Machine-learning-based prediction of oil recovery factor for experimental CO_2-Foam chemical EOR: Implications for carbon utilization projects

Hung Vo Thanh [a], Danial Sheini Dashtgoli [b,c], Hemeng Zhang [d,e], Baehyun Min [a,f,*]

[a] Center for Climate/Environment Change Prediction Research, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, 03760, Republic of Korea
[b] Department of Mathematics and Geosciences, University of Trieste, Italy
[c] National Institute of Oceanography and Applied Geophysics—OGS, Borgo Grotta Gigante 42/C, Trieste, Sgonico, 34010, Italy
[d] College of Safety Science and Engineering, Liaoning Technical University, Huludao, 125105, China
[e] Key Laboratory of Mine Thermodynamic Disasters and Control of Ministry of Education, Huludao, 125105, China
[f] Department of Climate and Energy Systems Engineering, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, 03760, Republic of Korea

ARTICLE INFO

Keywords:
CO_2-EOR
CO_2-Foam experiments
GRNN
CFNN-LM
CFNN-BR
XGBoost

ABSTRACT

Enhanced oil recovery (EOR) using CO_2 injection is promising with economic and environmental benefits as an active climate-change mitigation approach. Nevertheless, the low sweep efficiency of CO_2 injection remains a challenge. CO_2-foam injection has been proposed as a remedy, but its laboratory screening for specific reservoirs is costly and time-consuming. In this study, machine-learning models are employed to predict oil recovery factor (ORF) during CO_2-foam flooding cost-effectively and accurately. Four models, including general regression neural network (GRNN), cascade forward neural network with Levenberg–Marquardt optimization (CFNN-LM), cascade forward neural network with Bayesian regularization (CFNN-BR), and extreme gradient boosting (XGBoost), are evaluated based on experimental data from previous studies. Results demonstrate that the GRNN model outperforms the others, with an overall mean absolute error of 0.059 and an R^2 of 0.9999. The GRNN model's applicability domain is verified using a Williams plot, and an uncertainty analysis for CO_2-foam flooding projects is conducted. The novelty of this study lies in developing a machine-learning-based approach that provides an accurate and cost-effective prediction of ORF in CO_2-foam experiments. This approach has the potential to significantly reduce screening costs and time required for CO_2-foam injection, making it a more viable carbon utilization and EOR strategy.

1. Introduction

Petroleum resources have long been the primary source of fossil-fuel-based energy to meet global energy demands. Due to the limited reserves available, maximizing the extraction efficiency from oil reservoirs has become increasingly important. However, recovering residual oil from mature reservoirs in complex geological formations is still a challenge [1]. In order to address this difficulty, several enhanced oil recovery (EOR) techniques have been developed to extract residual oil further. Among these EOR techniques, gas injection into the reservoir is considered the most efficient approach for mobilizing trapped oil through various recovery processes [2]. Carbon dioxide (CO_2) is particularly effective for this purpose because it sweeps the residual oil via multiple contact miscibility processes suitable for both conventional and unconventional formations [3,4], thereby boosting oil production [5].

In addition, storing CO_2 deep underground also mitigates climate change [6]. Therefore, CO_2-EOR coupled with CO_2 storage is considered one of the most promising ways to reduce the cost of carbon capture, storage, and utilization [7,8]. Various methods have been proposed to utilize carbon dioxide (CO_2) for improving carbon storage and oil recovery performance. One of these methods is to combine CO_2 with bi-polymers to act as a carrier fluid. This approach could enhance the oil recovery efficiency and reduce carbon emissions [9]. Singh et al. [10] suggested a novel approach using natural surfactants for carbon utilization and cleaner production in hydrocarbon fields. This method aims to minimize the environmental impact of oil production and reduce carbon emissions by using natural surfactants. Another method

[*] Corresponding author. Department of Climate and Energy Systems Engineering, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, 03760, Republic of Korea.

E-mail address: bhmin01@ewha.ac.kr (B. Min).

Table 1
Summary of ML modeling studies for prediction of oil production performance.

ML models	Target	Number of Samples	Input variables	Reference
GB	Prediction Oil recovery factor in hydrocarbon fields	831	Reservoir geometry, geological information, transport, storage and fluid properties, saturation ratios and pressure, location	[48]
ANN	Prediction oil recovery factor for microbial enhanced recovery process	–	Microbial kinetic. operational data, and reservoir data	[49]
ANN	Prediction oil production performance of CO_2-WAG process	2100	Gas injection rate, oil saturation, water injection rate, water cut before gas flooding, water injection volume, cycle time, water injection time, production rate, injection pressure, permeability, porosity, thickness, grid size, bottom hole pressure	[50]
ANN	Prediction oil production and carbon storage performance of CO_2-WAG flooding	223	Initial saturation, WAG parameters, time, ratio of vertical and horizontal permeability	[51]
RF	Prediction oil recovery factor of low salinity flooding	1000	LSWI parameters, reservoir & injection temperature, volume injection, formation water composition, and injection water composition	[52]
CNN, LSTM, DNN	Screening EOR methods	735	porosity, depth, oil gravity, permeability, viscosity and temperature	[53]
ANN, DT, ERT, GB, RF, EXBoost	Estimation the CO_2 foam strength	157	Shear rate, temperature, pressure salinity, surfactant concentration foam quality	[54]
MARS, SVM and RF	Evaluation performance of CO_2 storage and oil production in residual oil zones	250	Thickness, depth, permeability, residual oil saturation, CO2 injection rate, bottom hole pressure, initial pressure, temperature	[55]
LSSVM	Prediction oil production performance of CO_2-EOR project	46	CO_2 injection rate, maximum and minimum bottom hole pressure of injection well, oil production rate, CO_2 concentration	[56]
SVM	Prediction shale gas production	573	Gas production, total injection, total proppant, number of stages, horizontal length, pressure, thickness, porosity, permeability, gas saturation	[57]
DT, EXBoost	Prediction oil recovery of experimental nanofluid injection	108	Size, oil density, viscosity, porosity, permeability, salinity, nanoparticles concentration	[58]
LR, MLP, SVM CMIS	Prediction oil recovery of experimental low salinity flooding	1316	Operational parameters, rock properties, oil properties, brine properties, connate water properties	[59]
ANN, SVM, DT	Estimation oil production performance of LSWI core flooding	117	Petrophysical properties, oil viscosity, oil density, residual oil saturation, temperature, brine properties	[60]
ANN	Prediction oil recovery factor of chemical EOR	847	Polymer concentration, salt concentration, rock type, initial oil saturation, petrophysical properties, pore volume flooding, temperature, salinity, molecular weight of polymer	[61]
ANN	Optimization of WAG injection strategy in subsurface reservoirs	166	Injection rate, production rate limit, start of depletion, end of depletion, average pressure	[62]
ANN	Optimization chemical EOR projects	988	Reservoir grid size, petrophysical properties, reservoir temperature, reservoir pressure, initial oil saturation, oil viscosity, oil gravity, salinity	[63]
RF	Optimization oil production and CO_2 storage in WAG process	216	Reservoir properties, WAG parameters, oil properties, depth, layer thickness, initial oil saturation, well operation	[64]

Table 2
Statistical parameters of the collected dataset for CO_2-foam flooding.

Statistical parameter	IOIP (%)	TPVT (cm^3)	Ø (%)	K (mD)	PV (–)	ORF (%)
Mean	93.53	43.68	29.48	12.13	9.76	30.12
Standard deviation	9.44	6.38	6.10	8.83	7.56	21.46
Minimum	50.00	22.00	16.06	0.10	0.20	1.00
Maximum	100.00	49.75	34.90	28.2	36.3	84.62

proposed by Pandey et al. [11] involves using a polymer-based carbonation process in an alkaline medium. This approach has implications for reducing carbon emissions and could contribute to carbon reduction strategies. By utilizing polymers, the process can increase the efficiency of carbon dioxide storage and reduce the amount of CO_2 released into the atmosphere.

While CO_2 injection has shown promise in EOR, it has some drawbacks, such as low sweep efficiency, asphaltene precipitation, and the corrosion of wells [12,13]. In response to these issues, CO_2 foam has been employed to improve the efficiency of CO_2-EOR flooding. However, before implementing CO_2-foam agents in target reservoirs, lab-scale assessments are necessary to identify potential uncertainties and risks, with pilot-scale studies often required before field implementation [14]. Numerous lab- and field-scale research has been conducted to explore the viability and practical characteristics of CO_2-foam EOR and identify essential recovery principles [15–18]. Recently, Chaturvedi et al. [19] conducted a comparative study of different CO_2-EOR methods, including water-alternating gas (WAG), CO_2-foam flooding and carbonated water injection. Their study revealed that CO_2-foam EOR achieved better performance compared to the other methods. This finding suggests that CO_2-foam EOR may be a more effective approach for enhancing oil recovery in certain conditions.

Simulation models have also been employed to understand variables influencing CO_2-foam flooding to improve oil recovery and CO_2 storage capacity [20–22]. However, lab-scale experiments and core-flooding simulations are costly and labor-intensive. Thus, machine learning (ML) models have been proposed as effective means for predicting objective functions in the absence of reservoir big data and mathematical formulations for the target phenomena.

ML models have been extensively employed for EOR research. For example, Cheraghi et al. [23] proposed the use of a deep artificial neural network (ANN) and random forest models to screen the most suitable EOR methods using data from oil and gas journals. Mohammadi et al. [24] evaluated the performance of several neural networks models, including multi-layer perceptron (MLP), cascade forward neural network (CFNN), generalized regression neural network (GRNN), and radial basis function (RBF), in predicting crude oil pyrolysis for thermal EOR based on 2000 samples. They found that a CFNN with Levenberg–Marquardt optimization (CFNN-LM) achieved the best prediction performance, with only 1% of data points labeled as outliers. Similarly, Mahdaviara et al. [25] employed MLP, GRNN, and CFNN models to predict the permeability of carbonate rock formations. They found that the CFNN-LM model exhibited the most accurate predictive performance with a root mean square error (RMSE) of 5.213. Meanwhile, Pan et al. [26] developed a predictive ML model based on extreme gradient boosting (XGBoost) to evaluate reservoir porosity from well log

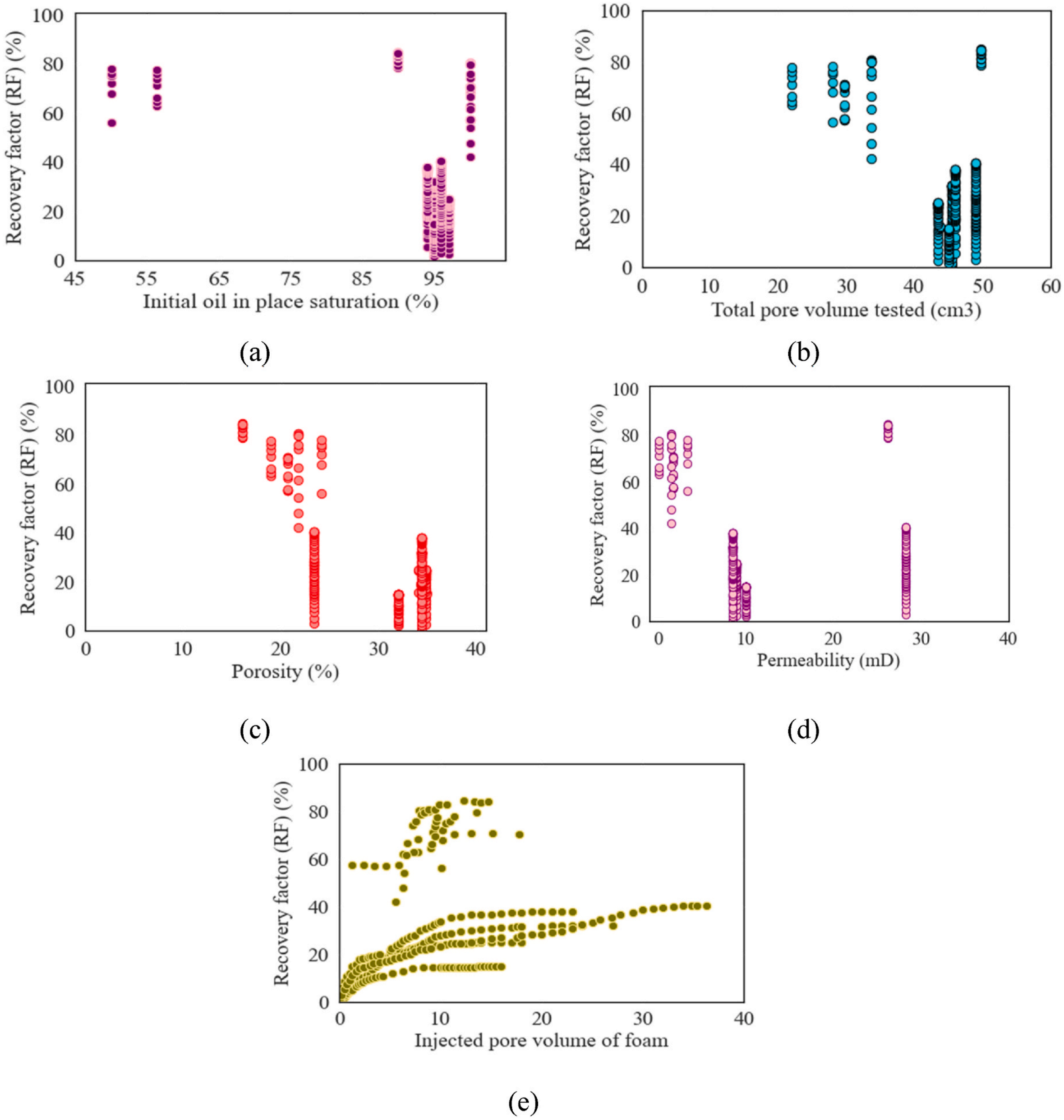

Fig. 1. Scatter plots for the input variables: (a) IOIP, (b) TPVT, (c) Ø, (d) K, and (e) PV for the CO_2-foam versus the ORF.

data. By using a grid search and nature-inspired method to optimize the XGBoost model, they achieved the best predictive results with an RMSE of 0.527. Huang et al. [27] assessed the performance of ANN, light gradient boosting machine (LightGBM) and XGBoost models to predict steam-assisted gravity drainage production. They concluded that training data with a high degree of unpredictability would benefit from the use of an ANN model.

In the petroleum industry and underground gas storage, ML-based models have been utilized for a variety of purposes, such as reserve appraisal in both traditional and non-traditional reservoirs [28–35], assessment of natural gas compressibility [36], prediction of reservoir quality [37], history matching of simulation models for oil production forecasts in fluvial channels [38], lithofacies and petrophysical predictions in carbonate reservoirs [39,40], prediction of cumulative oil production in shale formations [41], microbial enhanced oil recovery [42], pore pressure estimation using petrophysical well log data [43], and distribution 3D geostatistical models [44]. ML models have been used to predict oil recovery factors in several studies. Van Si et al. [45] built an ANN model for predicting the oil recovery factor (ORF) for CO_2-EOR. Esene et al. [46] performed the ORF prediction using ANN, least-squares support vector machine (LSSVM), and gene expression programming (GEP) for a carbonate water-injection process. In their work, the ANN yielded the most accurate prediction performance with an R^2 of 0.99. Recently, Larestani et al. [47] developed a series of ANN models and decision trees to predict the ORF and the net present value of chemical flooding projects, with the CFNN-LM model generating the highest predictive performance. These studies demonstrate the potential of ML models to predict recovery factors and optimize oil recovery processes. Table 1 highlights the employed machine learning approaches for prediction oil recovery performance in EOR projects.

Despite previous research on ML models, little attention has been given to using them for quickly predicting the ORF in CO_2-foam flooding systems, and the implications of the developed models have yet to be well-studied. Moreover, our literature survey reveals that the CFNN-LM, GRNN, XGBoost, and CFNN with Bayesian regularization (CFNN-BR) models are innovative approaches for ORF prediction in CO_2-foam

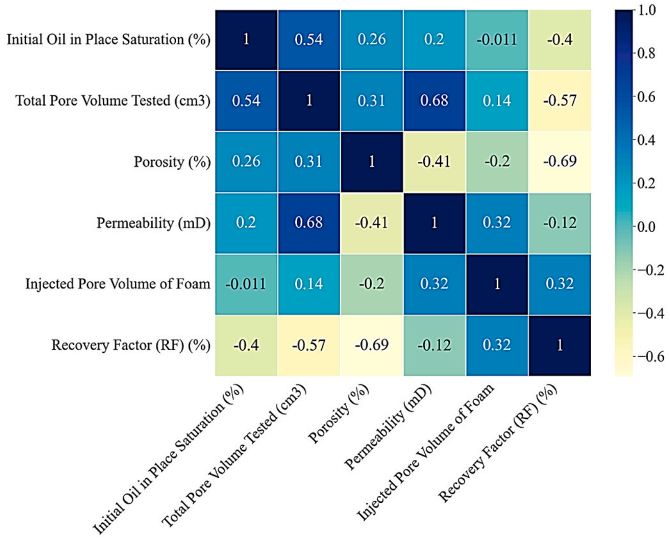

Fig. 2. Correlation heat map for the CO_2-foam dataset.

injection. As these models have demonstrated their effectiveness in various engineering and scientific applications [20–23], we seek to evaluate their performance in our study.

In this study, we aim to develop and evaluate various ML models for the swift and accurate prediction of the oil recovery factor (ORF) in CO_2-foam experiments. Our objective is to identify the optimal ML model reducing the time and cost of experimentation while maintaining prediction accuracy. For model testing, we consider various types of foam and compile a comprehensive dataset. We then utilize the selected ML model in uncertainty analysis for CO_2-foam experiments to determine the optimal ORF for 500 scenarios. Our proposed framework offers an effective solution for quickly predicting the ORF in CO_2-foam flooding systems and can be adapted for other EOR methods.

The remainder of the paper is organized as follows. Section 2 briefly overviews the CO_2-foam EOR process and describes the input features. Section 3 presents the GRNN, CFNN-BR, CFNN-BR, and XGBoost models. Section 4 outlines the research structure, data collection, ML model development, and statistical evaluation approach. Section 5 presents and discusses the numerical results. Finally, Section 6 summarizes the key findings of this study.

2. Methods

2.1. Data

To develop robust prediction models for CO_2-foam flooding, it is essential to establish a comprehensive dataset that reflects the diverse range of settings in which this EOR process can operate. In light of this, we collected 260 experimental data points for CO_2-foam flooding from published studies [17,65–69]. This dataset covers seven foam types with five main input variables, including initial oil in place (IOIP), total pore volume tested (TPVT), porosity (Ø), permeability (K), and injected pore volume (PV) of the foam. The objective of the ML models was to predict the ORF formulated as follows:

$$\mathrm{ORF} = f\,(\mathrm{IOIP}, \mathrm{TPVT}, \varnothing, K, \mathrm{PV}) \tag{1}$$

Previous studies have reported various measurements of CO_2 foam, but not all of these experiments calculated the (ORF), and the majority focused solely on foam stability. Consequently, their data could not be included in the dataset of this research. The final dataset included ionic, nonionic, and cationic surfactants and silica nanoparticles in the CO_2 foam to ensure a diverse range of experimental conditions.

A statistical summary of the input variables is presented in Table 2. The IOIP, TPVT, Ø, K, and PV for the foam ranged from 50 to 100%, 22–49.75 cm^3, 16.06–34.90%, 0.10 to 28.2 mD and 0.20 to 36.30, respectively. The relationship between the input variables and the ORF is illustrated in the scatter plots (Fig. 1).

The correlation heatmap in Fig. 2 shows the relationships between

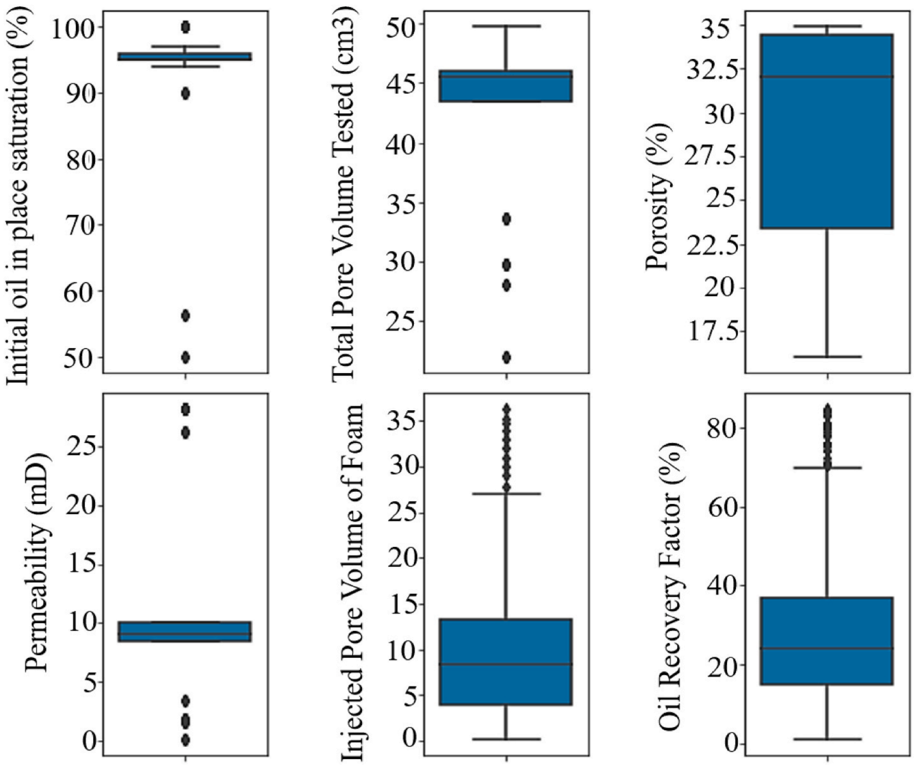

Fig. 3. Box plots to detect outliers in the dataset for the development of the ML models.

the variables in the dataset, with darker blue cells indicating a stronger positive correlation. Overall, permeability was positively correlated with TPVT, injected PV of the foam, and IOIP saturation, with correlation coefficients of 0.68, 0.32, and 0.20, respectively. IOIP saturation also had a positive correlation with TPVT (0.54). On the other hand, the OPF had a negative relationship with permeability, IOIP saturation, TPVT, and porosity, but a positive correlation with the injected PV of the foam.

Before using the dataset to train and test machine learning models, we detected any outliers. Fig. 3 shows the box plots for the input variables and the ORF of the dataset. Although a few outliers were observed, the collected dataset was deemed appropriate for testing models designed to predict the ORF in CO_2-foam flooding processes for EOR applications.

Table 3
Comparison of pros and cons of machine learning models [81].

Methods	Pros	Cons
Liner regression	Simple, easy to interpret, works well with small datasets	Assumes linear relationship between parameters, cannot capture complex nonlinear patterns
Logistic regression	Easy to interpret, works well with binary classification problems	Assumes linear relationship between variables, may not capture complex nonlinear patterns
Decision tree	Easy to interpret, handles nonlinear relationships, works well with both categorical and numerical data	Prone to overfitting, can create complex trees that are difficult to interpret
Random forest	Handles nonlinear relationships, works well with both categorical and numerical data, less prone to overfitting than decision trees	Can be slow and memory-intensive with large datasets
Support Vector Machine	Handles high-dimensional data well, works well with nonlinear relationships, robust to outliers	Can be computationally expensive, may require careful tuning of hyperparameters
General Regression Neural Network	Fast and easy to train, handles noisy data well, can handle non-linear relationships	May require tuning of hyperparameters, less interpretable than some other methods
Cascade Forward Neural Network with Levenberg–Marquardt optimization	Fast and easy to train, can handle complex relationships, can learn from data with noise	May require tuning of hyperparameters, may not always converge to a good solution
Cascade Forward Neural Network with Bayesian regularization	Handles overfitting well, works well with small datasets, can handle complex relationships	May require tuning of hyperparameters, computationally expensive
Extreme Gradient Boosting	Handles complex relationships, works well with both numerical and categorical data, computationally efficient	May require tuning of hyperparameters, less interpretable than some other methods

Table 4
Hyperparameter tuning for this study.

Model	Hyperparameter	Value
GRNN	Spread coefficient	0.075
CFNN-LM CFNN-BR	Activation function	Tansig
	Number of hidden layers	3
	Number of neurons in the hidden layers	10–18
XGBoost	Booster	gbtree
	Learning rate	0.5
	Max depth	9
	Min child weight	10
	n_estimators	400
	reg alpha	0.5
	reg lambda	8

Table 5
Comparison of statistical indicators for the four ML models.

Data	Indicator	GRNN	CFNN-LM	CFNN-BR	XGBoost
Training	R^2	0.9999	0.9954	0.9995	0.9987
	RMSE	0.480	1.097	0.354	0.769
	MAE	0.067	0.670	0.136	0.467
Testing	R^2	0.9999	0.9780	0.9970	0.998
	RMSE	0.186	3.571	1.397	0.971
	MAE	0.040	1.237	0.357	0.616
All	R^2	0.9999	0.9900	0.9985	0.998
	RMSE	0.414	2.161	0.820	0.910
	MAE	0.059	0.840	0.203	0.515

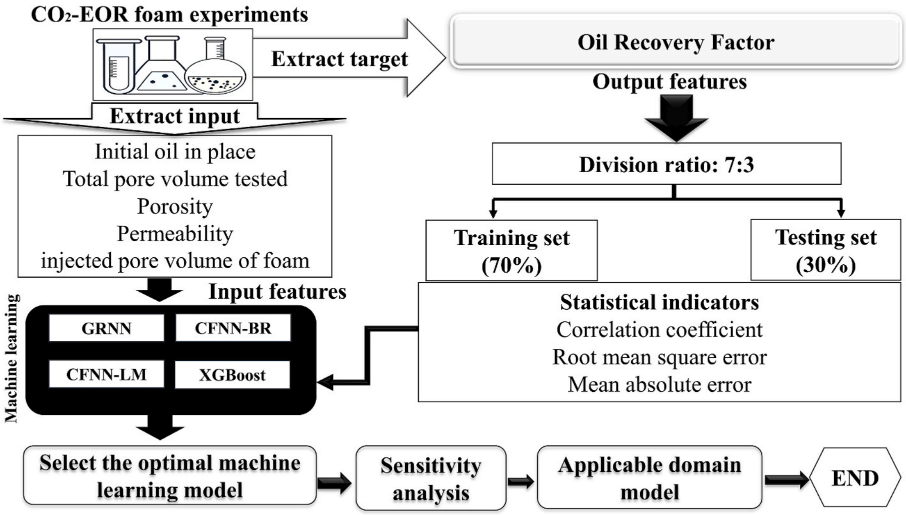

Fig. 4. Workflow of this study to estimate the ORF using four ML Models.

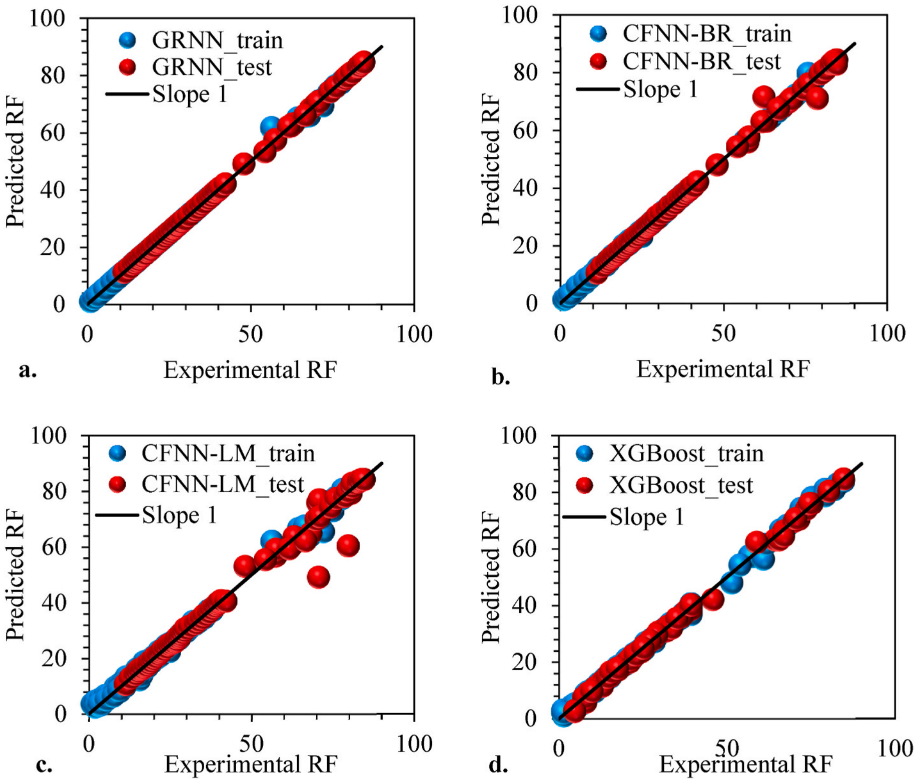

Fig. 5. Cross-plots for the relationship between the experiment and predicted ORF for the four ML models.

2.2. *Theory of machine learning techniques*

2.2.1. *Generalized regression neural network (GRNN)*

GRNN is a powerful form of ANN originally developed to predict continuous output variables [70]. GRNN employs kernel regression and can thus be defined as a normalized radial basis neural network [25]. This form of ANN topology has two benefits: rapid learning rates and low computational costs [71]. Unlike other ANN models, a GRNN does not rely on repetitive computations to predict the relationship between input and output matrices. It can accurately predict such relationships by only using training samples [72]. Comprising input, pattern, summation, and output layers, a GRNN receives data through its input layer and produces model output via its output layer. During the learning phase, the efficiency of the GRNN algorithm is fine-tuned solely by a spread variable (σ) [73].

2.2.2. *Cascaded forward neural network (CFNN)*

Improving network analysis can be achieved by identifying the connections between dependent and independent variables through the addition of more nodes to the feed-forward network [74]. A trainable CFNN is a form of back-propagation ANN that has a unique architecture compared to a conventional feed-forward network. The primary distinction between these topologies is the number of nodes between the output and dependent features [24]. In the first layer of the CFNN, a weighted connection enters from the input layer, but subsequent levels contain weighted connections from both the input layer and all preceding layers. Like other feed-forward networks, the CFNN has one or more linked hidden layers and activation functions, with neurons having biases and weighted connections [75]. The training stage is crucial for optimizing the ANN topology. Thus, two optimization techniques were used to train CFNN models in the present study: Bayesian Regularization (BR) and Levenberg–Marquardt (LM) optimizations [24].

2.2.3. *Extreme gradient boosting (XGBoost)*

XGBoost is a boosting algorithm widely utilized in numerous ML studies [76–78] and is one of the three types of ensemble methods (i.e., bagging, boosting, and stacking) [79]. Ensemble techniques aim to improve the generalization and stability of a single estimator by combining the results of many base estimators derived from a particular learning method [80]. Boosting both the regressor and the classifier reduces the training error by combining weak learners into a strong learner. A random data sample is selected, the model is trained, and then incremental boosting is employed, with each model attempting to compensate for the errors of its predecessor [80]. The XGBoost objective function consists of a loss function and a regularization term. The loss function determines the difference between the estimated value and the target value, while the regularization term prevents overfitting. The objective function for XGBoost is presented in Eq. (2) [78,79]:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\left\|\omega^2\right\| \tag{3}$$

where.

$\hat{y}_i$: predicted value
y_i: real value
$l(y_i, \hat{y}_i)$: loss function
f_k: a term to describe the decision tree structure
$\Omega(f_k)$: regularization term
n: the number of training samples
γ: a term to regulate the number of leaf nodes
T: the number of leaves
λ: a constant used to maintain the leaf node score within acceptable limits to avoid overfitting
ω: the leaf node score

Table 3 presents a comparative analysis between commonly used

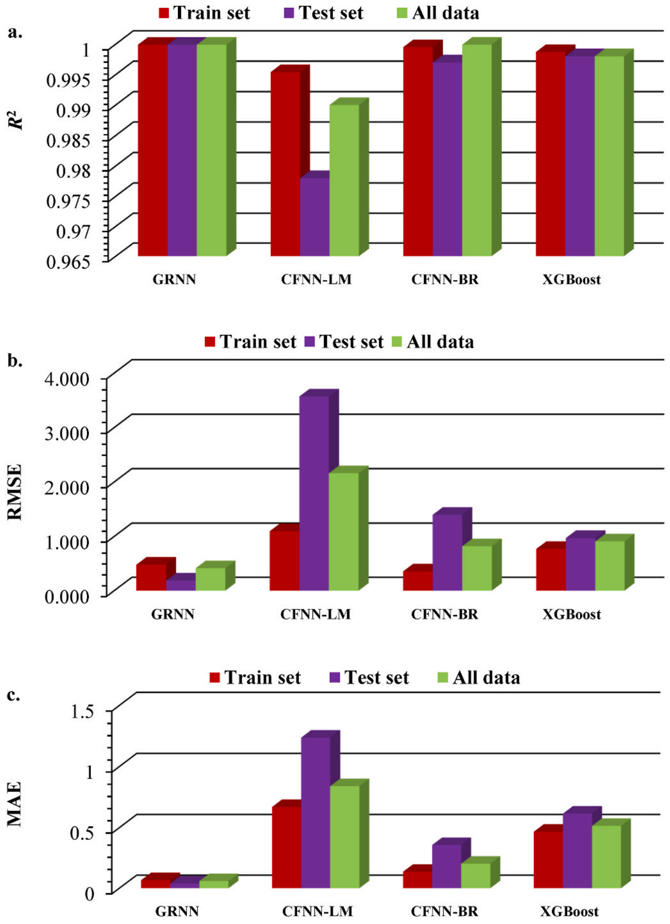

Fig. 6. Statistical performance of the four ML models: (a) R^2, (b) RMSE, and (c) MAE.

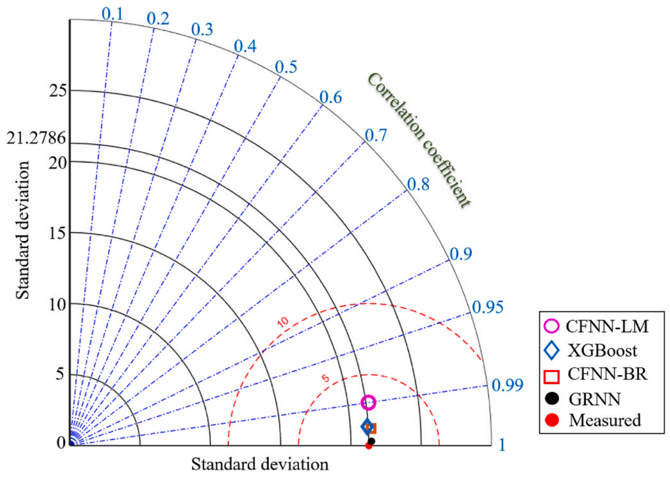

Fig. 7. Taylor diagram for the predictability of the four ML models.

machine learning models and four proposed smart schemes, which are intended to predict the oil recovery factor of CO_2-foam flooding. This comparison aims to demonstrate the advantages and disadvantages of the proposed methods in relation to the existing machine learning models. Overall, all of these methods have their own strengths and weaknesses. Both linear and logistic regression are simple and interpretable but may not be able to capture complex nonlinear relationships. Decision trees and random forests can handle nonlinear relationships but may be prone to overfitting. Support vector machines are robust to outliers but can be computationally expensive. General regression neural networks are fast and can handle noisy data but may require tuning of hyperparameters and are less interpretable. Cascade forward neural networks with Levenberg-Marquardt optimization can handle complex relationships and noise but not guarantee convergence toward the optimum at all times. Cascade forward neural networks with Bayesian regularization can handle overfitting and complex relationships but are computationally expensive. Extreme gradient boosting is computationally efficient and can handle complex relationships but may require tuning of hyperparameters and is less interpretable than some methods. In the context of predicting oil recovery based on CO_2 foam experiments, it would be significant to carefully consider the trade-offs between these different methods and choose the most appropriate one for a specific problem at hand.

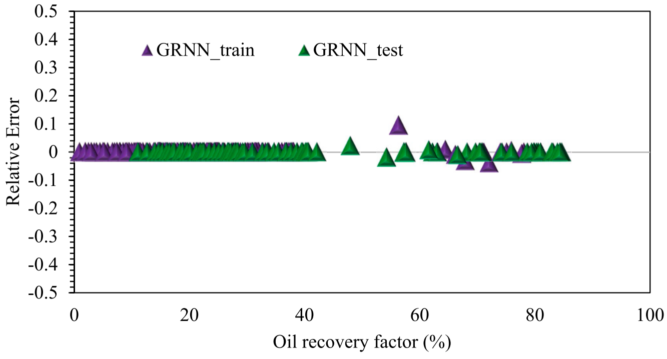

Fig. 8. Relative error distribution for the optimal GRNN model.

2.3. *Workflow*

The ML models were trained using the input variables such as IOIP, TPVT, porosity, permeability, and injected PV of the foam. Fig. 4 illustrates the key processes involved in the proposed methodology.

2.3.1. *Data preparation*

Let us recall that the experimental dataset for CO_2-foam flooding comprised 260 data points collected from previous studies [14,39–43]. The dataset was split into training (70%) and testing (30%) data, with both groups employed in the training and validation phases for developing the ML models.

2.3.2. *ML model development*

To predict the ORF in CO_2-foam experiments, four ML models were implemented: GRNN, CFNN-LM, CFNN-BR, and XGBoost. Their hyperparameters were tuned to obtain optimal prediction results. Table 2 presents a summary of the tuned hyperparameters for the four ML models. The GRNN model utilized the spread constant for the training data, while the CFNN-LM and CFNN-BR models were optimized with three hidden layers, each trained with a Tansig function composed of 10, 14, and 18 neurons, respectively. In the case of XGBoost, a random search was conducted to identify the optimal parameters for the training and testing models. The gbtree booster parameter was employed in the trained model, with 400 tress, a learning rate of 0.5, a maximum depth per tree of 9, L1 regularization on the weights (reg alpha) of 0.5, L2 regularization on the weights (reg lambda) of 8, and a minimum child weight of 10, as shown in Table 4.

2.3.3. *ML model evaluation*

During the ML model development process, model validation plays a critical role in determining the accuracy of the prediction results. In this study, three statistical indicators were used to assess the agreement between the predicted and experimental ORFs: the coefficient of

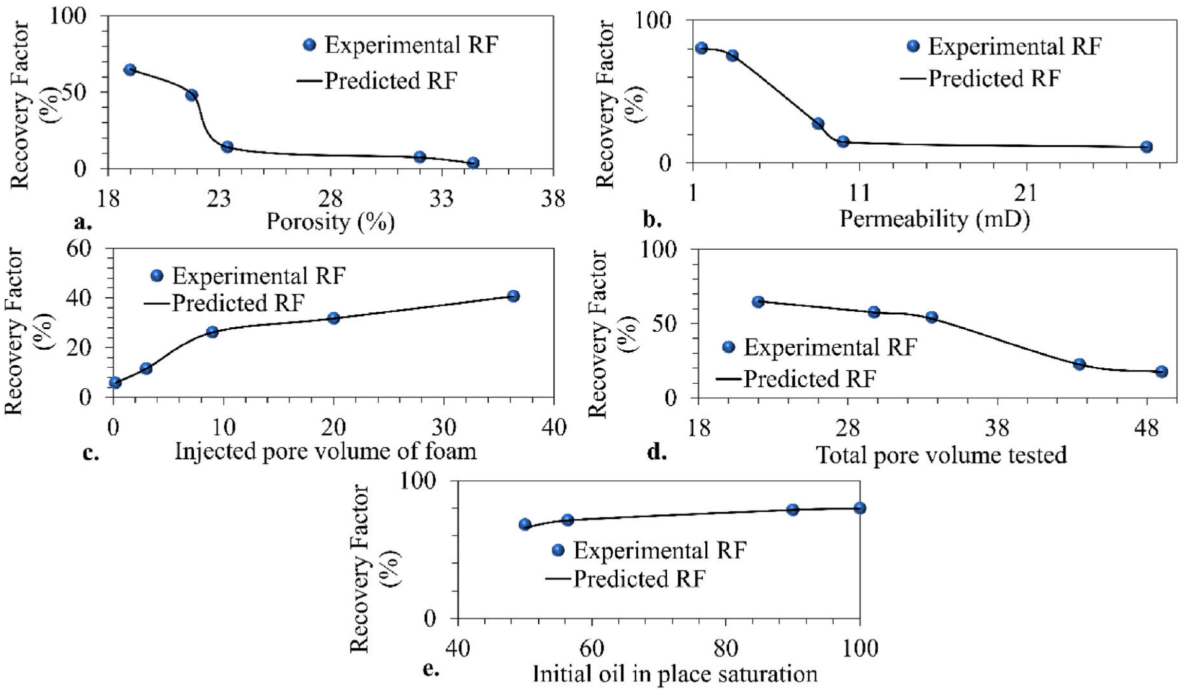

Fig. 9. Experimental ORFs and the ORFs predicted using the GRNN model in accordance with the (a) porosity, (b) permeability, (c) injected pore volume for the foam, (d) total pore volume tested, and (e) the initial oil in place.

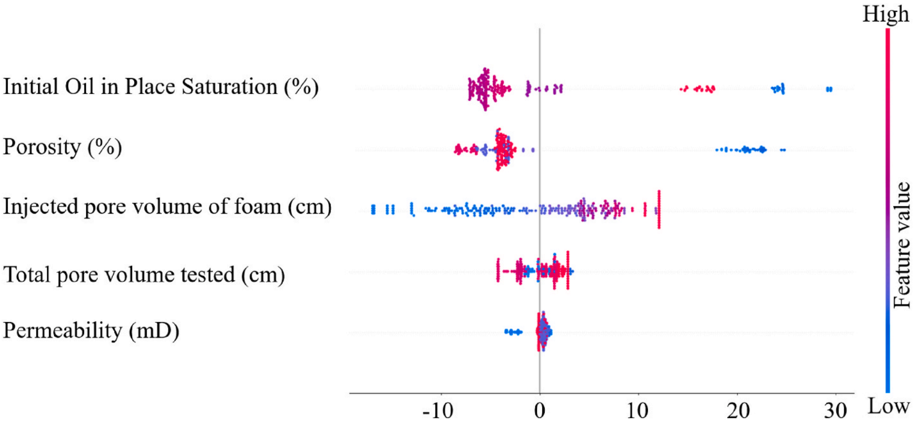

Fig. 10. SHAP values for the influential variables of the proposed model.

determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Equations (4)–(6) were employed to calculate these indicators, as follows

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (RF_i - RF_i^*)^2}{\sum_{i=1}^{n} (RF_i^* - \overline{RF})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (RF_i - RF_i^*)^2} \tag{5}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |RF_i - RF_i^*| \tag{6}$$

where n represents the number of experimental data points, RF_i is the predicted ORF, RF_i^* is the experimental ORF, and $\overline{RF}$ is the average ORF.

3. Results and discussion

3.1. Comparative performance of the four proposed ML models

The predictability of the four ML models was evaluated using statistical analysis and visual inspection. Table 5 shows the R^2, RMSE, and MAE metrics for each ML model. Overall, the GRNN model was the most reliable, closely followed by the CFNN-BR model. During the training phase, all four ML models achieved good prediction results ($R^2 > 0.998$), with the GRNN model performing the best ($R^2 = 0.9999$, RMSE = 0.48, and MAE = 0.67). Similarly, during the testing phase, the GRNN model exhibited excellent prediction performance, with an R^2 of 0.9999, RMSE of 0.186, and MAE of 0.67.

Fig. 5 presents the relationship between predicted and experimentally derived ORFs for CO_2-foam core flooding. The correlation coefficients for the predicted and measured ORFs from the training and testing data were mostly distributed along the fitted line (slope = 1), with the exception of the CFNN-LM model that exhibited a more

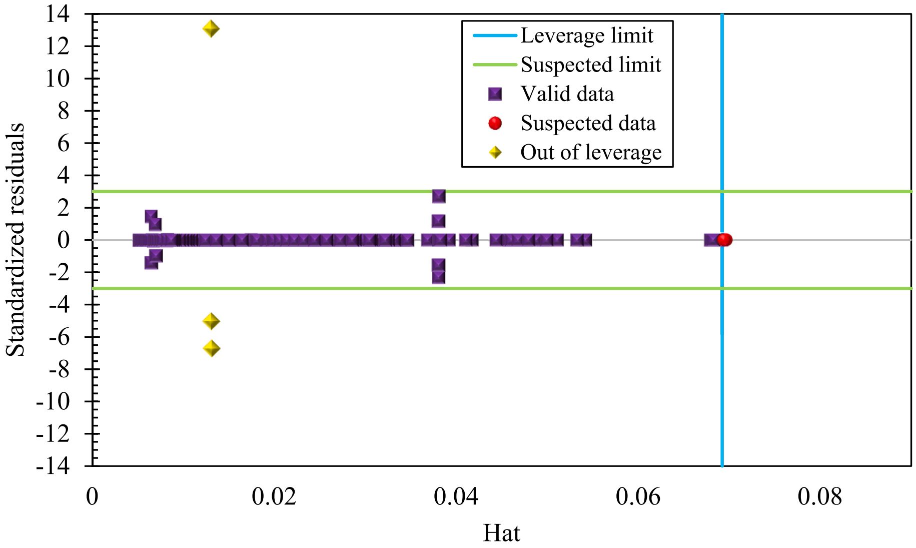

Fig. 11. Williams plot for the GRNN model in the prediction of the ORF for CO_2-foam experiments.

Table 6
Parameters considered for uncertainty analysis.

Variable	Range
Porosity (%)	15–40
Permeability (mD)	10–200
Initial oil in place saturation (%)	30–100
Total pore volume	10–50
Injected pore volume	1–40

scattered distribution. However, all four models achieved reasonable results in predicting the ORF in CO_2-foam flooding experiments. The quality of the training and testing data is illustrated in Fig. 6, which presents a comparison of the R^2, RMSE, and MAE for the four ML models. The GRNN model produced the best results for the training, testing, and combined data, while the CFNN-LM model was the lowest-performing model in both the training and testing phases. In brief, all of the ML models produced an excellent prediction performance, overall.

Fig. 7 presents a Taylor diagram that illustrates the accuracy of the predicted ORF in CO_2-foam experiments based on the correlation factor, R^2, and RMSE. All four ML models showed an excellent ORF prediction performance, though the GRNN model demonstrated the closest fit to the measured ORF data. This suggested that the GRNN model is the optimal choice for the accurate prediction of the ORF in CO_2-foam experiments. The relative error (RE) distribution for the GRNN model further confirms its superiority, with predictions closely clustered around the RE = 0 line and no RE surpassing 0.095 (Fig. 8). As a result, this study analyzed the GRNN model in more detail to assess its potential applicability in CO_2-EOR and carbon storage utilization.

3.2. *Effect of variation in the input variables on the GRNN model ORF predictions*

Fig. 9 illustrates the relationship between input variables and ORF predictions generated by the GRNN model, aiming to enhance our understanding of the topic. The figure displays the predicted and experimental ORFs concerning variations in porosity, permeability, injected PV of foam, TPVT, and IOIP saturation. Notably, there were only minor inconsistencies between the predicted and experimental ORFs each input parameter. These findings suggest that the GRNN model is a reliable tool for forecasting ORFs in diverse CO_2-foam laboratory experiments.

3.3. *Input variable impact analysis*

In this section, the Shapley Additive Explanations (SHAP) [55] technique was utilized to examine the influence of input parameters on the GRNN model. The SHAP values generated using the GRNN model are presented in Fig. 10. Porosity has a substantial negative effect on the ORF since higher porosity levels allow for easier CO_2-foam transport in porous media. Total pore volume and initial oil in place are also critical factors with a significant distribution of high SHAP values. In contrast, permeability and injected pore volume of foam have a weaker effect on ORF. Interestingly, the impact of permeability on CO_2-foam may vary depending on the specific conditions of the reservoir. Permeability is believed to have a minor effect on CO_2 foam performance for the following reasons: (i) CO_2 is highly compressible, enabling it to flow through both lowly and highly permeable porous media. Even in low-permeability formations, CO_2 can reach the oil-bearing regions and generate a foam; (ii) The surfactants used in CO_2-foam applications are designed to create a stable foam with a low critical micelle concentration (CMC), which means the foam generation even at low surfactant concentrations. This allows the foam to be effective even in low-permeability formations where the surfactant may not be able to penetrate the pores as easily; (iii) In some cases, lower permeability may actually be beneficial for consistent CO_2-foam performance preventing form preferential foam flooding through highly permeable conduits. Thus, when estimating the ORF in CO_2-foam experiments, special attention should be paid to porosity and total pore volume to optimize the core-flooding process.

3.4. *Applicability domain for the GRNN model*

In assessing the performance of an ML model, it is crucial to determine its applicability domain. Hence, outlier detection was carried out using the leverage method and a Williams plot [82–84] to evaluate both the GRNN model and the collected dataset. To accomplish it, the standardized residuals (r) derived from the model predictions were plotted against the hat (h) values, which represent the diagonal elements of the hat matrix [85,86]:

$$h = Y(Y^t Y)^{-1} Y^t \tag{7}$$

where Y represents a vector of equal dimensions to $n \times P$, P is the set of input parameters, and Y^t is the transpose X vector.

In the Williams plot, the leverage limitation (h*), which is calculated as $\frac{3(P+1)}{n}$, represents the applicability domain for the developed model

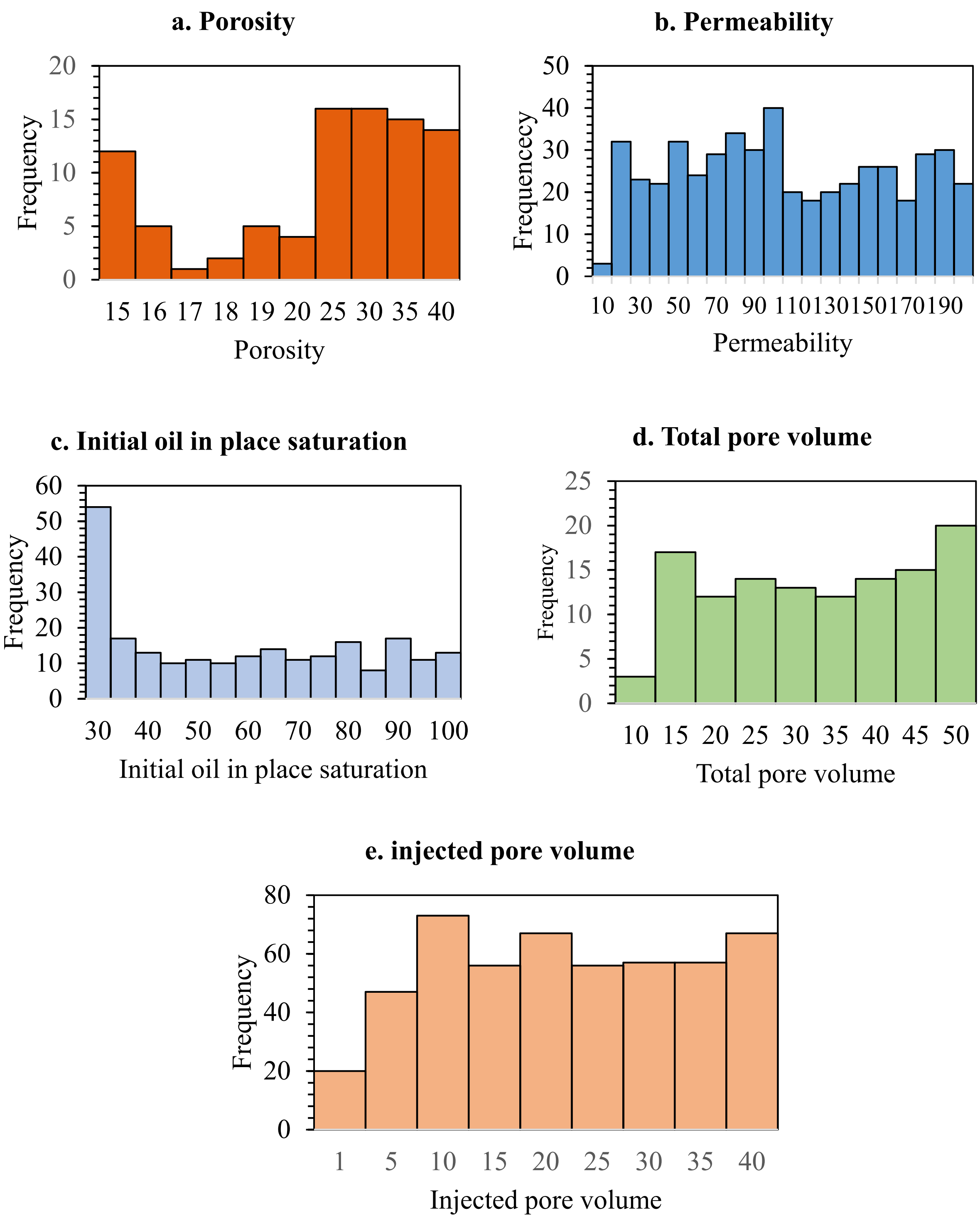

Fig. 12. Distribution of the input range for uncertainty analysis.

with n is for total number of data points [87]. Fig. 11 shows the Williams plot for the ORF predictions obtained using the GRNN model. The analysis revealed that 99.98% of the ORF data points fell within a suitable range ($0 \le h \le 0.06923$ and $-3 \le r \le 3$). These results demonstrate the applicability domain of the GRNN model and the dataset used in this study.

3.5. Implications for the GRNN model

As previous studies have not adequately examined the implications of their developed ML models, this study assessed the potential use of the GRNN model in uncertainty assessment for CO_2-foam flooding projects. The GRNN model can be utilized for conducting Monte Carlo simulations to analyze the uncertainty associated with the input parameters. In this study, we assumed that engineers need to assess the uncertainty of IOIP saturation, TPVT, porosity, permeability, and the injected PV of the foam in order to increase the ORF for a CO_2-foam flooding project. Table 6 summarizes the distribution range of these five input parameters.

The GRNN model has the potential to be a rapid and robust tool for CO_2-foam experiments in selecting suitable parameters of CO_2-EOR projects by assessing a large number of simulation scenarios. Fig. 12

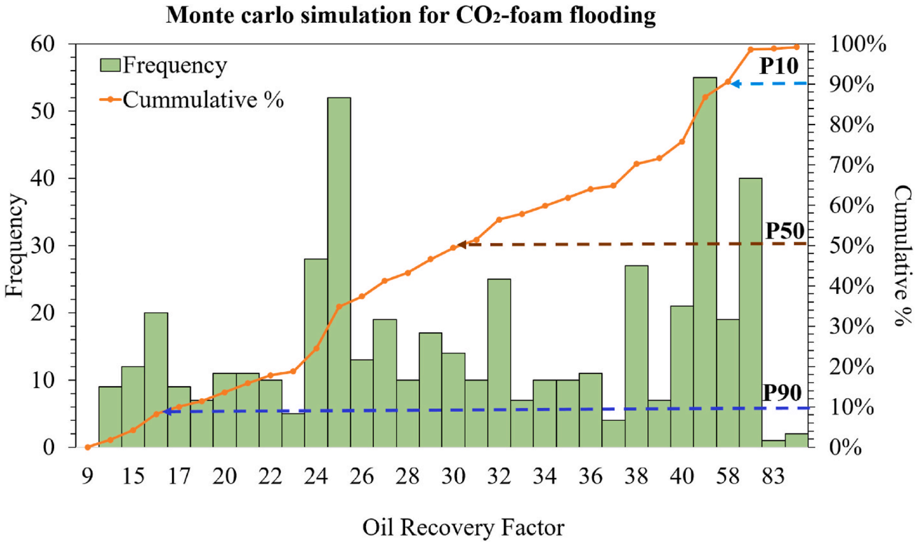

Fig. 13. ORF uncertainty assessment for five input factors in CO_2-foam flooding projects.

Table 7
Optimal design defined according to the uncertainty range.

Variable	Value
Porosity (%)	35
Permeability (mD)	30
Initial oil in place saturation (%)	35
Total pore volume	15
Injected pore volume	5

depicts the distribution of input variables for 500 scenarios generated at random. The GRNN model was then utilized to predict the ORF for these scenarios, and the predicted results were subjected to Monte Carlo simulations for uncertainty analysis of CO_2-foam flooding experiments (Fig. 13). The resulting ORFs for P90, P50, and P10 were 17%, 30%, and 58%, respectively, indicating that the GRNN model provided rapid predictive results for optimal ORF in chemical EOR projects. This finding highlights the potential of ML models to enhance the assessment of CO_2-foam experiments prior to field applications.

Based on the 500 simulation scenarios using the GRNN model, Table 7 shows the recommended design to achieve the desired ORF. This simulation process took only 98 s of CPU time, while corresponding experiments could take months or even years to complete. Therefore, if the GRNN model demonstrates improvement when applied to CO_2-foam experiments for specific reservoirs of CO_2-EOR projects, a rapid GRNN model could be developed to save time and labor costs for those projects.

4. Conclusion

This study evaluated the performance of four ML models (i.e., GRNN, CFNN-LM, CFNN-BR, and XGBoost) for predicting the oil recovery factor (ORF) in CO_2-foam flooding experiments using 260 data points from past research. Our key findings are as follows:

- We presented a method to develop ML models for predicting ORF quickly and saving time and experimental cost required for CO_2-foam experiments.
- Among the five input variables, porosity had the most significant impact on ORF predictions, followed by TPVT, IOIP, injected PV of foam, and permeability.
- The GRNN model was the most accurate in predicting ORF in CO_2-foam experiments, with the lowest MAE of 0.059, highest R^2 of 0.9999, and lowest RMSE of 0.414.
- We verified the applicability domain of the GRNN model using a Williams plot, with only 1.54% of the data points identified as outliers.
- Although the XGBoost, CFNN-BR, and CFNN-LM models were less accurate than the GRNN model, they still provided good prediction results for ORF in CO_2-foam laboratory experiments.
- Overall, this study suggests that ML modeling is a promising approach to reducing time and labor costs associated with CO_2-foam experiments while producing accurate predictions. We also employed the GRNN model to determine the optimal design based on 500 simulation scenarios, which only took 98 s to complete.

Future work will focus on improving the quality of the developed ML models to overcome their limitations. Specifically, we will explore how to extend the models' applicability to different statistical characterizations for predicting ORF in CO_2-foam flooding experiments.

Credit author statement

Hung Vo Thanh: Conceptualization; Data curation; Formal analysis; Methodology; Resources; Validation; Software; Roles/Writing - original draft, Data validation; Danial Sheini Dashtgoli: Methodology; Software; Validation; Roles/Writing - original draft, Data validation; Hemeng Zhang: Methodology; Software; Validation; Roles/Writing - original draft, Data validation; Baehyun Min: Investigation; Supervision; Funding acquisition; Project administration; Visualization; Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grants (No. 2022H1D3A2A01096314) and the Korea Institute of Energy Technology Evaluation and Planning (KETEP) funded by the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20214710100060 and No. 20212010200010).

Nomenclature

ANN	Artificial Neural Network
BR	Bayesian Regularization
CFNN	Cascade Forward Neural Network
CMIS	Committee Machine Intelligent System
CNN	Convolutional Neural Network
DNN	Deep Neural Network
EOR	Enhanced Oil Recovery
ERT	Extremely Randomized Trees
GRNN	Generalized Regression Neural Network
GEP	Gene Expression Programming
IOIP	Initial Oil in Place
LM	Levenberg–Marquardt
LR	Linear Regression
LSSVM	Least-Squares Support Vector Machine
LSTM	Long short-term memory
LWSI	Low water salinity injection
MARS	Multivariate Adaptive Regression Splines
ML	Machine Learning
MLP	Multi-Layer Perceptron
ORF	Oil Recovery Factor
PV	Pore Volume
RF	Random Forest
RMSE	Root Mean Square Error
SVM	Support Vector Machine
TPVT	Total Pore Volume Tested
WAG	Water alternating Gas
XGBoost	Extreme Gradient Boosting

References

[1] Kondori J, Miah MI, Zendehboudi S, Khan F, Heagle D. Hybrid connectionist models to assess recovery performance of low salinity water injection. J Pet Sci Eng 2021;197:107833. https://doi.org/10.1016/j.petrol.2020.107833.

[2] Belhaj H, Abukhalifeh H, Javid K. Miscible oil recovery utilizing N_2 and/or HC gases in CO_2 injection. J Pet Sci Eng 2013;111:144–52. https://doi.org/10.1016/j.petrol.2013.08.030.

[3] Xu Y, Sepehrnoori K, Zuloaga-Molero P, Li B, Yu W. Simulation study of CO_2-EOR in tight oil reservoirs with complex fracture geometries. Sci Rep 2016;6:1–11. https://doi.org/10.1038/srep33445.

[4] Farajzadeh R, Eftekhari AA, Dafnomilis G, Lake LW, Bruining J. On the sustainability of CO_2 storage through CO_2 – enhanced oil recovery. Appl Energy 2020;261. https://doi.org/10.1016/j.apenergy.2019.114467.

[5] Welkenhuysen K, Meyvis B, Swennen R, Piessens K. Economic threshold of CO_2-EOR and CO_2 storage in the North Sea: a case study of the Claymore, Scott and Buzzard oil fields. Int J Greenh Gas Control 2018;78:271–85. https://doi.org/10.1016/j.ijggc.2018.08.013.

[6] Vo Thanh H, Sugai Y, Nguele R, Sasaki K. Integrated work flow in 3D geological model construction for evaluation of CO_2 storage capacity of a fractured basement reservoir in Cuu Long Basin. Vietnam. Int J Greenh Gas Control 2019;90:102826. https://doi.org/10.1016/j.ijggc.2019.102826.

[7] Vo Thanh H, Sugai Y, Sasaki K. Application of artificial neural network for predicting the performance of CO_2 enhanced oil recovery and storage in residual oil zones. Sci Rep 2020;10:18204. https://doi.org/10.1038/s41598-020-73931-2.

[8] You J, Ampomah W, Sun Q. Development and application of a machine learning based multi-objective optimization workflow for CO_2-EOR projects. Fuel 2020;264:116758. https://doi.org/10.1016/j.fuel.2019.116758.

[9] Sharma T, Joshi A, Jain A, Chaturvedi KR. Enhanced oil recovery and CO_2 sequestration potential of Bi-polymer polyvinylpyrrolidone-polyvinyl alcohol. J Pet Sci Eng 2022;211:110167. https://doi.org/10.1016/j.petrol.2022.110167.

[10] Singh A, Chaturvedi KR, Sharma T. Natural surfactant for sustainable carbon utilization in cleaner production of fossil fuels: extraction, characterization and application studies. J Environ Chem Eng 2021;9:106231. https://doi.org/10.1016/j.jece.2021.106231.

[11] Pandey A, Chaturvedi KR, Trivedi J, Sharma T. Assessment of polymer based carbonation in weak/strong alkaline media for energy production and carbon storage: an approach to address carbon emissions. J Clean Prod 2021;328:129628. https://doi.org/10.1016/j.jclepro.2021.129628.

[12] Hemmati-sarapardeh A, Ayatollahi S, Zolghadr A, Ghazanfari M, Masihi M. Experimental determination of equilibrium interfacial tension for nitrogen-crude oil during the gas injection process : the role of temperature , pressure , and composition. J Chem Eng Data 2014;59:3461–9.

[13] Barati-harooni A, Naja A, Hoseinpour S, Tatar A. Estimation of minimum miscibility pressure (MMP) in enhanced oil recovery (EOR) process by N2 flooding using di ff erent. computational schemes 2019;235:1455–74. https://doi.org/10.1016/j.fuel.2018.08.066.

[14] Dang C, Nghiem L, Nguyen N, Chen Z, Nguyen Q. Evaluation of CO_2 low salinity water-alternating-gas for enhanced oil recovery. J Nat Gas Sci Eng 2016;35:237–58. https://doi.org/10.1016/j.jngse.2016.08.018.

[15] Zhao J, Torabi F, Yang J. The role of emulsification and IFT reduction in recovering heavy oil during alkaline - surfactant-assisted CO_2 foam flooding : an experimental study. Fuel 2022;313. https://doi.org/10.1016/j.fuel.2021.122942.

[16] Sibaweihi N, Awotunde AA, Sultan AS, Al-Yousef HY. Sensitivity studies and stochastic optimization of CO_2 foam flooding. Comput Geosci 2015;19:31–47. https://doi.org/10.1007/s10596-014-9446-7.

[17] Al Yousef ZA, Almobarky MA, Schechter DS. Surfactant and a mixture of surfactant and nanoparticles to stabilize CO_2/brine foam, control gas mobility , and enhance oil recovery. J Pet Explor Prod Technol 2020;10:439–45. https://doi.org/10.1007/s13202-019-0695-9.

[18] Zhao J, Torabi F, Yang J. The synergistic role of silica nanoparticle and anionic surfactant on the static and dynamic CO_2 foam stability for enhanced heavy oil recovery : an experimental study. Fuel 2021;287. https://doi.org/10.1016/j.fuel.2020.119443.

[19] Chaturvedi KR, Sharma T. Comparative analysis of carbon footprint of various CO_2-enhanced oil recovery methods: a short experimental study. Chem Eng Commun 2023. https://doi.org/10.1080/00986445.2023.2185518.

[20] Zhang Y, Wang Y, Xue F, Wang Y, Ren B, Zhang L, et al. CO_2 foam flooding for improved oil recovery : reservoir simulation models and influencing factors. J Pet Sci Eng 2015;133:838–50. https://doi.org/10.1016/j.petrol.2015.04.003.

[21] Wei J, Zhou X, Zhou J, Li J, Wang A. Experimental and simulation investigations of carbon storage associated with CO_2 EOR in low-permeability reservoir. Int J Greenh Gas Control 2021;104. https://doi.org/10.1016/j.ijggc.2020.103203.

[22] Fernø MA, Eide Ø, Steinsbø M, Langlo SAW, Christophersen A, Skibenes A, et al. Mobility control during CO_2 EOR in fractured carbonates using foam : laboratory evaluation and numerical simulations. J Pet Sci Eng 2015;135:442–51. https://doi.org/10.1016/j.petrol.2015.10.005.

[23] 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. https://doi.org/10.1016/j.petrol.2021.108761.

[24] Mohammadi MR, Hemmati-Sarapardeh A, Schaffie M, Husein MM, 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;205:108836. https://doi.org/10.1016/j.petrol.2021.108836.

[25] Mahdaviara M, Larestani A, Nait Amar M, Hemmati-Sarapardeh A. On the evaluation of permeability of heterogeneous carbonate reservoirs using rigorous data-driven techniques. J Pet Sci Eng 2022;208:109685. https://doi.org/10.1016/j.petrol.2021.109685.

[26] Pan S, Zheng Z, Guo Z, Luo H. An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. J Pet Sci Eng 2022;208. https://doi.org/10.1016/j.petrol.2021.109520.

[27] Huang Z, Chen Z. Comparison of different machine learning algorithms for predicting the SAGD production performance. J Pet Sci Eng 2021:202. https://doi.org/10.1016/j.petrol.2021.108559.

[28] Miah MI, Ahmed S, Zendehboudi S. Connectionist and mutual information tools to determine water saturation and rank input log variables. J Pet Sci Eng 2020;190. https://doi.org/10.1016/j.petrol.2019.106741.

[29] Esmaili S, Mohaghegh SD. Full field reservoir modeling of shale assets using advanced data-driven analytics. Geosci Front 2016;7:11–20. https://doi.org/10.1016/j.gsf.2014.12.006.

[30] Yasin Q, Sohail GM, Ding Y, Ismail A, Du Q. Estimation of petrophysical parameters from seismic inversion by combining particle swarm optimization and multilayer linear calculator. Nat Resour Res 2020;29:3291–317. https://doi.org/10.1007/s11053-020-09641-3.

[31] Al-qaness MAA, Ewees AA, Vo Thanh H, AlRassas AM, Dahou A, Elaziz MA. Predicting CO_2 trapping in deep saline aquifers using optimized long short-term memory. Environ Sci Pollut Res 2022. https://doi.org/10.1007/s11356-022-24326-5.

[32] Al-Mudhafar WJ, Rao DN, Srinivasan S, Vo Thanh H, Lawe EM Al. Rapid evaluation and optimization of carbon dioxide-enhanced oil recovery using reduced-physics proxy model. Energy Sci Eng 2022;10:4112–35.

[33] Vo Thanh H, Taremsari SE, Ranjbar B, Rahimi E, Rahimi MAE. Hydrogen storage on porous carbon adsorbents : rediscovery by nature-derived algorithms in random forest machine. Energies 2023.

[34] Vo Thanh H, Zamanyad A, Safaei-Farouji M, Ashraf U, Hemeng Z. Application of hybrid artificial intelligent models to predict deliverability of underground natural gas storage sites. Renew Energy 2022;200:169–84. https://doi.org/10.1016/j.renene.2022.09.132.

[35] Al-qaness MAA, Ewees AA, Vo Thanh H, Mutahar A, Abd M. An optimized neuro-fuzzy system using advance nature-inspired Aquila and Salp swarm algorithms for smart predictive residual and solubility carbon trapping efficiency in underground storage formations. J Energy Storage 2022;56:106150. https://doi.org/10.1016/j.est.2022.106150.

[36] Hemmati-sarapardeh A, Hajirezaie S, Reza M, Mosavi A, Nabipour N, Shamshirband S, et al. Mechanics Modeling natural gas compressibility factor using

a hybrid group method of data handling. Eng Appl Comput FLUID Mech 2020;14: 27–37. https://doi.org/10.1080/19942060.2019.1679668.

[37] Rashid M, Luo M, Ashraf U, Hussain W, Ali N, Rahman N, et al. Reservoir quality prediction of gas-bearing carbonate sediments in the qadirpur field: insights from advanced machine learning approaches of SOM and cluster analysis. Minerals 2023;13. https://doi.org/10.3390/min13010029.

[38] Vo Thanh H, Sugai Y. Integrated modelling framework for enhancement history matching in fluvial channel sandstone reservoirs. Upstream Oil Gas Technol 2021; 6:100027. https://doi.org/10.1016/j.upstre.2020.100027.

[39] Al-Mudhafar WJ. Integrating lithofacies and well logging data into smooth generalized additive model for improved permeability estimation : zubair formation , South Rumaila oil field. Mar Geophys Res 2019;40:315–32. https://doi.org/10.1007/s11001-018-9370-7.

[40] Al-Mudhafar WJ. Integrating machine learning and data analytics for geostatistical characterization of clastic reservoirs. J Pet Sci Eng 2020;195:107837. https://doi.org/10.1016/j.petrol.2020.107837.

[41] Al-mudhafar WJ. Polynomial and nonparametric regressions for efficient predictive proxy metamodeling : application through the CO_2-EOR in shale oil reservoirs. J Nat Gas Sci Eng 2019;72:103038. https://doi.org/10.1016/j.jngse.2019.103038.

[42] Ansah EO, Vo Thanh H. Microbe induced fluid viscosity variation : field scale simulation , sensitivity and geological uncertainty. J Pet Explor Prod Technol 2020: 1–24. https://doi.org/10.1007/s13202-020-00852-1.

[43] Zhang G, Davoodi S, Shamshirband S, Ghorbani H. A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques. Energy Rep 2022;8:2233–47. https://doi.org/10.1016/j.egyr.2022.01.012.

[44] Vo Thanh H, Sugai Y, Sasaki K. Impact of a new geological modelling method on the enhancement of the CO_2 storage assessment of E sequence of Nam Vang field, offshore Vietnam. Energy Sources. Part A Recover Util Environ Eff 2020;42: 1499–512. https://doi.org/10.1080/15567036.2019.1604865.

[45] Van Si L, Chon BH. Effective prediction and management of a CO_2 flooding process for enhancing oil recovery using artificial neural networks. J Energy Resour Technol Trans ASME 2018;140:1–14. https://doi.org/10.1115/1.4038054.

[46] Esene C, Zendehboudi S, Shiri H, Aborig A. Deterministic tools to predict recovery performance of carbonated water injection. J Mol Liq 2020:301. https://doi.org/10.1016/j.molliq.2019.111911.

[47] Larestani A, Pezhman S, Hadavimoghaddam F. Predicting the surfactant-polymer flooding performance in chemical enhanced oil recovery : cascade neural network and gradient boosting decision tree. Alex Eng J 2022;61:7715–31. https://doi.org/10.1016/j.aej.2022.01.023.

[48] Makhotin I, Orlov D, Koroteev D, Burnaev E, Karapetyan A, Antonenko D. Machine learning for recovery factor estimation of an oil reservoir: a tool for derisking at a hydrocarbon asset evaluation. Petroleum 2022;8:278–90. https://doi.org/10.1016/j.petlm.2021.11.005.

[49] Pavan PS, Arvind K, Nikhil B, Sivasankar P. Predicting performance of in-situ microbial enhanced oil recovery process and screening of suitable microbe-nutrient combination from limited experimental data using physics informed machine learning approach. Bioresour Technol 2022;351:127023. https://doi.org/10.1016/j.biortech.2022.127023.

[50] Lv W, Tian W, Yang Y, Yang J, Dong Z, Zhou Y, et al. Method for potential evaluation and parameter optimization for CO_2-WAG in low permeability reservoirs based on machine learning. IOP Conf Ser Earth Environ Sci 2021:651. https://doi.org/10.1088/1755-1315/651/3/032038.

[51] Le Van S, Chon BH. Evaluating the critical performances of a CO_2–Enhanced oil recovery process using artificial neural network models. J Pet Sci Eng 2017;157: 207–22. https://doi.org/10.1016/j.petrol.2017.07.034.

[52] Hidayat F, Astsauri TMS. Applied random forest for parameter sensitivity of low salinity water Injection (LSWI) implementation on carbonate reservoir. Alex Eng J 2022;61:2408–17. https://doi.org/10.1016/j.aej.2021.06.096.

[53] Kumar Pandey R, Gandomkar A, Vaferi B, Kumar A, Torabi F. Supervised deep learning-based paradigm to screen the enhanced oil recovery scenarios. Sci Rep 2023;13:1–8. https://doi.org/10.1038/s41598-023-32187-2.

[54] Iskandarov J, Fanourgakis GS, Ahmed S, Alameri W, Froudakis GE, Karanikolos GN. Data-driven prediction of in situ CO_2 foam strength for enhanced oil recovery and carbon sequestration. RSC Adv 2022;12:35703–11. https://doi.org/10.1039/d2ra05841c.

[55] Chen B, Pawar RJ. Characterization of CO_2 storage and enhanced oil recovery in residual oil zones. Energy 2019;183:291–304. https://doi.org/10.1016/j.energy.2019.06.142.

[56] Ahmadi MA, Zendehboudi S, James LA. Developing a robust proxy model of CO_2 injection: coupling Box–Behnken design and a connectionist method. Fuel 2018; 215:904–14. https://doi.org/10.1016/j.fuel.2017.11.030.

[57] Wang M, Hui G, Pang Y, Wang S, Chen S. Optimization of machine learning approaches for shale gas production forecast. Geoen 2022;226:211719. https://doi.org/10.2139/ssrn.4205046.

[58] Ali F, Khan MA, Haider G, ul-Haque A, Tariq Z, Nadeem A. Predicting the efficiency of bare silica-based nano-fluid flooding in sandstone reservoirs for enhanced oil recovery through machine learning techniques using experimental data. Appl Nanosci 2022;12:2367–77. https://doi.org/10.1007/s13204-022-02529-z.

[59] Tatar A, Askarova I, Sha A, Rayhani M. Data-Driven connectionist models for performance prediction of low salinity water flooding in sandstone reservoirs. ACS Omega 2021;6:32304–26. https://doi.org/10.1021/acsomega.1c05493.

[60] Salimova R, Pourafshary P, Wang L. Data-driven analyses of low salinity waterflooding in carbonates. Appl Sci 2021;11. https://doi.org/10.3390/app11146651.

[61] Saberi H, Esmaeilnezhad E, Choi HJ. Artificial neural network to forecast enhanced oil recovery using hydrolyzed polyacrylamide in sandstone and. Polymers (Basel) 2021;13.

[62] Matthew DAM, Ghahfarokhi AJ, Shang C, Ng W, Amar MN. Proxy model development for the optimization of water alternating CO_2 gas for enhanced oil recovery. Energies 2023:16.

[63] Si L Van, Chon BH. Artificial neural network model for alkali-surfactant-polymer flooding in viscous oil reservoirs: generation and application. Energies 2016;9. https://doi.org/10.3390/en9121081.

[64] Li H, Gong C, Liu S, Xu J. Machine learning-assisted prediction of oil production and CO_2 storage effect in CO_2-water-alternating-gas injection (CO_2-WAG). Appl Sci 2022:12.

[65] Ydsteb T. Enhanced oil recovery by CO_2 and CO_2-foam in fractured carbonates. The University of Bergen; 2013.

[66] Turta AT, Singhal AK. Field foam applications in enhanced oil recovery projects: screening and design aspects. J Can Pet Technol 2002;41.

[67] Li RF, Yan W, Liu S, Oil O, Hirasaki GJ, Miller CA. Foam mobility control for surfactant enhanced oil recovery. SPE J 2010;15:928–48.

[68] Yang J, Wang X, Peng X, Du Z, Zeng F. Experimental studies on CO_2 foam performance in the tight cores. J Pet Sci Eng 2019;175:1136–49. https://doi.org/10.1016/j.petrol.2019.01.029.

[69] Zhao J. Comprehensive experimental study on foam flooding for enhancing heavy oil recovery. University of Regina; 2017.

[70] Specht DF, others. A general regression neural network. IEEE Trans Neural Network 1991;2:568–76.

[71] Zeng J, Jamei M, Amar MN, Hasanipanah M, Bayat P. A novel solution for simulating air overpressure resulting from blasting using an efficient cascaded forward neural network. Eng Comput 2021:1–13. https://doi.org/10.1007/s00366-021-01381-z.

[72] Cigizoglu HK. Application of generalized regression neural networks to intermittent flow forecasting and estimation. J Hydrol Eng 2005;10:336–41. https://doi.org/10.1061/(asce)1084-0699(2005)10:4(336).

[73] Cigizoglu HK, Alp M. Generalized regression neural network in modelling river sediment yield. Adv Eng Software 2006;37:63–8. https://doi.org/10.1016/j.advengsoft.2005.05.002.

[74] Jesús O De, Hagan MT. Backpropagation Algorithms for a Broad Class of Dynamic Networks 2007;18:14–27.

[75] Lashkarbolooki M, Vaferi B, Shariati A, Zeinolabedini Hezave A. Investigating vapor-liquid equilibria of binary mixtures containing supercritical or near-critical carbon dioxide and a cyclic compound using cascade neural network. Fluid Phase Equil 2013;343:24–9. https://doi.org/10.1016/j.fluid.2013.01.012.

[76] Meng M, Zhong R, Wei Z. Prediction of methane adsorption in shale : classical models and machine learning based models. Fuel 2020;278. https://doi.org/10.1016/j.fuel.2020.118358.

[77] Vo Thanh H, Yasin Q, Al-mudhafar WJ, Lee K. Knowledge-based machine learning techniques for accurate prediction of CO_2 storage performance in underground saline aquifers. Appl Energy 2022;314. https://doi.org/10.1016/j.apenergy.2022.118985.

[78] Gholami H, Mohamadifar A, Collins AL. Spatial mapping of the provenance of storm dust: application of data mining and ensemble modelling. Atmos Res 2020; 233:104716. https://doi.org/10.1016/j.atmosres.2019.104716.

[79] Zhang D, Qian L, Mao B, Huang CAN, Huang BIN. A data-driven design for fault detection of wind turbines using random forests and XGboost. IEEE Access 2018;6: 21020–31. https://doi.org/10.1109/ACCESS.2018.2818678.

[80] Chen T, Guestrin C. XGBoost : a scalable tree boosting system. San Francisco, CA, USA: ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.; 2016. p. 785–94.

[81] Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. Artif Intell Rev 2006;26:159–90. https://doi.org/10.1007/s10462-007-9052-3.

[82] Nait Amar M. Prediction of hydrate formation temperature using gene expression programming. J Nat Gas Sci Eng 2021;89:103879. https://doi.org/10.1016/j.jngse.2021.103879.

[83] Nait Amar M, Zeraibi N. Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO_2-EOR process. Petroleum 2018. https://doi.org/10.1016/j.petlm.2018.08.001. 0–1.

[84] Nait Amar M, Shateri M, Hemmati-Sarapardeh A, Alamatsaz A. Modeling oil-brine interfacial tension at high pressure and high salinity conditions. J Pet Sci Eng 2019; 183:106413. https://doi.org/10.1016/j.petrol.2019.106413.

[85] Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci 2007;26:694–701. https://doi.org/10.1002/qsar.200610151.

[86] Rousseeuw PJ, Leroy AM. Robust regression and outlier detectionvol. 589. John wiley & sons; 2005.

[87] Vo Thanh H, Nait M, Lee K. Robust machine learning models of carbon dioxide trapping indexes at geological storage sites. Fuel 2022;316:123391. https://doi.org/10.1016/j.fuel.2022.123391.