figure 4, is a tree-based regression[81] and classification[85] technique that aggregates a great number of decision trees. This method was introduced by Breiman in 2001 utilizing ensemble trees to predict the target variable which is calculated as the average of the predictions of the individual regression trees in the ensemble.[86] Also, the part of the training data that is not taken by the bootstrap sampling to build the tree is defined as the out-of-bag (OOB) sample which is used for incorporating a validation step within the fitting procedure.[86−88] These OOB errors are the estimation errors when the tuned RF network is employed in the OOB samples.[88] RF is based on the randomness of various kinds of decision trees generated from different data subpools. This tuning procedure can extremely reduce the model's variance, control overfitting, and improve the model's efficiency.[89] The number of trees and features (predictor) are two crucial variables for tuning the RF model.[90] After considering the optimum number of trees and assigning a bootstrap sample from the training subset to each tree, the model draws features from the training subset with f haphazardly selected features for the split point in each neuron. Then, the split point and the best variable from the predictor are separated, and every neuron is divided into two subneurons. Lately, the new forecast value is estimated by averaging the predicted values of every single tree.[91] The most significant difference between RF and ET algorithms is that RF selects the optimum split for growing in the tree's path, while in the ET algorithm, splits are chosen randomly.[92−94]

2.2.3. Gradient Boosting (GBoost). The GBoost algorithm combines weak learners, i.e., learners better than random, into powerful learners in an iterative procedure.[95] As shown in Figure 5, GBoost is an ensemble tree-based technique which is widely applied in regression and classification tasks.[96,97] In this method, the training data are strategically resampled to supply the most beneficial information for every consecutive model.[98] GBoost builds the answer and improves the overfitting issue by reducing the loss functions.[99] For this goal, it is suggested to select a function $h(x, \theta_t)$ to be the most parallel to the negative gradient $(g_t(x_i))_{i=1}^{N}$. By choosing an iterative model, we can overcome the difficulty of the parameters' prediction. The function $g_t(x)$ for each observed data is defined as follows:

$$g_t(x) = E_y\left[\frac{\partial \Psi\,(y, f(x))}{\partial f(x)}\middle| x\right]_{f(x)=\hat{f}^{t-1}(x)} \quad (1)$$

For allowing the replacement of a difficult tuning problem, one can easily select the new function increment to be the

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

most matched with $-g_t(x)$ using classic least-squares optimization as below:[99]

$$(\rho_t, \theta_t) = \arg\min_{\rho,\theta} \sum_{i=1}^{N} [-g_t(x_i) + \rho h(x_i, \theta)]^2 \quad (2)$$

The following steps represent a general tuning path of the GBoost model:

(a) initializing the $\hat{f}_0$ as a constant;
(b) calculate the negative gradient of $-g_t(x)$;
(c) conform a new base-learner function $h(x, \theta_t)$;
(d) recognize the optimum gradient descent step-size ρ_t as below

$$\rho_t = \arg\min_{\rho} \sum_{i=1}^{N} \Psi [y_i, \hat{f}_{t-1}(x_i) + \rho h(x_i, \theta_t)] \quad (3)$$

(e) Update the model prediction:

$$\hat{f}_t \leftarrow \hat{f}_{t-1} + \rho_t h(x, \theta_t) \quad (4)$$

In this method, the base-learner phase is just one node and the loss function is the standard squared error. By proceeding with the training of the model, the best structure is obtained.[96,99]

2.2.4. Light Gradient Boosting Machine (LightGBM). LightGBM, as shown in Figure 6, is a type of Gradient

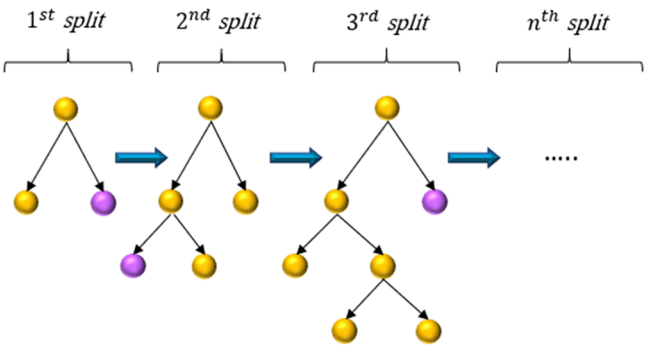


Figure 6. Simple structure of the LightGBM.

Boosting algorithm which is based on the decision tree training approach.[100] This algorithm is a powerful framework for solving various machine-learning problems.[101] Applying less memory storage is a notable advantage of the LightGBM in comparison to other machine learning approaches.[102] The LightGBM technique contains two creative methods, namely, exclusive features bundling (EFB), which is proposed to handle very large data features without overfitting concerns, and gradient-based one-side sampling (GOSS), which is regarded to filter samples to detect split values. Thus, samples with lower gradient values are satisfactorily trained and have less training errors. The vital parameters of LightGBM have the ability to manage numerous data, high speed, and higher precision in forecasts.[102] The following equation is described the training subset of the LightGBM algorithm:[103]

$$X = \{(x_j, y_j)\}_{j=1}^{N} \quad (5)$$

Then, $\hat{f}_{(x)}$ will predict by minimizing the loss function L:

$$L(y, f(x)): \hat{f}(x) = \arg\min E_{y,x} \cdot L(y, f(x)) \quad (6)$$

Finally, the training step of each individual tree can be described as follows:[103]

$$W_{q(x)}, \quad q \in \{1, 2, 3, ..., N\} \quad (7)$$

In the above equation, N expresses the leaf number in a tree, q denotes used decision rules in a single tree, and W describes the weight term of every leaf node.[103] Applying Newton's law for minimizing the objective function, the training final result of each step is tuned as below:

$$G_t \cong \sum_{i=1}^{N} L[y_i, F_{t-1}(x_i) + f_t(x_i)] \quad (8)$$

3. RESULTS AND DISCUSSION

3.1. Models' Development. This research presents four advanced and robust intelligent models (ET, RF, GBoost, and LightGBM) to predict the viscosity of the CO_2−N_2 gas mixture using pressure, temperature, and mole fraction of CO_2 as input parameters. For this purpose, a widespread data bank consisting of 3036 data points was gathered. Python programming language applying *pandas* and *numpy* libraries were used to train the intelligent paradigms. To optimize ET, RF, and GBoost models, wide ranges of hyperparameters, such as the maximum depth of the trees, the minimum number of data in each leaf (min sample leaf), the minimum number of data required to split an internal node (min sample split), the number of estimators (the number of trees), and learning rate, especially for GBoost (the step size which means each weight in all trees will be multiplied by this value), were tested to get the optimum structure of the models. Accordingly, extensive ranges including maximum depths from 2 to 100, min sample leaf from 1 to 5, min sample split from 1 to 6, and the number of trees from 10 to 1000 were tested. Furthermore, the mean square error (MSE) was regarded as the lost function during the tuning of the models. As a result, the best structures of the prementioned models were constructed by applying the hyperparameter values reported in Table 3.

Table 3. Hyperparameter Optimal Values of the ET, RF, and GBoost Algorithms

model	maximum depth	min sample leaf	min sample split	*n*_estimators	learning rate
ET	23	1	2	80	
RF	37	1	2	120	
GBoost	7	1	2	66	0.097

In addition, for developing the LightGBM technique, the optimum structure was earned by the maximum depth of 8, a learning rate of 0.4, and a number of leaves (the maximum number of leaves per tree) of 15.

3.2. Statistical Evaluation. Statistical analyses were utilized to assess the performance of developed models. Statistical evaluation indices include APRE, RMSE, AAPRE, standard deviation (SD), and correlation coefficient (R^2). Eqs 9−13 provide the mathematical formulas of the statistical criteria, where n is defined as the number of data, and y_i^{exp} and y_i^{cal} represent the experimental and calculated gas mixture viscosity by the proposed model, respectively. Also, $\bar{y}$ denotes the average value of actual data points.

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

Table 4. Statistical Error Analysis of the Developed Models

models	status	APRE (%)	AAPRE (%)	RMSE (μPa·s)	SD	R^2
ET	train	−0.80	1.70	1.994	0.110	0.9994
	test	−1.70	3.03	3.328	0.141	0.9981
	all	−0.98	1.97	2.323	0.117	0.9992
RF	train	−0.57	1.36	1.961	0.098	0.9994
	test	−0.82	2.47	3.044	0.064	0.9987
	all	−0.62	1.58	2.221	0.092	0.9993
GBoost	train	−1.06	3.28	2.094	0.059	0.9994
	test	−1.28	4.40	3.873	0.080	0.9974
	all	−1.11	3.51	2.551	0.064	0.9990
LightGBM	train	−0.33	2.70	2.437	0.050	0hat.9991
	test	0.03	3.95	4.210	0.086	0.9971
	all	−0.25	2.95	2.881	0.059	0.9987

$$\text{APRE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left(\frac{y_i^{\text{exp}} - y_i^{\text{cal}}}{y_i^{\text{exp}}}\right) \times 100\right) \tag{9}$$

$$\text{AAPRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{y_i^{\text{exp}} - y_i^{\text{cal}}}{y_i^{\text{exp}}}\right) \times 100\right| \tag{10}$$

$$\text{RMSE} = \left(\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{exp}} - y_i^{\text{cal}})^2\right)^{0.5} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{\text{exp}} - y_i^{\text{cal}})^2}{\sum_{i=1}^{n}(y_i^{\text{exp}} - \bar{y})^2} \tag{12}$$

$$\text{SD} = \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i^{\text{exp}} - y_i^{\text{cal}}}{y_i^{\text{exp}}}\right)^2\right)^{0.5} \tag{13}$$

Table 4 provides a summary of the error analysis of the developed models. As shown, all models had good agreement with experimental measurements. By AAPRE of 1.58%, the RF was the most accurate model in all error indices. ET equally displayed the next high precision of prediction. Furthermore, GBoost and LightGBM were placed in the following positions by AAPRE values of 3.51 and 2.95%, respectively. Besides, the comparison between the results of the proposed RF model and previous correlations in the literature is presented in Table 5.

Table 5. Comparison between the Proposed RF Model in This Research and Existing Correlations

study	Chen and Ruth	Sanjari et al.	Standing	Heidaryan et al.	this work (RF)
RMSE	23.93	54.25	41.01	49.84	3.04

As can be seen in Table 5, it is obvious that the RF model could outperform all prior correlations existing in the literature.

3.3. Graphical Evaluation. Graphical evaluation is one of the visual approaches used to assess the models. Different types of plots were introduced for this investigation. The cross plot is one graphical criterion that compares calculated values versus experimental data. A tight distribution around the $X = Y$ straight line indicates that the model reflects a higher degree of accuracy and there is a better agreement between the estimated and measured values. Figure 7 depicts the training and testing steps of all obtained models. As illustrated in the figure, while all approaches demonstrated acceptable compactness around a unit slope line, the RF approach displayed a concentrated area of points around this line in the entire viscosity range. It validated the method's high accuracy of prediction.

The error distribution plots of the developed models are shown in Figure 8. The error distribution diagrams illustrate the relative error percentage for each data point. More concentration of data points near the zero line indicates a more accurate prediction model. Consequently, as shown in the figure, although all models deviated from measured points at low viscosity values, a reliable prediction was offered for the rest of the range. In three models of ET, RF, and GBoost, a more compact distribution was reported than the LightGBM approach.

The group error is another effective visual descriptor that illustrates the average of an error indicator in a separate range of calculated values. The viscosity range was split into five equal 100 μPa·s ranges and the AAPRE for each category was plotted for four models, as shown in Figure 9. This diagram offers the opportunity to compare the effectiveness of each model in different ranges with itself and other models and provides a perspective for obtaining the most accurate method in each range. According to Figure 9, all applied models could estimate viscosity with an absolute percent relative error of less than 4.1% in all ranges. The RF model exhibited a relatively consistent accurate performance in all ranges with an acceptable absolute error and provided the most accurate prediction in viscosity less than 100 μPa·s, while in the other three models, this range was the least accurately predicted region. In the viscosity range between 100 and 200 μPa·s, the performance of all models was close, and gas mixture viscosity was calculated with an AAPRE of less than 1.18%. Increasing viscosity enhanced the performance of the GBoost method, which provided the highest precise prediction at viscosities greater than 200 μPa·s.

The cumulative frequency graph demonstrates what percentage of data already has what maximum absolute percent error. In this technique, the cumulative frequency of data is plotted versus the absolute percent error. Each model located higher in this graph is more accurate and offers a more credible forecast of output and shows that a greater proportion of data has the same error value in comparison to other models. The cumulative frequency plot of intelligent models is depicted in Figure 10. As the figure reflects for the RF model, more than 94.96% of data had an absolute percent error of less than 5%, and in the ET model, 92.98% of data were predicted with an error less than 5.2%. While the performance of

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

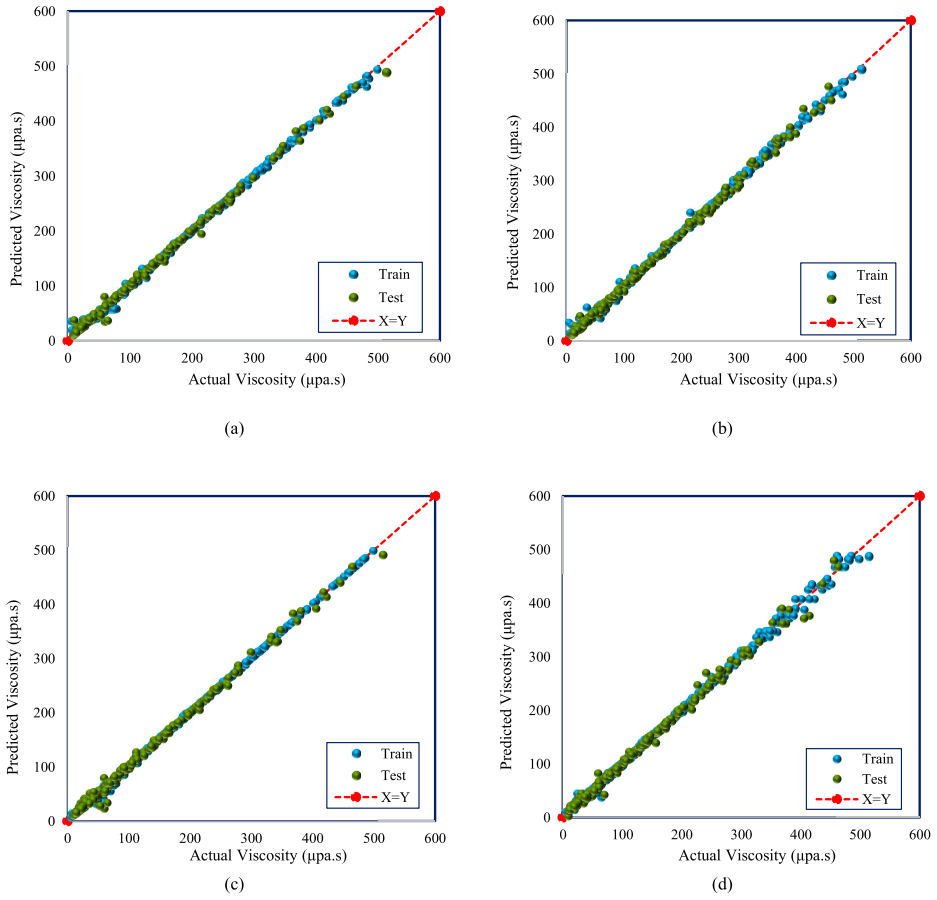


Figure 7. Cross plots of the developed models in this study; (a) ET, (b) RF, (c) GBoost, and (d) LightGBM.

LightGBM and GBoost models was stranded in lower positions compared to that of ET and RF, they also could estimate 80% of points by errors lower than 4.14 and 5.3%, respectively.

3.4. Trend Analysis of the Developed RF Model. In each model's trend chart, the pattern of changes in the predicted and measured values due to increasing one variable parameter can be evaluated by keeping all inputs constant except one. Trend analysis illustrates the impact of increasing a parameter on the error value in addition to indicating that the measured values reflect a logical trend. The effect of temperature and pressure on the CO_2−N_2 mixture was studied in the RF model. Figure 11a illustrates the effect of pressure on this gas mixture at a constant CO_2 mole fraction and at a temperature of 74 K. Increasing the pressure increases the attraction force between gas molecules, and increasing the kinetic energy leads to an increase in gas viscosity. The temperature effect under different pressure ranges is depicted in Figure 11b,c, which illustrates that at high pressure, gas fluid properties become closer to a liquid, and an increase in temperature accelerates molecular movement and decreases viscosity by lessening internal friction. Rising temperature impedes fluid flow at low pressure, which promotes intermolecular interaction and increases viscosity. The model accurately predicted viscosity at varying pressures and temperatures by matching the patterns of physical phenomena.

3.5. Variable Impact Analysis. The relevancy factor is one of the methodologies of sensitivity analysis. The relevancy factor determines the impact of each input on the output based on the Pearson technique, which is calculated by the r value. The mathematical calculation of r is given in eq 14. This parameter is limited between −1 and 1. A positive relevancy factor for one input indicates that the dependent and independent variables have a parallel trend. This means that increasing inputs leads to a growing output trend. However, for $r < 0$, input has a reverse impact on output; in other words, an increasing trend in target value comes from decreasing independent parameters. Furthermore, how much the absolute value of r is close to 1, there is a stronger relationship between the two parameters.[104]

$$r(I_k, y)\frac{\sum_{i=1}^{n}(I_i^k - I_{ave}^k)(y_i - y_{ave})}{\sqrt{\sum_{i=1}^{n}(I_i^k - I_{ave}^k)^2 \sum_{i=1}^{n}(y_i - y_{ave})^2}} \quad (14)$$

In the above formula, I_i^k and I_{ave}^k denote the average value and the ith value of the kth input, respectively (k could be temperature, mole fraction of CO_2 or pressure). Besides, y_i shows the ith value of the predicted viscosity, and y_{ave} illustrates the average of the predicted viscosity. The impact of the input variables of the developed models on the viscosity of CO_2−N_2 mixtures is illustrated in Figure 12. As can be seen, the relevancy factor of each input parameter maintained the same across all models. Pressure by the relevancy factor of more than 0.73 showed a positive dominance effect. While pressure and CO_2 mole fraction were in a positive relationship by the target value, the temperature had a negative impact on the viscosity by −0.37 of the r value. The negative value for temperature r value means that most of the data were at high pressure; thus, increasing temperature decreases viscosity.

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

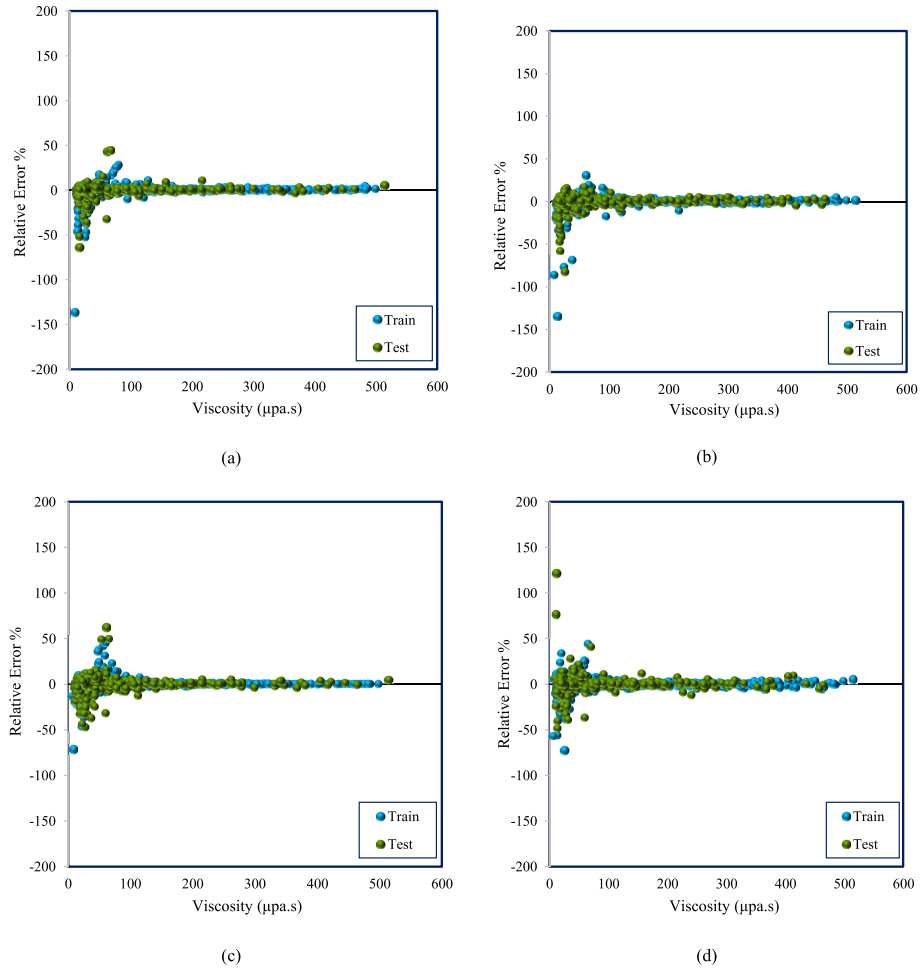


Figure 8. Error distribution plots for (a) ET, (b) RF, (c) GBoost, and (d) LightGBM.

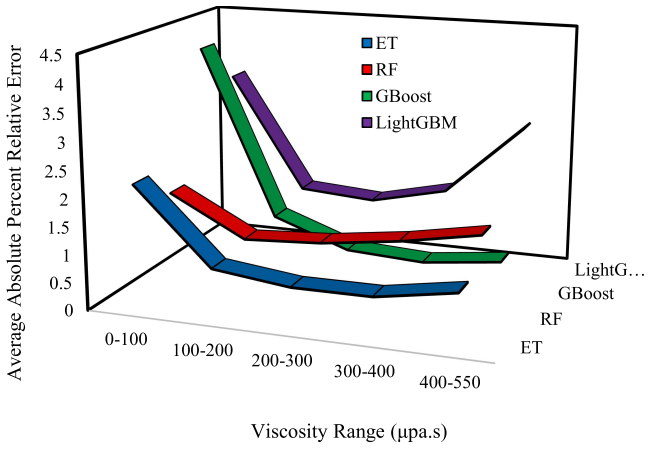


Figure 9. Group error for five viscosity ranges for four intelligent models developed in this research.

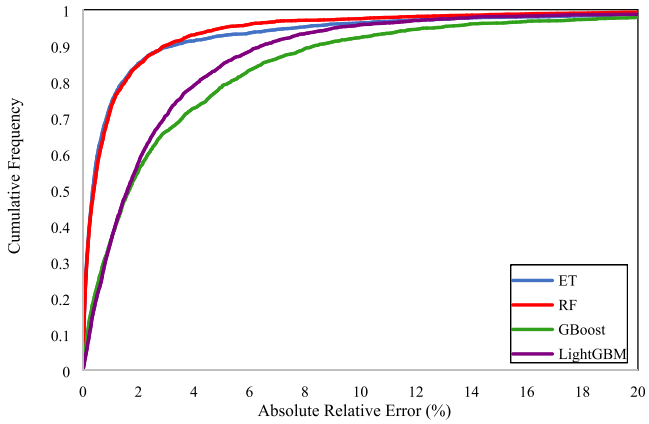


Figure 10. Cumulative frequency plot for the developed predictive models.

3.6. Outlier Detection of the proposed RF model. The leverage approach is one of the well-known techniques for evaluating a model's applicability area by identifying outlier points by finding the leverage values of each compound.[105,106] In this approach, to assess the reliability of the dataset and model's applicability, analytical and visual tools are utilized.[107] William's plot identifies outlier data by locating them far from the bulk of the data.[108] In this plot, the standardized residuals, which measure the variance of predicted values from experimental data, are sketched versus the diagonal components of the hat matrix, referred to as hat values. The formulas to calculate the hat matrix and standardized residuals are given in eqs 15−17, where X is a two-dimensional $(N \times k)$ matrix, N and k represent the number of data points and input parameter variables, respectively, and T symbolizes the transpose matrix. In standardized residual calculation, e_i stands for the difference between the i-th model's predicted and experimental value,

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

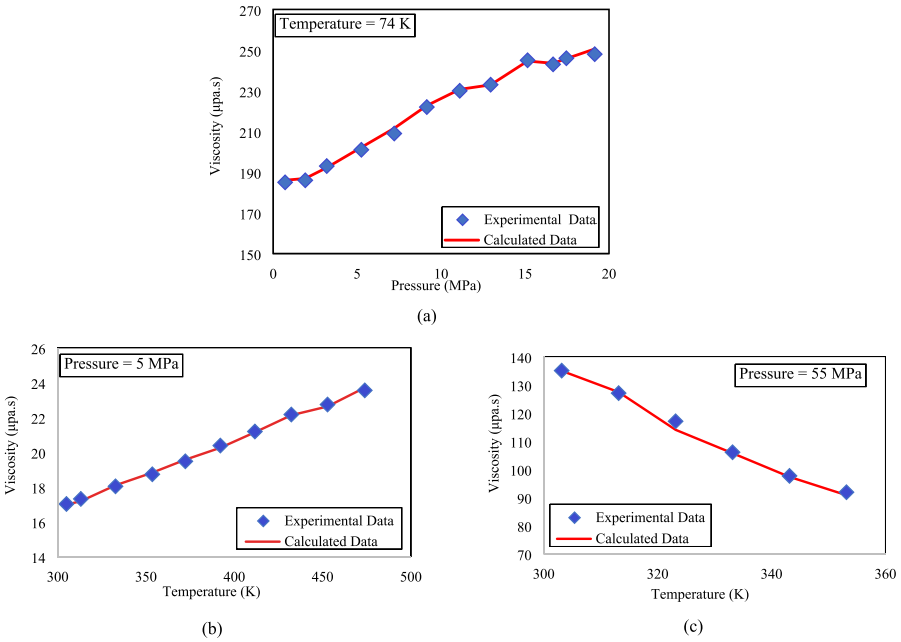


Figure 11. Effect of the pressure and temperature on the gaseous mixture's viscosity of the RF model.

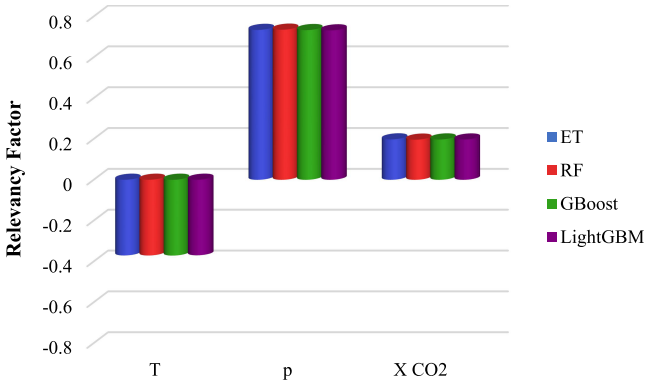


Figure 12. Relative importance of temperature, pressure, and CO_2 mole fraction on the developed models.

RMSE shows the root mean square error, and H_{ii} indicates the hat value of the i-th data point.[109]

$$H = X(X^{T}X)^{-1}X^{T},\ X = N \times k \tag{15}$$

$$\mathrm{SR} = \frac{e_i}{(\mathrm{MSE}(1 - H_{ii}))^{0.5}} \tag{16}$$

$$H^{*} = \frac{3 \times (k + 1)}{n} \tag{17}$$

In the visual representation of the William's plot, the data are considered valid if it is placed in the zone that is limited between $0 \leq H \leq H^{*}$ and $-3 \leq \mathrm{SR} \leq 3$. In this zone, points are considered valid data and are included in the model's applicability domain. Outlier data points are placed below the leverage limit (H^{*}), but they are not in the range of $-3 \leq \mathrm{SR} \leq 3$, recognized as suspected data. The data show a hat value higher than H^{*} based on their SR divided into good high leverage and bad high leverage. Good high leverage refers to the zone in which the standardized residual is located in the range of $-3 \leq \mathrm{SR} \leq 3$, implying that the measurements of these data points are valid, but they are outside the applicability domain of the model. The data points with SR > 3 or SR < − 3 are bad high-leverage points, which are outside of the model's applicability domain and are not predicted well.[109]

William's plot of the RF model for predicting viscosity is shown in Figure 13. The leverage limit value is obtained as

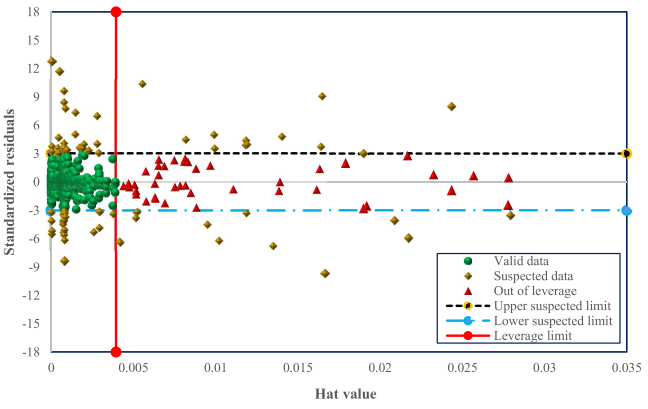


Figure 13. William's plot for outlier detection of the proposed RF model.

0.0039. As demonstrated in the plot, a major portion (nearly of 96%) of the databank was located in the valid data region. Therefore, the leverage approach validated the dataset and certified the applicability domain of the applied model.

4. CONCLUSIONS

In the present research, the viscosity of CO_2−N_2 mixtures was modeled utilizing smart modeling approaches considering pressure, temperature, and mole fraction of CO_2 as inputs. For this goal, a widespread databank consisting of 3036 data points was collected from different literature sources. The main conclusions of the present study are as follows:

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

1. The results showed the superiority of the RF model in comparison to other developed models with an AAPRE of 1.58% and a correlation coefficient of 0.9993.
2. All the developed models in this study demonstrated a high precision and satisfactory agreement between experimental and predicted viscosity values.
3. Cross-validation analysis showed that all developed models in this research could outperform the prior related work, without overfitting.
4. The developed smart schemes can be ordered in terms of their AAPRE values as below:

RF < ET < LightGBM < GBoost

5. The trend analysis of gas mixture viscosity change curves indicated that the RF model could forecast the real behavior of the CO_2−N_2 viscosity variation accurately.
6. Lately, outlier detection using the Leverage method showed that nearly 96% of the data points were placed in the valid area, and the used databank was valid and reliable.

AUTHOR INFORMATION

Corresponding Authors

Haimin Zheng − *Engn & Design Dept, Proc Sect, CNOOC Research Institute Co., Beijing 100027, P.R. China*; Email: zhenghm4@cnooc.com.cn

Abdolhossein Hemmati-Sarapardeh − *Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman 1234567891, Iran; State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China;* orcid.org/0000-0002-5889-150X; Email: hemmati@uk.ac.ir, aut.hemmati@gmail.com

Authors

Atena Mahmoudzadeh − *Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran 1234567812, Iran*

Behnam Amiri-Ramsheh − *Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman 1234567891, Iran*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c00228

Author Contributions

[⊥]H.Z. and A.M. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge the respected editor and anonymous reviewers for taking their time to study this paper and provide us with their constructive comments.

REFERENCES

(1) Wang, X.; Gu, Y. Oil recovery and permeability reduction of a tight sandstone reservoir in immiscible and miscible CO2 flooding processes. *Ind. Eng. Chem. Res.* **2011**, *50*, 2388−2399.

(2) Zhou, D.; Yang, D. Scaling criteria for waterflooding and immiscible CO2 flooding in heavy oil reservoirs. *J. Energy Resour. Technol.* **2017**, *139*, No. 022909.

(3) Massarweh, O.; Abushaikha, A. S. A review of recent developments in CO2 mobility control in enhanced oil recovery. *Petroleum* **2022**, *8*, 291−317.

(4) Nematzadeh, M.; Khanamiri, H.; Aghajani, M.; Kharrat, R.; Gandomkar, A.; Motealleh, M.; Ghazanfari, M. An experimental study of secondary WAG injection in a low-temperature carbonate reservoir in different miscibility conditions. *Pet. Sci. Technol.* **2012**, *30*, 1359−1368.

(5) Carpenter, C. Experimental Program investigates miscible CO2 WAG injection in carbonate reservoirs. *J. Pet. Technol.* **2019**, *71*, 47−49.

(6) Dai, Z.; Middleton, R.; Viswanathan, H.; Fessenden-Rahn, J.; Bauman, J.; Pawar, R.; Lee, S.-Y.; McPherson, B. An integrated framework for optimizing CO2 sequestration and enhanced oil recovery. *Environ. Sci. Technol. Lett.* **2014**, *1*, 49−54.

(7) Casteel, J.; Djabbarah, N. Sweep improvement in CO2 flooding by use of foaming agents. *SPE Reservoir Eng.* **1988**, *3*, 1186−1192.

(8) Yan, W.; Miller, C. A.; Hirasaki, G. J. Foam sweep in fractures for enhanced oil recovery. *Colloids Surf., A* **2006**, *282*−*283*, 348−359.

(9) Heller, J.; Dandge, D.; Card, R.; Donaruma, L. Direct thickeners for mobility control of CO2 floods. *Soc. Pet. Eng. J.* **1985**, *25*, 679−686.

(10) Bae, J.; Irani, C. A laboratory investigation of viscosified CO2 process. *SPE Adv. Technol. Ser.* **1993**, *1*, 166−171.

(11) Xiong, Y.; Kiran, E. Miscibility, density and viscosity of poly (dimethylsiloxane) in supercritical carbon dioxide. *Polymer* **1995**, *36*, 4817−4826.

(12) Jafari, S.; Khezrnejad, A.; Shahrokhi, O.; Ghazanfari, M. H.; Vossoughi, M. Experimental investigation of heavy oil recovery by continuous/WAG injection of CO2 saturated with silica nanoparticles. *Int. J. Oil, Gas Coal Technol.* **2015**, *9*, 169−179.

(13) Sun, Q.; Zhang, N.; Li, Z.; Wang, Y. Nanoparticle-Stabilized Foam for Mobility Control in Enhanced Oil Recovery. *Energy Technol.* **2016**, *4*, 1084−1096.

(14) Yekeen, N.; Manan, M. A.; Idris, A. K.; Padmanabhan, E.; Junin, R.; Samin, A. M.; Gbadamosi, A. O.; Oguamah, I. A comprehensive review of experimental studies of nanoparticles-stabilized foam for enhanced oil recovery. *J. Pet. Sci. Eng.* **2018**, *164*, 43−74.

(15) Majidaie, S.; Onur, M.; Tan, I. M. An experimental and numerical study of chemically enhanced water alternating gas injection. *Pet. Sci.* **2015**, *12*, 470−482.

(16) Schramm, L. L.; Novosad, J. J. Micro-visualization of foam interactions with a crude oil. *Colloids Surf.* **1990**, *46*, 21−43.

(17) Sun, W.; Sun, B.; Li, Y.; Huang, X.; Fan, H.; Zhao, X.; Sun, H.; Sun, W. Thickening supercritical CO2 with π-stacked Co-polymers: molecular insights into the role of intermolecular interaction. *Polymer* **2018**, *10*, 268.

(18) Padamsey, R.; Railton, J. CO2 capture and use for EOR in Western Canada 4. Economic results and conclusions. *Energy Convers. Manage.* **1993**, *34*, 1165−1175.

(19) Srivastava, R. K.; Huang, S. S.; Dong, M. Comparative effectiveness of CO2, produced gas, and flue gas for enhanced heavy-oil recovery. *SPE Reservoir Eval. Eng.* **1999**, *2*, 238−247.

(20) Dong, M.; Huang, S. Flue gas injection for heavy oil recovery. *J. Can. Pet. Technol.* **2002**, *41*, No. PETSOC-02-09-04.

(21) Wu, Z.; Liu, H.; Wang, X. 3D experimental investigation on enhanced oil recovery by flue gas coupled with steam in thick oil reservoirs. *Energy Fuels* **2018**, *32*, 279−286.

(22) Song, C.; Pan, W.; Srimat, S. T.; Zheng, J.; Li, Y.; Wang, Y.-H.; Xu, B.-Q.; Zhu, Q.-M. Tri-reforming of methane over Ni catalysts for CO2 conversion to syngas with desired H2/CO ratios using flue gas of power plants without CO2 separation. *Stud. Surf. Sci. Catal.* **2004**, *153*, 315−322.

(23) Johnson, H.; Schmidt, L.; Thrash, L. A Flue Gas Huff'n'Puff Process for Oil Recovery From Shallow Formations. In *SPE/DOE Enhanced Oil Recovery Symposium*; OnePetro, 1990.

(24) Zhang, Y.; Sayegh, S.; Huang, S.; Dong, M. Laboratory investigation of enhanced light-oil recovery by CO/Flue gas huff-n-

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

puff process. In *Canadian International Petroleum Conference*; OnePetro, 2004.
(25) Baniasadi, H.; Khamehchi, E. Development of a new model to predict gas viscosity using artificial neural networks. *Int. J. Pet. Geosci. Eng.* **2014**, *2*, 195−207.
(26) Kestin, J.; Leidenfrost, W.; Liu, C. On relative measurements of the viscosity of gases by the oscillating-disk method. *Z. Angew. Math. Phys.* **1959**, *10*, 558−564.
(27) Kestin, J.; Kobayashi, Y.; Wood, R. The viscosity of four binary, gaseous mixtures at 20 and 30 C. *Physica* **1966**, *32*, 1065−1089.
(28) Kestin, J.; Ro, S. T. The viscosity of nine binary and two ternary mixtures of gases at low density. *Ber. Bunsenges. Phys. Chem.* **1974**, *78*, 20−24.
(29) Gururaja, G.; Tirunarayanan, M.; Ramachandran, A. Dynamic viscosity of gas mixtures. *J. Chem. Eng. Data* **1967**, *12*, 562−567.
(30) Golubev, I. I. A. F. *Viscosity of gases and gas mixtures: A handbook*; Israel Program for Scientific Translations, 1970; [available from the US Department].
(31) Khosravi, B.; Betken, B.; Jakobsen, J. P.; Løvseth, S. W.; Span, R. Viscosity measurements of CO2-rich; CO2+ N2 and CO2+ H2 mixtures in gas or supercritical phase at temperatures between 273 and 473 K and pressures up to 8.7 MPa. *Fluid Phase Equilib.* **2022**, No. 113519.
(32) Carr, N. L.; Kobayashi, R.; Burrows, D. B. Viscosity of hydrocarbon gases under pressure. *J. Pet. Technol.* **1954**, *6*, 47−55.
(33) Lohrenz, J.; Bray, B. G.; Clark, C. R. Calculating viscosities of reservoir fluids from their compositions. *J. Pet. Technol.* **1964**, *16*, 1171−1176.
(34) Whitson, C. H.; Brule, M. R. *Phase behavior*; Henry L. Doherty Memorial Fund of AIME. Society of Petroleum Engineers, 2000; vol *20*.
(35) Dean, D. E.; Stiel, L. I. The viscosity of nonpolar gas mixtures at moderate and high pressures. *AIChE J.* **1965**, *11*, 526−532.
(36) Standing, M. B. *Volumetric and phase behavior of oil field hydrocarbon systems*; Society of Petroleum Engineers of AIME, 1977.
(37) Vesovic, V.; Wakeham, W. The prediction of the viscosity of dense gas mixtures. *Int. J. Thermophys.* **1989**, *10*, 125−132.
(38) Chen, Z.; Ruth, D. On viscosity correlations of natural gas. In *Annual Technical Meeting*; OnePetro, 1993.
(39) Dranchuk, P.; Islam, M.; Bentsen, R. A mathematical representation of the Carr, Kobayashi and Burrows natural gas viscosity correlations. *J. Can. Pet. Technol.* **1986**, *25*, No. PETSOC-86-01-03.
(40) Jarrahian, A.; Heidaryan, E. A simple correlation to estimate natural gas viscosity. *J. Nat. Gas Sci. Eng.* **2014**, *20*, 50−57.
(41) Heidaryan, E.; Esmaeilzadeh, F.; Moghadasi, J. Natural gas viscosity estimation through corresponding states based models. *Fluid Phase Equilib.* **2013**, *354*, 80−88.
(42) Sanjari, E.; Nemati Lay, E.; Peymani, M.; Afshar, N. An efficient reliable method for calculation of natural gas viscosity. *Pet. Sci. Technol.* **2014**, *32*, 1300−1308.
(43) Yang, X.; Zhang, S.; Zhu, W. A new model for the accurate calculation of natural gas viscosity. *Nat. Gas Ind.* **2017**, *4*, 100−105.
(44) Yang, X.; Zhu, W. A theoretical model for the density and temperature dependent viscosity of hydrocarbon gases. *Pet. Sci. Technol.* **2016**, *34*, 765−770.
(45) Sanjari, E.; Lay, E. N.; Peymani, M. An accurate empirical correlation for predicting natural gas viscosity. *J. Nat. Gas Chem.* **2011**, *20*, 654−658.
(46) Heidaryan, E.; Moghadasi, J.; Salarabadi, A. A new and reliable model for predicting methane viscosity at high pressures and high temperatures. *J. Nat. Gas Chem.* **2010**, *19*, 552−556.
(47) Lay, E. N.; Peymani, M.; Sanjari, E. Prediction of Natural Gas Viscosity using Artificial Neural Network Approach. *Int. J. Chem. Mol. Eng.* **2012**, *6*, 577−583.
(48) Sanaei, A.; Yousefi, S.; Naseri, A.; Khishvand, M. A novel correlation for prediction of gas viscosity. *Energy Sources, Part A* **2015**, *37*, 1943−1953.
(49) Abooali, D.; Khamehchi, E. Estimation of dynamic viscosity of natural gas based on genetic programming methodology. *J. Nat. Gas Sci. Eng.* **2014**, *21*, 1025−1031.
(50) Deumah, S. S.; Yahya, W. A.; Al-khudafi, A. M.; Ba-Jaalah, K. S.; Al-Absi, W. T. Prediction of Gas Viscosity of Yemeni Gas Fields Using Machine Learning Techniques. In *SPE Symposium: Artificial Intelligence-Towards a Resilient and Efficient Energy Industry*; OnePetro, 2021.
(51) AlQuraishi, A. A.; Shokir, E. M. Artificial neural networks modeling for hydrocarbon gas viscosity and density estimation. *J. King Saud Univ.-Eng. Sci.* **2011**, *23*, 123−129.
(52) 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.
(53) Lee, A. L.; Gonzalez, M. H.; Eakin, B. E. The viscosity of natural gases. *J. Pet. Technol.* **1966**, *18*, 997−1000.
(54) Londono, F. E.; Archer, R. A.; Blasingame, T. A. Correlations for Hydrocarbon-Gas Viscosity and Gas Density—Validation and Correlation of Behavior Using a Large-Scale Database. *SPE Reservoir Eval. Eng.* **2005**, *8*, 561−572.
(55) Lucas, K. *Die druckabhängigkeit der viskosität von flüssigkeiten-eine einfache abschätzung*; 1981.
(56) Nazeri, M.; Chapoy, A.; Burgass, R.; Tohidi, B. Viscosity of CO2-rich mixtures from 243 K to 423 K at pressures up to 155 MPa: New experimental viscosity data and modelling. *J. Chem. Thermodyn.* **2018**, *118*, 100−114.
(57) Seibt, D.; Vogel, E.; Bich, E.; Buttig, D.; Hassel, E. Viscosity measurements on nitrogen. *J. Chem. Eng. Data* **2006**, *51*, 526−533.
(58) Seibt, D.; Herrmann, S.; Vogel, E.; Bich, E.; Hassel, E. Simultaneous measurements on helium and nitrogen with a newly designed viscometer− densimeter over a wide range of temperature and pressure. *J. Chem. Eng. Data* **2009**, *54*, 2626−2637.
(59) Schäfer, M.; Richter, M.; Span, R. Measurements of the viscosity of carbon dioxide at temperatures from (253.15 to 473.15) K with pressures up to 1.2 MPa. *J. Chem. Thermodyn.* **2015**, *89*, 7−15.
(60) Pensado, A.; Padua, A. A.; Comuñas, M.; Fernandez, J. Viscosity and density measurements for carbon dioxide+ pentaerythritol ester lubricant mixtures at low lubricant concentration. *J. Supercrit. Fluids* **2008**, *44*, 172−185.
(61) Evers, C.; Lösch, H.; Wagner, W. An absolute viscometer-densimeter and measurements of the viscosity of nitrogen, methane, helium, neon, argon, and krypton over a wide range of density and temperature. *Int. J. Thermophys.* **2002**, *23*, 1411−1439.
(62) Iwasaki, H.; Takahashi, M. Viscosity of carbon dioxide and ethane. *J. Chem. Phys.* **1981**, *74*, 1930−1943.
(63) Kestin, J.; Paykoc, E.; Sengers, J. On the density expansion for viscosity in gases. *Physica* **1971**, *54*, 1−19.
(64) Kestin, J.; Ro, S.; Wakeham, W. Viscosity of the binary gaseous mixtures He−Ne and Ne−N 2 in the Temperature Range 25−700 C. *J. Chem. Phys.* **1972**, *56*, 5837−5842.
(65) Kestin, J.; Khalifa, H.; Ro, S.; Wakeham, W. The viscosity and diffusion coefficients of eighteen binary gaseous systems. *Phys. A* **1977**, *88*, 242−260.
(66) Saunders, M. W. On the Measurement of the Absolute Viscosity of Nitrogen and Air Over the Temperature Range of 100 to 400° K and at Pressures of 1 to 150 Atmospheres. 1972.
(67) Bailey, B. The viscosity of carbon dioxide and acetylene at atmospheric pressure. *J. Phys. D: Appl. Phys.* **1970**, *3*, 550.
(68) Kestin, J.; Yata, J. Viscosity and diffusion coefficient of six binary mixtures. *J. Chem. Phys.* **1968**, *49*, 4780−4791.
(69) DiPippo, R.; Kestin, J.; Oguchi, K. Viscosity of three binary gaseous mixtures. *J. Chem. Phys.* **1967**, *46*, 4758−4764.
(70) Breetveld, J.; DiPippo, R.; Kestin, J. Viscosity and Binary Diffusion Coefficient of Neon—Carbon Dioxide Mixtures at 20° and 30° C. *J. Chem. Phys.* **1966**, *45*, 124−126.
(71) Goldman, K. Viscosity of nitrogen at low temperatures and high pressures. *Physica* **1963**, *29*, 499−516.

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875

(72) Ross, J. F.; Brown, G. M. Viscosities of gases at high pressures. *Ind. Eng. Chem.* **1957**, *49*, 2026−2033.
(73) Michels, A.; Botzen, A.; Schuurman, W. The viscosity of carbon dioxide between 0 C and 75 C and at pressures up to 2000 atmospheres. *Physica* **1957**, *23*, 95−102.
(74) Michels, A.; Gibson, R. The measurement of the viscosity of gases at high pressures—The viscosity of nitrogen to 1000 atms. *Proc. R. Soc. London, Ser. A* **1931**, *134*, 288−307.
(75) Haepp, H. J. Messung der Viskosität von Kohlendioxid und Propylen. *Wärme Stoffübertragung* **1976**, *9*, 281−290.
(76) Van Der Gulik, P. Viscosity of carbon dioxide in the liquid phase. *Phys. A* **1997**, *238*, 81−112.
(77) Humberg, K.; Richter, M.; Trusler, J. M.; Span, R. Measurement and modeling of the viscosity of (nitrogen + carbon dioxide) mixtures at temperatures from (253.15 to 473.15) K with pressures up to 2 MPa. *J. Chem. Thermodyn.* **2018**, *120*, 191−204.
(78) Hunter, I.; Marsh, G.; Matthews, G.; Smith, E. Argon+ carbon dioxide gaseous mixture viscosities and anisotropic pair potential energy functions. *Int. J. Thermophys.* **1993**, *14*, 819−833.
(79) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3−42.
(80) Hammed, M. M.; AlOmar, M. K.; Khaleel, F.; Al-Ansari, N. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. *Math. Probl. Eng.* **2021**, *2021*, No. 7001710.
(81) Amiri-Ramsheh, B.; Safaei-Farouji, M.; Larestani, A.; Zabihi, R.; Hemmati-Sarapardeh, A. Modeling of wax disappearance temperature (WDT) using soft computing approaches: Tree-based models and hybrid models. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109774.
(82) Kondori, J.; Miah, M. I.; Zendehboudi, S.; Khan, F.; Heagle, D. Hybrid connectionist models to assess recovery performance of low salinity water injection. *J. Pet. Sci. Eng.* **2021**, *197*, No. 107833.
(83) Okoro, E. E.; Obomanu, T.; Sanni, S. E.; Olatunji, D. I.; Igbinedion, P. Application of artificial intelligence in predicting the dynamics of bottom hole pressure for under-balanced drilling: extra tree compared with feed forward neural network model. *Petroleum* **2022**, *8*, 227−236.
(84) Larestani, A.; Hemmati-Sarapardeh, A.; Naseri, A. Experimental measurement and compositional modeling of bubble point pressure in crude oil systems: Soft computing approaches, correlations, and equations of state. *J. Pet. Sci. Eng.* **2022**, *212*, No. 110271.
(85) Cheraghi, Y.; Kord, S.; Mashayekhizadeh, V. Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities. *J. Pet. Sci. Eng.* **2021**, *205*, No. 108761.
(86) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(87) Ao, Y.; Li, H.; Zhu, L.; Ali, S.; Yang, Z. Identifying channel sand-body from multiple seismic attributes with an improved random forest algorithm. *J. Pet. Sci. Eng.* **2019**, *173*, 781−792.
(88) Attanasi, E. D.; Freeman, P. A.; Coburn, T. C. Well predictive performance of play-wide and Subarea Random Forest models for Bakken productivity. *J. Pet. Sci. Eng.* **2020**, *191*, No. 107150.
(89) Kim, S.; Lee, M.; Lee, K.; Ahn, T.; Lee, J. Data-driven estimation of three-phase saturation during gas hydrate depressurization using CT images. *J. Pet. Sci. Eng.* **2021**, *205*, No. 108916.
(90) Boulesteix, A. L.; Janitza, S.; Kruppa, J.; König, I. R. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery* **2012**, *2*, 493−507.
(91) Gholizadeh, M.; Jamei, M.; Ahmadianfar, I.; Pourrajab, R. Prediction of nanofluids viscosity using random forest (RF) approach. *Chemom. Intell. Lab. Syst.* **2020**, *201*, No. 104010.
(92) Shateri, M.; Sobhanigavgani, Z.; Alinasab, A.; Varamesh, A.; Hemmati-Sarapardeh, A.; Mosavi, A. S. Comparative analysis of machine learning models for nanofluids viscosity assessment. *Nanomaterials* **2020**, *10*, 1767.
(93) Abdi, J.; Hadavimoghaddam, F.; Hadipoor, M.; Hemmati-Sarapardeh, A. Modeling of CO2 adsorption capacity by porous metal organic frameworks using advanced decision tree-based models. *Sci. Rep.* **2021**, *11*, 24468.
(94) Talebkeikhah, M.; Amar, M. N.; Naseri, A.; Humand, M.; Hemmati-Sarapardeh, A.; Dabir, B.; Seghier, M. E. A. B. Experimental measurement and compositional modeling of crude oil viscosity at reservoir conditions. *J. Taiwan Inst. Chem. Eng.* **2020**, *109*, 35−50.
(95) Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937−1967.
(96) Larestani, A.; Mousavi, S. P.; Hadavimoghaddam, F.; Hemmati-Sarapardeh, A. Predicting formation damage of oil fields due to mineral scaling during water-flooding operations: Gradient boosting decision tree and cascade-forward back-propagation network. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109315.
(97) Velthoen, J.; Dombry, C.; Cai, J.-J.; Engelke, S. Gradient boosting for extreme quantile regression. 2021, *arXiv preprint arXiv:2103.00808*.
(98) Zhang, Y.; Haghani, A. A gradient boosting method to improve travel time prediction. *Transport. Res., Part C: Emerg. Technol.* **2015**, *58*, 308−324.
(99) Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 21.
(100) Gong, M.; Bai, Y.; Qin, J.; Wang, J.; Yang, P.; Wang, S. Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin. *J. Build. Eng.* **2020**, *27*, No. 100950.
(101) Csizmadia, G.; Liszkai-Peres, K.; Ferdinandy, B.; Miklósi, Á.; Konok, V. Human activity recognition of children with wearable devices using LightGBM machine learning. *Sci. Rep.* **2022**, *12*, 5472.
(102) Rai, M.; Mandoria, H. L. Network Intrusion Detection: A comparative study using state-of-the-art machine learning methods. In *2019 international conference on issues and challenges in intelligent computing techniques (ICICT)*; IEEE, 2019; pp 1−5.
(103) 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*, 17911.
(104) 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.
(105) Rousseeuw, P.; Leroy, A. *Robust regression and outlier detection*; Wiley Interscience: New York, 1987.
(106) Mohammadi, A. H.; Eslamimanesh, A.; Gharagheizi, F.; Richon, D. A novel method for evaluation of asphaltene precipitation titration data. *Chem. Eng. Sci.* **2012**, *78*, 181−185.
(107) Goodall, C. R. *Computation using the QR decomposition*; 1993.
(108) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694−701.
(109) Rousseeuw, P. J.; Van Zomeren, B. C. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* **1990**, *85*, 633−639.

https://doi.org/10.1021/acsomega.3c00228
ACS Omega 2023, 8, 13863−13875