Advances in
Geo-Energy Research

Vol. 12, No. 2, p. 141-155, 2024

Original article

Accurate determination of nano-confined minimum miscible pressure to aid CO_2 enhanced oil recovery and storage in unconventional reservoirs

Yujiao He[1], Bing Wei[1]*, Jinzhou Zhao[1], Junyu You[2]*, Valeriy Kadet[3], Jun Lu[4]

[1] *State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu 610500, P. R. China*
[2] *School of Petroleum and Natural Gas Engineering, Chongqing University of Science and Technology, Chongqing 401331, P. R. China*
[3] *Gubkin Russia State University of Oil and Gas, Moscow 119991, Russia*
[4] *McDougall School of Petroleum Engineering, The University of Tulsa, Tulsa 74104, USA*

Keywords:
CO_2 enhanced oil recovery and storage
unconventional reservoirs
nano-confinement
minimum miscibility pressure
interpretable machine learning

Cited as:
He, Y., Wei, B., Zhao, J. You, J., Kadet, V., Lu, J. Accurate determination of nano-confined minimum-miscible-pressure to aid CO_2 enhanced oil recovery and storage in unconventional reservoirs. Advances in Geo-Energy Research, 2024, 12(2): 141-155.
https://doi.org/10.46690/ager.2024.05.06

Abstract:
The precise determination of minimum miscible pressure is of great importance for CO_2 enhanced oil recovery and storage as it directly influences the efficiency of pore-scale oil displacement and CO_2 trapping. In this study, an interpretable machine learning framework is developed, enabling the reliable evaluation of nano-confined minimum miscible pressure. Four machine learning algorithms (Random Forest, Multi-layer Perceptron, Support Vector Regression, and eXtreme Gradient Boosting) are employed to accurately predict the nano-confined minimum miscible pressure of a CO_2-oil system. The results demonstrate that, excluding support vector regression, the determination coefficients for all models surpass 94%, signifying the robust predictive performance of our model. Subsequently, Shapley Additive exPlanations is used to analyze the feature importance ranking and the impact of each input feature on minimum miscible pressure in these models. Based on the interpretation results, our multi-layer perceptron model is superior in mining the input-output relationship and reflecting the petrophysical laws, rendering it highly suitable for predicting the minimum miscible pressure while considering nano-confinement. In addition, it is found that pore size significantly influences minimum miscible pressure prediction and that minimum miscible pressure decreases with decreasing pore size when the pore size is ≤ 75 nm. Single-factor sensitivity analysis is applied to validate the trend patterns between input features and minimum miscible pressure in the multi-layer perceptron model.

1. Introduction

In recent years, unconventional oil and gas resources, such as tight/shale oil, have become crucial areas of global oil and gas exploration and development (Wei et al., 2021). Due to the low primary recovery rate of standard oil recovery from these resources, which is only about one-third of that from conventional reservoirs, enhanced oil recovery (EOR) techniques for unconventional reservoirs have garnered significant attention (Sambo et al., 2023). Gas injection has emerged as one of the leading technologies for EOR around the world. The potential of enhancing hydrocarbon recovery in unconventional reservoirs by injecting CO_2 (You et al., 2020) or a hydrocarbon mixture has been demonstrated through experimental measurements (Tovar et al., 2021), numerical simulation (Wei et al., 2022b), and field pilot tests (Alharthy et al., 2018). Meanwhile, CO_2-EOR has also been considered as an important means of achieving carbon neutrality (Cai et al., 2020), since injected CO_2 is partially stored in subsurface formations.

The precise determination of minimum miscible pressure (MMP) of the CO_2-oil system is of great significance for the

---

*Corresponding author.

E-mail address: yujiaoh1228@163.com (Y. He); bwei@swpu.edu.cn (B. Wei); zhaojz@swpu.edu.cn (J. Zhao); youjunyu2013@gmail.com (J. You); kadet.v@gubkin.ru (V. Kadet); jun-lu@utulsa.edu (J. Lu).

Yandy Scientific Press

Received March 11, 2024; revised April 4, 2024; accepted April 21, 2024; available online April 27, 2024.

development of unconventional reservoirs (Zhao et al., 2022), because oil recovery can be improved beyond MMP when CO_2 is miscible with the crude oil in place. However, in unconventional reservoirs with well-developed nanoscale pores, the confinement of crude oil within these nanospaces can significantly alter the thermodynamic phase behavior and related fluid parameters (Hamada et al., 2007; Feng et al., 2020). These changes directly impact the MMP of CO_2-oil systems in confined spaces, because MMP is closely associated with the fluid phase behavior (Teklu et al., 2014b; Wei et al., 2022a). Therefore, more attention should be directed to exploring precise MMP prediction methods for CO_2-crude oil systems in nano-confined spaces.

Existing methods for predicting the MMP of CO_2-oil systems typically involve experimental tests (Novosad et al., 1990; Nguyen et al., 2015; Zhang et al., 2019), empirical modeling (Shokir, 2007; ZareNezhad, 2016), and numerical simulations (Zhao and Fang, 2020; Ge et al., 2021). Experimental techniques, such as the slim-tube experiment (Hao et al., 2020; Lu et al., 2022), rising bubble apparatus (Li and Luo, 2017), and vanishing interfacial tension tests (Safaei et al., 2023), have been widely utilized to measure the MMP under different injected gases. However, these methods are generally applied to measure the MMP of fluids in the bulk phase, rather than in nanopores (Zhang et al., 2017). Over the years, researchers have improved traditional MMP determination methods to predict the MMP of the CO_2-oil system in nano-confined spaces. Bo et al. (2021) created a nanofluidic device capable of gauging fluorescence intensity changes within nanoscale confinement channels to ascertain MMP below 50 nm. Numerous scholars developed theoretical and empirical models for forecasting MMP in nano-confined spaces based on diverse principles. These have included modified vapor/liquid equilibrium calculations combined with the Peng-Robinson equation of state and the vanishing interfacial tension method (Wei et al., 2022b), a multiple mixing cells model coupled with modified vapor/liquid equilibrium procedure (Teklu et al., 2014a), and a modified multiple mixing cells method and Peng-Robinson equation of state (Sun and Li, 2021). In general, the experimental testing of MMP in nanopores is time-consuming, costly, and demands rigorous experimental conditions. In contrast, the outcomes of theoretical models are heavily dependent on the coupled equation of state models, and also lack a standardized format for modified models, thereby constraining their precision and practicality. In summary, few studies exist on MMP within nano-confined space, while the application of existing methods is constrained by factors such as experimental conditions and insufficient theoretical foundation. These shortcomings increase the urgency to explore suitable approaches to study the effect of nano-confinement on MMP.

Machine learning methods have demonstrated to be promising in predicting MMPs owing to their data dependency, high accuracy and high processing speed. Researchers have utilized diverse machine learning algorithms to develop MMP prediction models based on reservoir conditions, crude oil properties, and injection gas compositions, such as the hybrid neural-genetic algorithm (Dehghani et al., 2008), hybrid genetic algorithm-backpropagation neural network (Chen et al., 2014), adaptive boosting support vector regression (Dargahi-Zarandi et al., 2020), hybrid genetic algorithm-support vector machine (Chemmakh et al., 2021), and random forest (Lv et al., 2023). However, many current machine learning models primarily focus on accuracy as the sole evaluation criterion, neglecting the importance of capturing the underlying physical relationships between model input and output. Chen et al. (2022) established a novel support vector regression model, which achieved high accuracy but exhibited notable deviations from actual patterns in practical applications. Thus, besides focusing on prediction accuracy, more effort should be given to enhancing the scientific fundamentals behind these models.

Simple models like linear models or single decision trees are inherently interpretable. In contrast, complex models such as ensemble learning algorithms or neural networks often exhibit better predictive performance but decreased interpretability (Molnar, 2020). Concerning the complex high-dimensional non-linear challenge of predicting MMP in nano-confined spaces, machine learning models may simultaneously present advanced predictive performance and poor model interpretability. The prediction process of these models is categorized as a “black box”, meaning that the process of information flow from inputs to outputs is always invisible. Therefore, interpretable methods are needed to verify that the established model can accurately capture the relationship between the input and output variables and to ensure that it conforms to physical laws. Shapley Additive exPlanations (SHAP), as a game theory-based and model-independent post-hoc interpretable method, can significantly enhance the interpretability of machine learning models without compromising their performance. By providing explanations for individual predictions, SHAP can simplify the interpretation of the model predictions, making them easier to understand. This improved interpretability can help to gain a better comprehension of factors that influence the model predictions, leading to increased trust in the modeling results. In the petroleum industry, SHAP has been successfully used to elucidate the correlations among tight gas production (Ma et al., 2023), coalbed methane production (Min et al., 2023), and MMP derived from the slim-tube experiment and rising bubble apparatus method (Huang et al., 2023), showcasing its outstanding applicability and performance. Accordingly, SHAP is selected and used in this work to balance model performance and interpretability.

We introduce an interpretable machine learning model that incorporates nano-confinement. Pore size is taken for the first time as an input for the prediction of MMP in a pure CO_2-oil system. An in-depth analysis of the model prediction by SHAP reveals the influence of various factors on MMP in the confined space. Initially, 348 datasets are gathered from published works (Teklu et al., 2014a; Hawthorne et al., 2016; Wang et al., 2016). Subsequently, to simplify and enhance the applicability of the established interpretable machine learning framework, four research-based machine learning algorithms rooted in different principles are used for MMP predictions, including Random Forest (RF), Multilayer Perceptron (MLP), Support Vector Regression (SVR), and eXtreme Gradient Boosting (XGBoost). The performance of the models is assessed based on evaluation metrics such as coefficient of determination (R^2), mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE). More importantly, SHAP analysis is performed to consolidate the feature importance of each model and the influence of each input feature on MMP, followed by a comparison with the governing physical laws. The model that best agrees with the dataset is identified by assessing both predictive performance and adherence to SHAP explanation outcomes. Finally, validation is conducted through a single-factor sensitivity analysis. This work offers a novel approach to investigate MMP in nanopores and provides valuable insights into the phase behaviors of CO_2-oil systems at the nanoscale level.

2. Methodology

2.1 Random forest

Since RF was proposed by Breiman (2001) based on ensemble learning, it has found extensive application in regression issues. The term ”forest” in RF refers to an ensemble of multiple decision trees, while the term “random” signifies the use of random sampling and random variable space. For RF regression modeling, the following steps are involved in model construction using a training dataset with N samples and M features:

1) Using the Bagging method, N samples are randomly selected with replacement;
2) A subset of m variables ($m << M$) is randomly chosen as candidate predictors for node splitting, as a contribution to the construction of an individual decision tree. By repeating this process, a large number of regression decision trees are generated;
3) The prediction result of the model is obtained by averaging the predictions made by these regression decision trees.

During the model construction process, the principle for selecting variables at tree nodes is based on minimizing the mean squared error. Specifically, for any variable A and splitting point s, the objective is to minimize the sum of mean squared errors by partitioning the data into two subsets, D_1 and D_2. This process helps to identify the corresponding variable and splitting point, as shown below:

$$\underbrace{\min}_{A,s}\left[\underbrace{\min_{x_i\in D_1(A,s)}\sum_{y_i}(y_1)^2}_{c_1}+\underbrace{\min}_{c_2}\sum_{x_i\in D_2(A,s)}(y_i-c_2)^2\right] \tag{1}$$

where x_i represents the feature attribute; y_i denotes the true value of the sample; c_1 and c_2 represent the sample output mean of dataset D_1 and dataset D_2, respectively.

2.2 Multi-layer perceptron

MLP is a kind of feedforward network consisting of three or more layers: An input layer, one or more hidden layers, and an output layer. It utilizes a nonlinear base function to establish mapping relationships through linear combinations of inputs. The transformation from the input layer to the hidden layer is generally nonlinear, whereas that from the hidden layer to the output layer is linear. Units within the same layer are fully connected, and so are all units between any two layers.

Each unit node in the MLP only receives inputs from the adjacent previous layer. The output of every unit node, except for those in the input layer, is determined by applying an activation function F to the sum of the weighted inputs from the previous layer. Given an input sample x, the output of the i-th unit in the hidden layer r_i can be represented as:

$$r_i=\frac{1}{1+\exp(x)},\quad i=1,2,\cdots,H \tag{2}$$

where H denotes the number of units in hidden layer. Thus, the formula for calculating the output of each unit in the output layer will be:

$$y_j=\sum_{i=1}^{H}w_{ij}r_i(x)+\bar{g}_j,\quad j=1,2,\cdots,Z \tag{3}$$

where w_{ij} represents the weight parameter of the connection between the i-th unit in the hidden layer and the j-th unit in the output layer. The output layer consists of Z units, with y_i being the output of the j-th unit in the output layer. Furthermore, $\bar{g}$ is an estimator of the mean value of y_i.

2.3 Support vector regression

SVR is an application of Support Vector Machine to regression tasks. It is a supervised learning method based on statistical principles, specifically created for solving binary classification issues. SVR utilizes Support Vector Machine to fit a curve to achieve the regression analysis of data, effectively addressing issues such as small sample size, nonlinearity and high dimensionality. The primary steps of SVR are as follows. Given a training sample set D, construct a regression model $f(x)$ that closely approximates y:

$$D=\left[(X,Y_i)\,|X\in R^d,Y_i\in R,i=1,2,\ldots,n\right] \tag{4}$$

$$f(x)=W^D\phi(x)+b \tag{5}$$

where X and Y_i denote the feature and label matrix, R^d represents the d-dimensional real space, W and b denote the parameters to be determined.

The loss is zero only when $f(x)$ exactly matches y. However, SVR assumes that the allowable deviation between $f(x)$ and y is at most ε. The loss is calculated only if the absolute difference between $f(x)$ and y exceeds ε, akin to constructing an interval band of width 2ε centered on $f(x)$. If the training sample falls within this interval, it is considered as correct prediction.

2.4 eXtreme Gradient Boosting

XGBoost is a versatile supervised gradient boosting algorithm that improves the computational efficiency of traditional boosting methods by the parallel combination of weak tree learners with a stronger learner. It incorporates second-

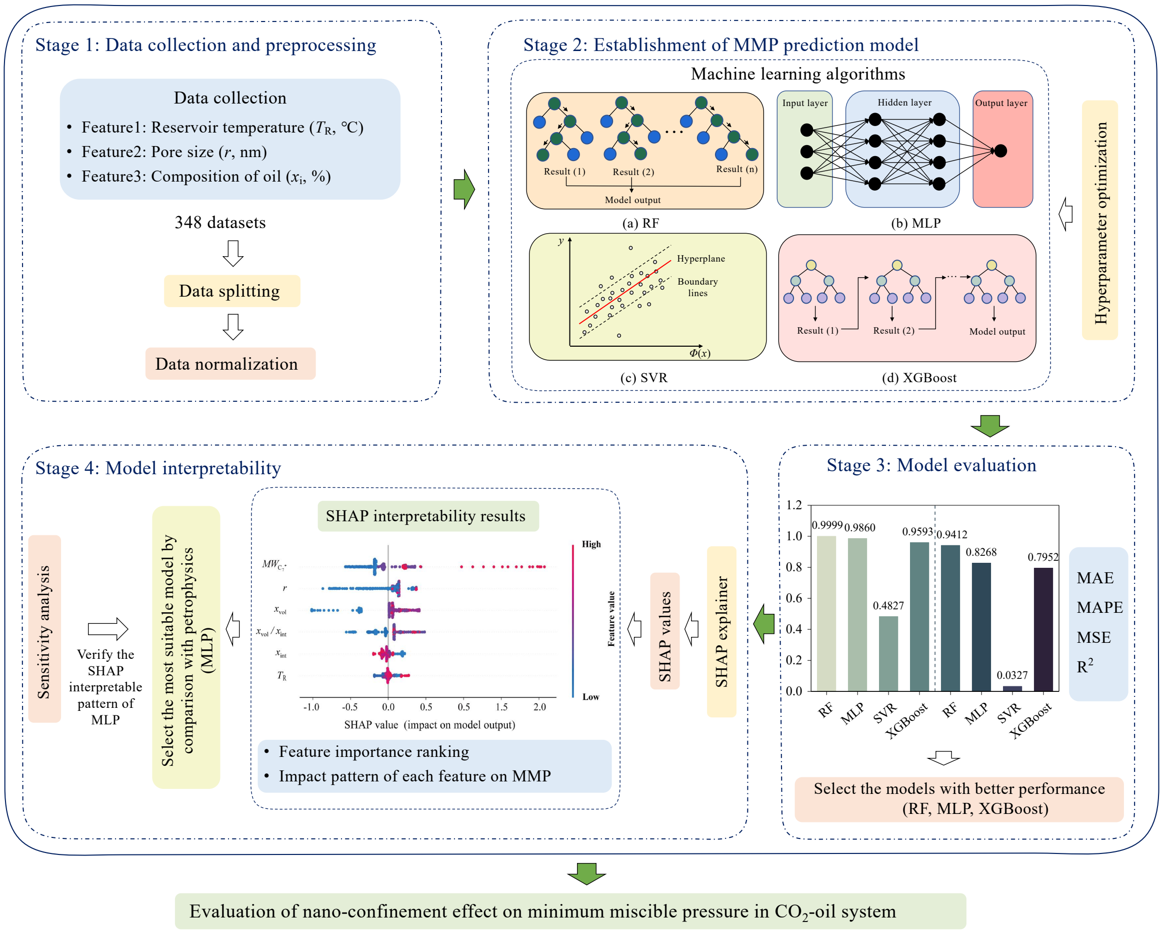


Fig. 1. Workflow of the interpretable machine learning framework.

order Taylor expansion to enhance precision and enables the customization of loss functions through gradient descent. This algorithm features the complexity of tree models in regularization terms to mitigate overfitting. Specifically, it employs shrinkage to reduce the influence of individual trees and eliminate deviations. Compared with traditional boosting approaches, XGBoost introduces model randomization to reduce variance (Li et al., 2022). The primary aim of XGBoost is to minimize the objective function L, depicted by:

$$\begin{aligned} L^{(t)} &= \sum_{i=1}^{n} l\left(y_i,\hat{y}_i\right) + \sum_{k=1}^{K}\Omega\left(f_k\right) \\ &= \sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(t-1)} + f_t\left(x_i\right)\right) + \Omega\left(f_t\right) \end{aligned} \tag{6}$$

where l denotes the loss function, $\hat{y}_i$ indicates the predicted value of the i-th sample generated by the model; t is the number of iterations; K denotes the number of trees; x_i represents the feature vector of the i-th data point; f_k corresponds to the structure q and leaf weights w of the k-th independent tree. Furthermore:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda||\delta||^2 \tag{7}$$

where γ and λ denote the regularization parameters applied to the model; δ represents the scores of every leaf; T indicates the leaves number of the tree.

2.5 Shapley Additive exPlanations

SHAP, proposed by Lundberg and Lee (2017), is a post-hoc interpretability method designed to explain the feature-contribution mechanism of machine learning models by computing the contribution of each input feature to model output. SHAP is represented as a linear additive model of feature contributions, such as:

$$g\left(z'\right) = \phi_0 + \sum_{j=1}^{M}\phi_j z'_j \tag{8}$$

where g denotes the explaining model; $z' \in \{0,1\}^M$ denotes whether the corresponding feature exists (0 or 1); ϕ_0 is a constant; ϕ_j represents the Shapley value of the j-th feature.

Using SHAP for interpretability enhancement involves the application of the Shapley value concept from cooperative game theory, attributing the model output to the contribution of each feature. The Shapley value represents the average marginal contribution of a feature to prediction across all possible feature subsets. Thus, it offers a fair method to distribute the worth or payoff generated by cooperation among participating players, and is defined as:

$$\phi_j = \sum_{S\subseteq(x_1,\ldots,x_M)\backslash x_j} \frac{|S|!(M-|S|-1)!}{M!}\left[\text{val}\left(S\cup x_j\right)-\text{val}(S)\right] \tag{9}$$

where $\text{val}(*)$ represents the value function representing the model and prediction value; $\{x_i\}_{i=1}^{M}$ denotes the features; $(x_i,\cdots,x_M)\backslash x_j$ denotes the set excluding x_j.

3. Establishment of the interpretable machine learning framework

In order to explore the nano-confinement effect on MMP in a pure CO_2-oil system, a SHAP-based interpretable machine learning framework is developed to understand the impact patterns between input features and MMP. This framework comprises three key steps to reveal and predict MMP behavior under nano-confinement, as illustrated in Fig. 1.

Step 1: Data collection and preprocessing. 348 sets of data from pure CO_2 injection processes are first collected from published works and then divided into two subsets: 80% training set and 20% testing set. Subsequently, min-max scaling normalization is applied to bring all data within the range of $[0,1]$.

Step 2: Establishment and evaluation of MMP prediction models. The intricate relationship between model input features and MMP is extracted based on four machine learning algorithms: RF, MLP, SVR, and XGBoost. Grid Search with Cross-Validation (GridSearchCV) is employed to optimize the model hyperparameters. Subsequently, the testing set is input into the trained model using the optimized hyperparameters to obtain predictions. After normalizing the input data, MMP predictions are inversely normalized to generate final forecasts corresponding to the original observations. Finally, the model is evaluated based on the predefined evaluation criteria, and the better-performing models are taken forward for subsequent research.

Step 3: Model interpretation. SHAP is chosen to explain the predictive outcomes of the tested models. Initially, both the dataset and better-performing models are input into the SHAP explainer to obtain the SHAP values of individual samples. Subsequently, summary plot and dependence plot are drawn to illustrate the feature importance ranking and the pattern of impact of each feature on MMP. Moreover, the most suitable model is chosen by comparing the SHAP interpretable results with petrophysics among the better-performing models. Sensitivity analysis is performed to verify the impact patterns of the selected model.

3.1 Dataset description

The input features and data processing method are illustrated in Fig. 2. 348 sets of MMP data from pure CO_2 injection processes are collected, with each set containing three types of group information, including reservoir temperature (T_R), pore size (r), and crude oil composition. The influencing factors include T_R, r, volatile component content (x_{vol}), intermediate hydrocarbon component content (x_{int}), the ratio of volatile to intermediate component content in crude oil (x_{vol}/x_{int}), and the molecular weight of C_{7+} components ($MW_{C_{7+}}$).

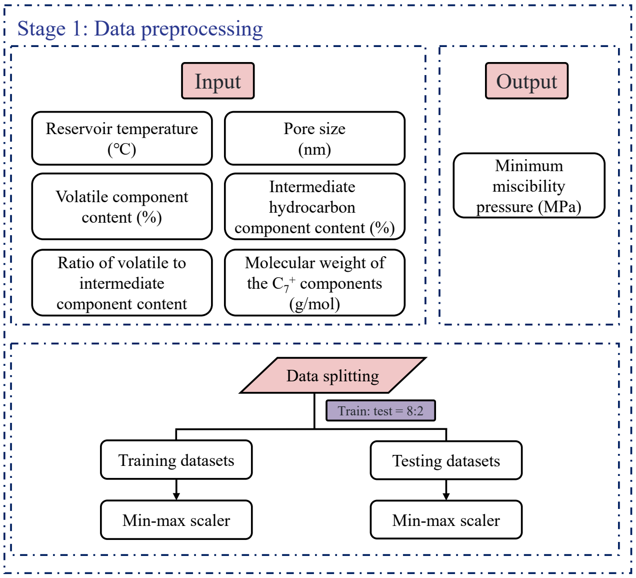


Fig. 2. The data preprocessing step.

3.2 Data normalization

This step involves rescaling the data to achieve zero mean and unit variance. Normalizing input data before training can substantially reduce estimation errors and save computation time. It is a crucial task because many machine learning algorithms assume a normal distribution of data with zero mean and unit variance. The min-max scaler method is employed for data normalization, as defined by:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x_{norm} denotes the standardized value, while x refers to the original value; $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of a specific feature, respectively.

3.3 Hyperparameter optimization

The performance of machine learning models has been proven to be significantly influenced by the chosen parameter configuration (Yang and Shami, 2020). To maximize the performance of established prediction models, the grid search algorithm is employed to explore various parameter combinations, creating a grid where each point represents potential values. Coupled with cross-validation, the grid search algorithm evaluates model performance through dividing the dataset into training and validation subsets. Through iterative training and evaluation, the hyperparameter combination that features the highest accuracy on the validation set is chosen as the optimal configuration. The hyperparameters for optimization are presented in Table 1, employing a 5-fold cross-validation.

3.4 Evaluation criteria

Several evaluation metrics are taken to assess the predictive performance, including MAE, MSE, MAPE, and R^2. MAE quantifies the average absolute discrepancy between the predicted and actual values, revealing the magnitude of prediction

Table 1. Hyperparameters and optimized values of machine learning models.

Algorithm	Hyperparameter	Definition	Optimized value
RF	*n_estimators*	Mumber of decision trees	5
	max_depth	Maximum depth of decision tree	5
	min_samples_split	Minimum number of divisible samples of a node	2
	min_samples_leaf	Minimum number of samples a leaf node contains	3
MLP	*n_layers*	Number of hidden layers	3
	hidden_layer_sizes	Hidden layer size	(70, 30, 20)
	activation	Activation function	tanh
	solver	Solver	lbfgs
SVR	*kernel*	Kernels	rbf
	C	Penalty factor	1,000
	ε	Epsilon	0.01
XGBoost	*max_depth*	Maximum depth of decision tree	5
	learning_rate	Learning rate	1
	subsample	Subsample	1
	colsample_bytree	Feature random sampling ratio	0.75

errors. MSE calculates the mean squared error by averaging the squared differences between predicted and actual values, capturing the variance in prediction errors. MAPE assesses the average absolute percentage difference between predicted and actual values, which indicates the relative size of the prediction errors in relation to the true values. R^2 reflects model fitting, with values closer to 1 indicating better fitting. The expressions for these metrics are as follows:

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{11}$$

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{12}$$

$$\mathrm{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{13}$$

$$\mathrm{R}^2=1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i-y_i)^2} \tag{14}$$

where n denotes the number of samples; y_i represents the actual values; $\hat{y}_i$ signifies the predicted values; $\bar{y}_i$ denotes the mean of the actual values.

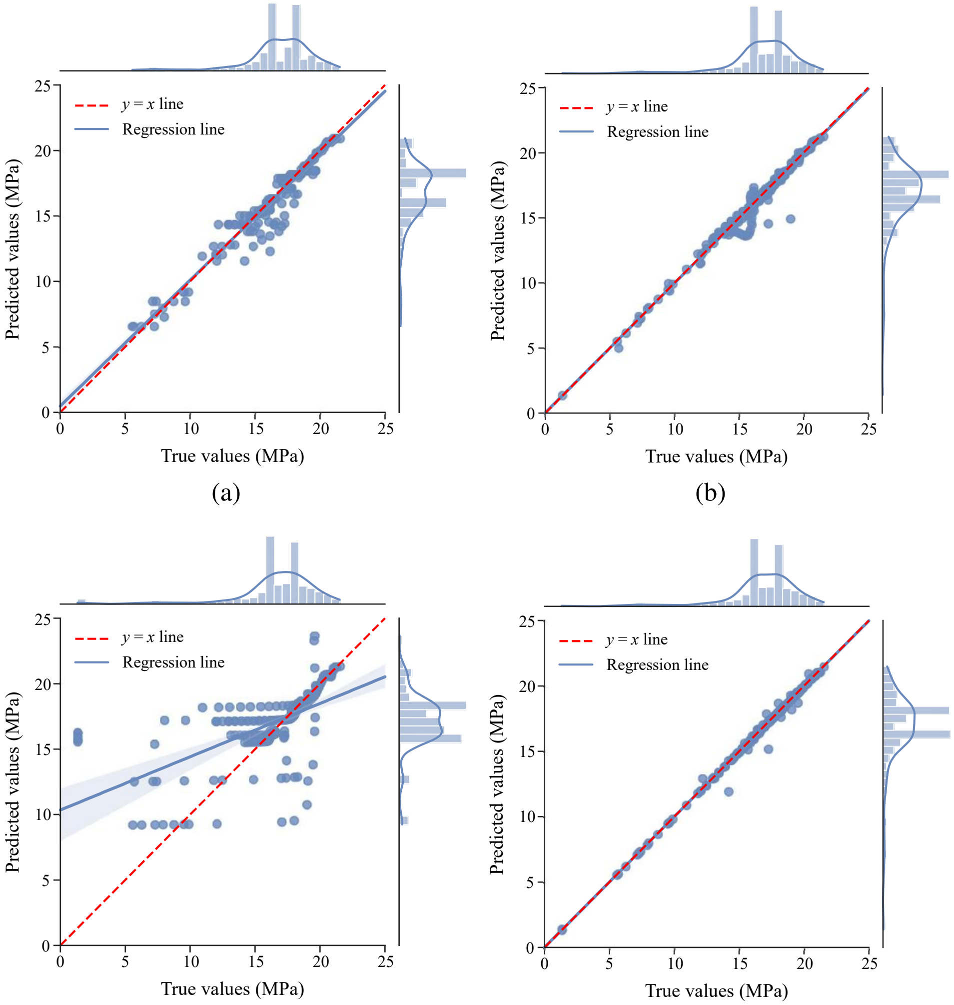
4. Results and analysis

4.1 Performance of prediction models

Four models containing six features, T_R, r, x_{vol}, x_{int}, x_{vol}/x_{int} and $MW_{\mathrm{C}_{7+}}$, are constructed using the dataset as described previously. The raw data are preprocessed and used to train the MMP prediction models. The hyperparameters are optimized by the grid search algorithm, with the results shown in Table 1. The testing set is then fed to the models with optimized hyperparameters to predict the MMP. The model performance is evaluated using the aforementioned four assessment parameters.

The outcomes of the prediction models are presented in Fig. 3, with Figs. 3(a) to 3(d) representing the results for the RF, MLP, SVR, and XGBoost models, respectively. Each data point corresponds to a set of true values and the matching predicted values for MMP, and the red dashed line represents the ideal curve. The proximity of the points to the ideal curve indicates the degree of precision at those specific points. Moreover, a regression line (blue solid line) is fitted to all data points, and the deviation of this line from the ideal curve reflects the magnitude of prediction error. A larger deviation signifies a larger error in the predictions. The bar charts in each subplot display the marginal distributions of true values and predicted values, with the upper bar of the graph representing the distribution of the true values and the right side depicting the distribution of the predicted values. Find that the data points in Figs. 3(a), 3(b) and 3(d) deviate much less from the curve and the regression line almost completely overlaps with the curve. From these results, aside from the SVR model, it can be deduced that the performance of the tested models is relatively good.

The bar charts in Fig. 4 illustrate the performance of the four models. A smaller value of MAE, MSE and MAPE, along with larger R^2, indicates a better performance of the prediction model. A comparison of these parameters reveals that the SVR model exhibits poorer performance than the other three models, both on the training and the testing datasets. Specifically, the SVR model demonstrates the lowest R^2 value,



(a) (b) (c) (d)

Fig. 3. Results of MMP prediction across the tested models. (a) RF model, (b) MLP model, (c) SVR model, and (d) XGBoost model.

Table 2. Values of model evaluation metrics for the total dataset.

Prediction model	Evaluation metrics			
	MAE	MSE	MAPE	R^2
RF	0.4111	0.3968	0.0268	0.9431
MLP	0.2412	0.2079	0.0152	0.9730
SVR	1.1227	7.1389	0.2579	0.3460
XGBoost	0.0560	0.0411	0.0040	0.9950

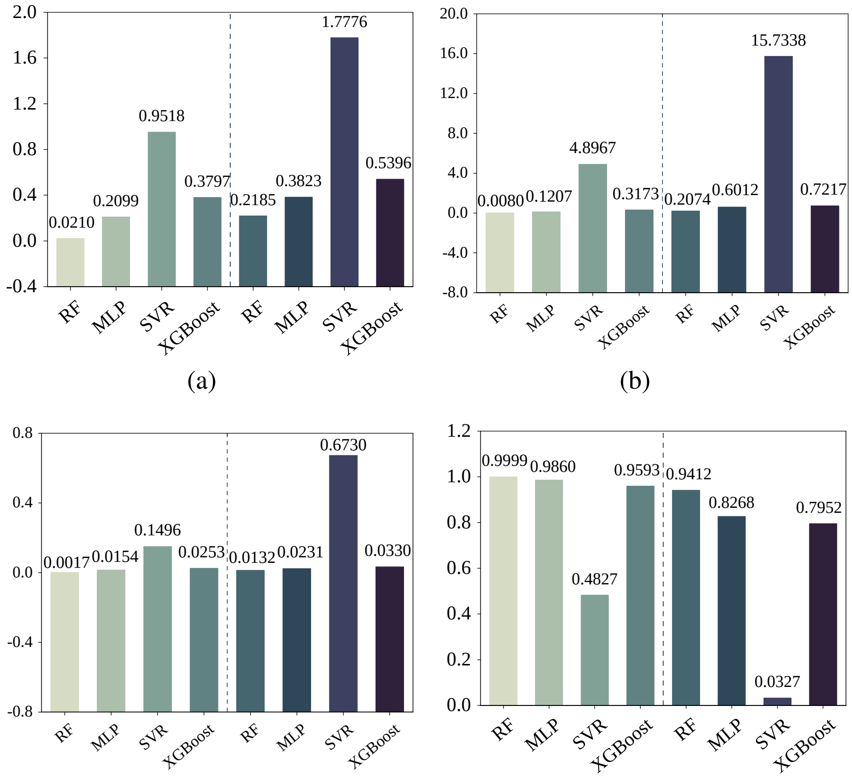
accompanied by the highest MAE, MSE and MAPE values, which is possibly caused by the uneven distribution of data. The other three models show more robust performance, thereby meeting the predetermined performance criteria. The evaluation of the total dataset in Table 2 reveals that these three models achieve R^2 values exceeding 0.94 on the overall dataset while simultaneously demonstrating comparable and relatively lower MAE, MSE and MAPE values. In summary, based on R^2, the predictive performance of the four models can be ranked as: XGBoost $>$ MLP $>$ RF $>$ SVR, which is also consistent with the other three evaluation parameters.

4.2 Model interpretation based on SHAP

4.2.1 SHAP summary plot

As indicated by the preceding analysis, machine learning models based on different principles exhibit varying performance in MMP prediction. The RF, MLP and XGBoost models demonstrate favorable performance and are proven to be capable of meeting the stringent prediction requirements. However, determining the optimal predictive model based solely on predictive performance remains unreasonable. Consequently, SHAP is integrated to further refine model selection. Given that the predictive performance of the SVR model falls short of the requirements, this model is excluded from the subsequent model interpretation.

Fig. 5 depicts the global interpretability results of SHAP, which primarily reflects the feature importance with feature



(a) (b)

(c) (d)

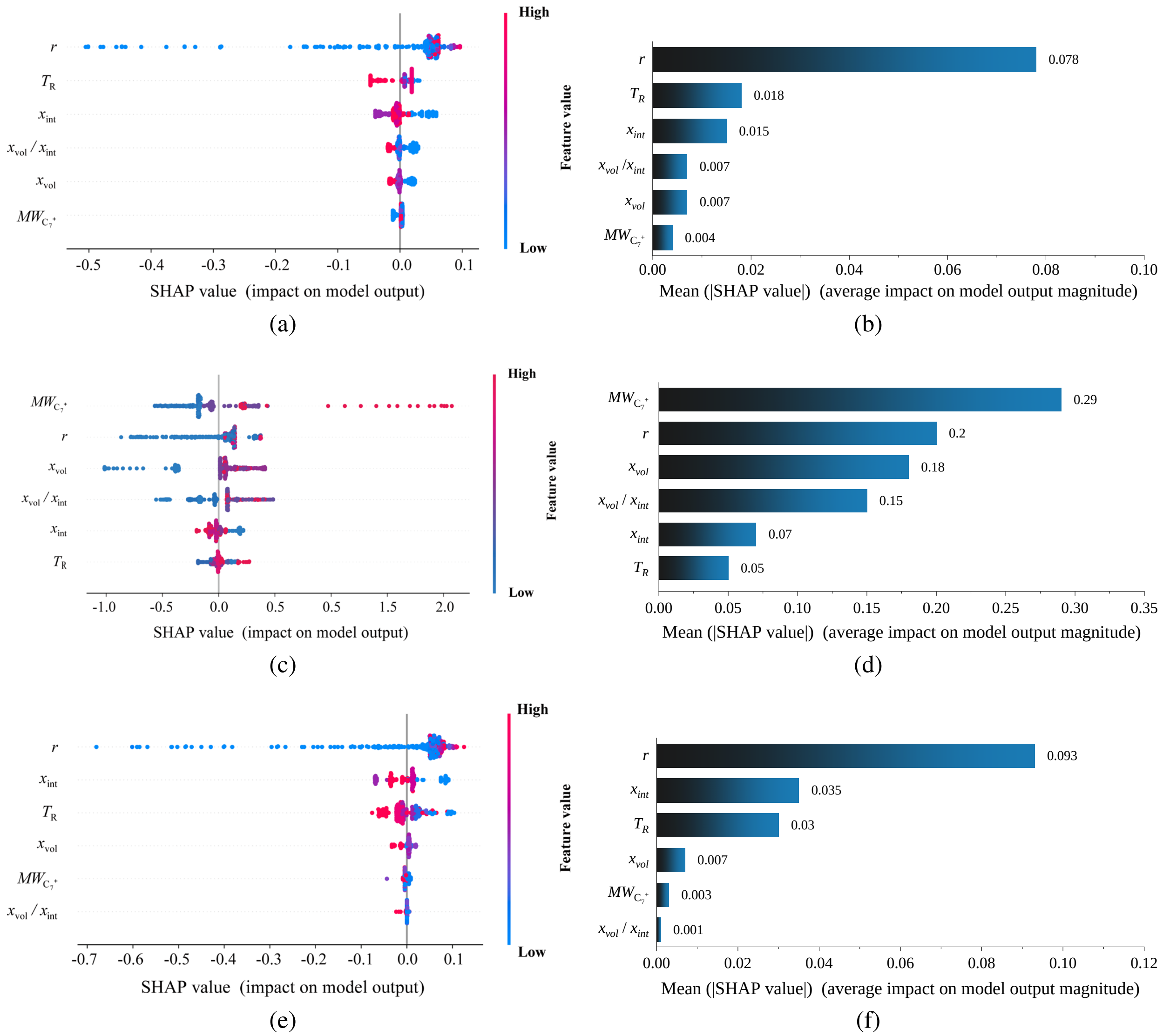
Fig. 4. Values of model evaluation parameters for both the training and testing sets. The left side of the dotted line represents the training set, while the right side represents the testing set. (a) MAE, (b) MSE, (c) MAPE, and (d) R^2.

Table 3. Feature importance ranking of three MMP prediction models.

Prediction model	Feature importance ranking	Range of mean absolute SHAP values
RF	$r > T_R > x_{int} > x_{vol}/x_{int} > xvol > MW_{\mathrm{C}_{7+}}$	0.004-0.078
MLP	$MW_{\mathrm{C}_{7+}} > r > x_{vol} > x_{vol}/x_{int} > x_{int} > T_R$	0.05-0.29
XGBoost	$r > x_{int} > T_R > x_{vol} > MW_{\mathrm{C}_{7+}} > x_{vol}/x_{int}$	0.001-0.093

effects. Figs. 5(a), 5(c) and 5(e) present the SHAP values for each feature within these three MMP prediction models. The x-axis represents the magnitude of the SHAP value, encompassing both positive and negative values. Positive SHAP values mean that the input feature contributes to an increase in the output, whereas negative values suggest that the input feature leads to a decrease in the output. The y-axis denotes distinct input features. Each point within the subplots corresponds to the SHAP value of a specific feature within a given sample. The vertical color bar transitions from blue to red, indicating the progression of the feature's raw values from small to large. The color distribution horizontally along the x-axis for each feature provides insight into the general relationship between a feature's raw values and its SHAP values. For example, when the color of the data point changes from red to blue, this indicates a decreasing impact of the corresponding input feature on the model output, whereas a change in color from blue to red represents an increasing impact of the feature. In places with a high density of SHAP values, the points are stacked vertically. Examining the distribution of SHAP values reveals how a variable may influence the actual model's predictions. In addition, Figs. 5(b), 5(d) and 5(f) show that feature importance in the model output is more intuitive based on the average absolute SHAP value of each feature across all input samples.

As depicted in the subplot of Fig. 5, the features influencing MMP are ranked in decreasing order according to their contribution level. For example, if the contribution of r to MMP is more than that of T_R, this suggests that the interpretability based on r is much more reliable than T_R, as shown in Fig. 5(a). The feature importance ranking of the three MMP prediction models is listed in Table 3. These data demonstrate that there is significant variability in the feature importance rankings among the three models when SHAP explanation is used. As for the RF model, the most influential feature is r, with the mean absolute SHAP value being 0.078 and the relative value of $MW_{\mathrm{C}_{7+}}$ being 0.004, constituting a difference of 0.074. Only the importance of



(a) (b)

(c) (d)

(e) (f)

Fig. 5. SHAP summary plots of the prediction models. (a), (c), (e) denotes the SHAP values of individual samples, while (b), (d), (f) denotes the mean absolute SHAP values of each parameter. (a)-(b) for RF model; (c)-(d) for MLP model and (e)-(f) for XGBoost model.

feature r demonstrates consistent ranking across all models, consistently ranked in the top three (first in RF and XGBoost and second in MLP).

4.2.2 SHAP dependence plot

Given the notable disparities in the feature importance elucidated by SHAP across the RF, MLP and XGBoost models, a series of SHAP dependence plots are drawn to illustrate the correlation between each feature and MMP, as shown in Figs. 6-8. These plots display the SHAP values of input features (y-axis) and the raw values of features (x-axis) for individual samples. The subplots (a) to (f) represent the dependence plots of T_R, r, x_{vol}, x_{int}, x_{vol}/x_{int} and $MW_{C_{7+}}$, respectively.

The SHAP values represent the average marginal contribution of each input feature toward the prediction of MMP. They quantify the degree of each feature value deviating from the predicted value of MMP from the baseline prediction (Min et al., 2023), which is the average prediction value of the dataset. As the predictions are based on the average value, the dependence plot effectively depicts the consistency between input features and their corresponding SHAP values, which is in agreement with their observed impact on MMP predictions.

In the dependence plots, the position of points reflects the impact of feature values on the predicted MMP values. For instance, as depicted in Fig. 6(a), within the range of 80-110 °C, the corresponding SHAP values are greater than zero, indicating that increasing T_R leads to an increase in the predicted MMP. On the other hand, within the range of 110-130 °C, an increase in T_R results in a decrease in the predicted MMP value. Furthermore, it is important to highlight the vertical dispersion of SHAP values for each point in Figs. 6-8, mainly caused by the interactions among the features. For example, when x_{vol} has a value of 0.16, the range of SHAP

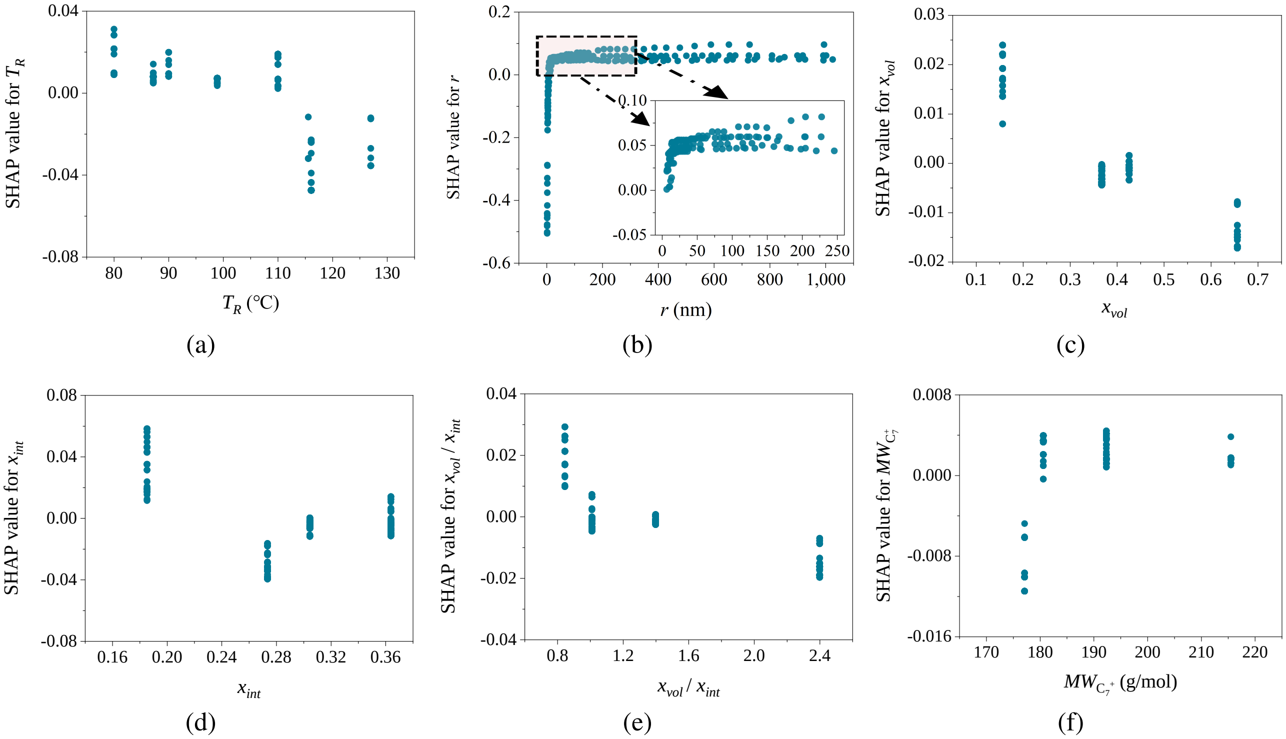


Fig. 6. SHAP dependence plots of the RF model for various input features. (a) T_R, (b) r, (c) x_{vol}, (d) x_{int}, (e) x_{vol}/x_{int}, and (f) $MW_{C_{7+}}$.

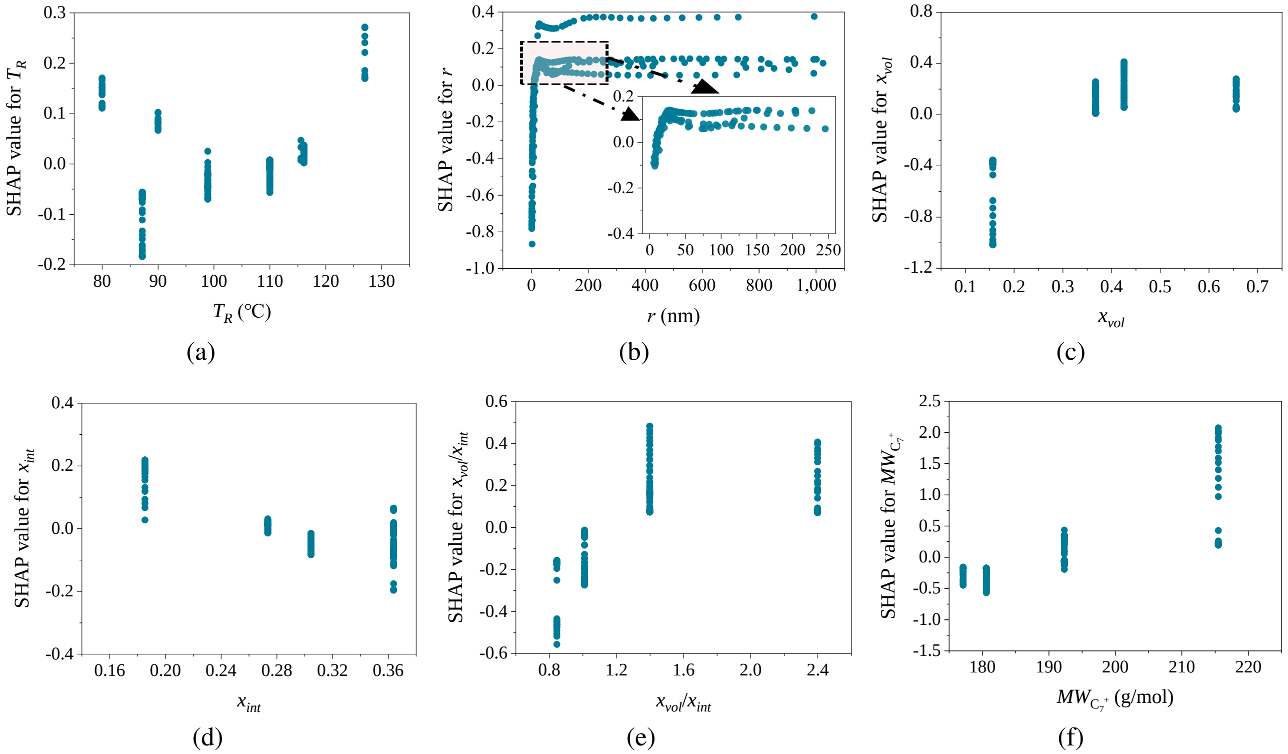
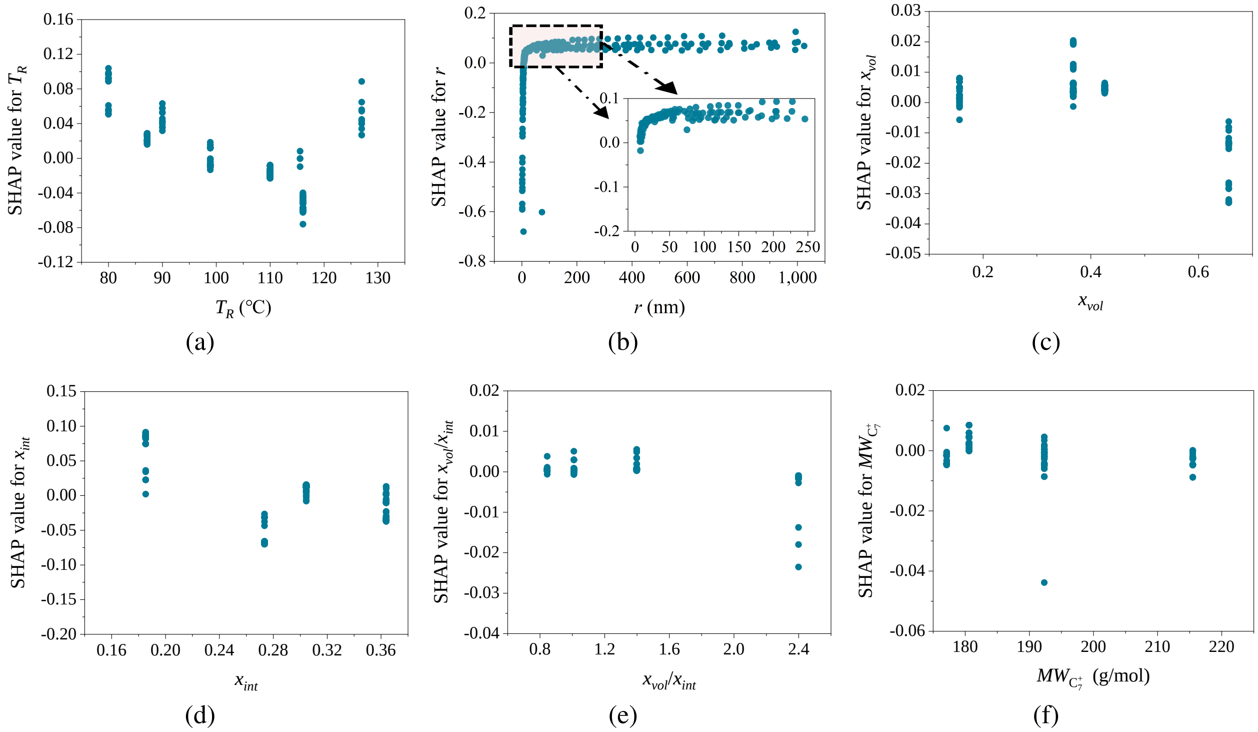


Fig. 7. SHAP dependence plots of the MLP model for various input features. (a) T_R, (b) r, (c) x_{vol}, (d) x_{int}, (e) x_{vol}/x_{int}, and (f) $MW_{C_{7+}}$.



(a) (b) (c)

(d) (e) (f)

Fig. 8. SHAP dependence plots of the XGBoost model for various input features. (a) T_R, (b) r, (c) x_{vol}, (d) x_{int}, (e) x_{vol}/x_{int}, and (f) $MW_{C_{7+}}$.

Table 4. Interpretability determined by SHAP.

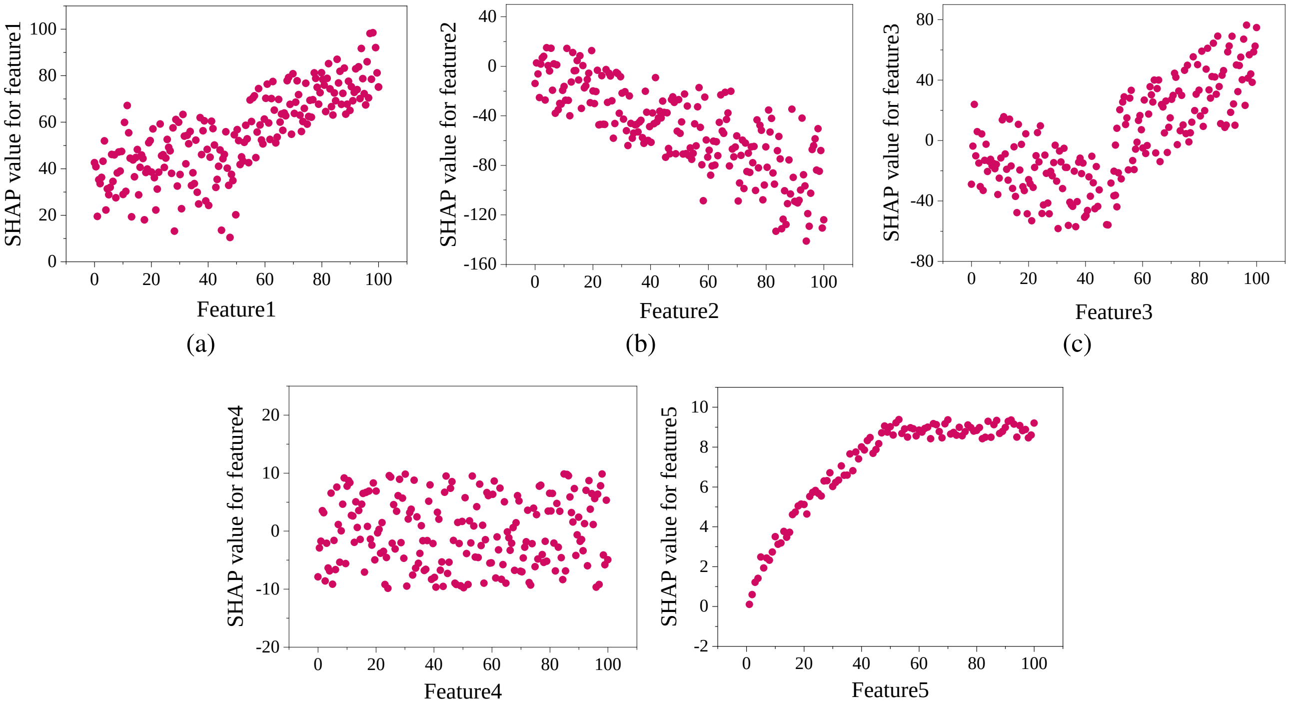
Input feature	RF	MLP	XGBoost
Reservoir temperature	***N***	***M***	***M***
Pore size	***I***	***I***	***I***
Volatile component content	***N***	***P***	***N***
Intermediate hydrocarbon component content	***M***	***N***	***M***
Ratio of volatile to intermediate component content	***N***	***P***	***N***
Molecular weight of C_7+ components	***P***	***P***	***C***

Notes: ***P*** and ***N*** denote the influencing factor taking the positive and negative interpretability for MMP, respectively; ***M*** denotes the influencing factor taking the positive interpretability when there is an appropriate value; ***C*** denotes the influencing factor taking the confusing interpretability; ***I*** denotes the influencing factor taking the positive interpretability within a range, beyond which there is no effect.

values varies from 0.008 to 0.024, depending on the specific values of other features at those points, as illustrated in Fig. 6(c).

Analyzing the overall trends between the nodes in the dependence plot allows to readily establish the relationship between feature values and MMP. Five distinct patterns of trends among points in the graph are identified, which follows data reported in the literature (Min et al., 2023). These patterns are illustrated in Fig. 9. A positive correlation between the feature value and the corresponding SHAP value is depicted in Fig. 9(a), denoted as ***P***, while a negative correlation is observed in Fig. 9(b), denoted as ***N***. As can be seen in Fig. 9(c), as the feature value increases, the corresponding SHAP value exhibits an initial decrease followed by an increase, representing a changing trend denoted as ***M***. Fig. 9(d) reveals no distinct correlation or inconclusive relationship between these two variables, denoted as ***C***. Fig. 9(e) illustrates a pattern denoted as ***I***, where the SHAP values initially increase and then remain stable as the feature value increases.

Next, the trends shown in Figs. 6-8 are analyzed according to the aforementioned recognition patterns. Fig. 7(f) demonstrates that as $MW_{C_{7+}}$ increases, there is a corresponding increase in the SHAP value, which agrees with pattern ***P***. As can be seen in Fig. 7(d), the increase in x_{int} results in a decrease in the SHAP value, which aligns with pattern ***N***. With an increase in T_R in Fig. 8(a), the SHAP value initially decreases and subsequently increases, following pattern ***M***. Fig. 8(f) is classified as pattern ***C***, as the SHAP value exhibits multiple inflection points with an increase in the $MW_{C_{7+}}$ value, and the relationship is relatively inconclusive. In addition, as shown in Fig. 6(b), an increase in r makes the SHAP value initially increase before reaching a plateau, matching pattern ***I***. The results of all these features in the three models are summarized in Table 4.



(a) (b) (c)

(d) (e)

Fig. 9. Schematic diagram illustrating the five trend patterns between points in the SHAP dependence plot (Figs. 6-8). (a) $\boldsymbol{P}$, (b) $\boldsymbol{N}$, (c) $\boldsymbol{M}$, (d) $\boldsymbol{C}$ and (e) $\boldsymbol{I}$.

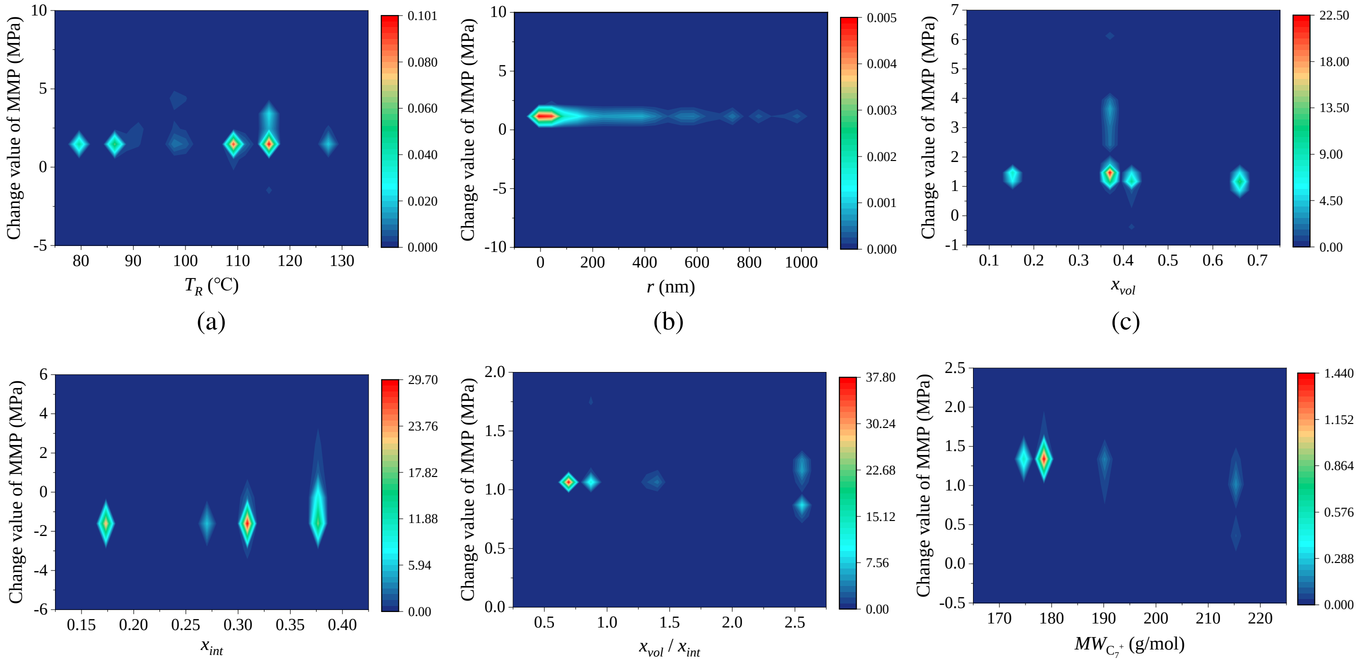
Table 4 reveals significant differences among the three models in terms of interpretability, with some contrasting trends. As reported previously (Liao et al., 2014), with increasing temperature, the thermal motion of CO_2 molecules intensifies, making them less prone to blend with crude oil, thereby increasing the MMP. In addition, the volatile components in crude oil are more likely to evaporate into the CO_2 phase during this process, leading to an overall increase in gas volume and MMP. In contrast, the presence of intermediate components (C_2-C_6), due to their molecular similarities with CO_2, leads to a significant reduction in MMP, following the principles of similarity dissolution (Liang et al., 2017; Shen et al., 2023). Laboratory experimental evidence confirms that the MMP of crude oil in nanoscale pores is somewhat lower compared to that in bulk phase, with a rapid decrease upon pore size reduction (Bo et al., 2021). In summary, MMP is prone to enlarge with increasing T_R, r, x_{vol}, x_{vol}/x_{int} and $MW_{C_{7+}}$, and it is prone to decline with increasing x_{int}.

When comparing these outcomes with those presented in Table 4, the interpretability result obtained from the MLP model (Fig. 7) appears highly consistent with the theoretical knowledge of the CO_2-EOR process. The trends exhibited by the various features, except for the non-monotonic changes in T_R, are congruent with the actual physical laws, suggesting that the MLP model provides the most reasonable interpretability of data among the models.

After conducting a comprehensive analysis of the SHAP interpretation patterns of these three models, the trends exhibited by r in each model are the least controversial. The observed trends closely agree between the RF model in Fig. 6(b), the MLP model in Fig. 7(b), and the XGBoost model in Fig. 8(b). These data indicate that MMP increases initially and then remains unchanged as r increases. Notably, the value of r corresponding to the stationary point is approximately 75 nm for the RF model, 75 nm for the MLP model, and 100 nm for XGBoost, Therefore, 75 nm is maintained as the threshold representing the inflection point where the influence of nano-confinement on MMP undergoes a significant change. This implies that the miscible pressure between crude oil and CO_2 in unconventional reservoirs, which are predominantly characterized by nanopores, is lower than that in conventional reservoirs. This correlation is particularly strong when the pore size in the reservoir is less than 75 nm. Bo et al. (2021) reached similar conclusions through experiments, as did Wei et al. (2021) using modified state equation calculations. This discovery firmly establishes the pore size for the bulk phase MMP theory at 75 nm, offering substantial evidence to support the extension of the validity of classical theory to the nanoscale.

4.3 Sensitivity analysis of the MLP model

Based on the SHAP interpretation results, significant differences are found in the correlations between the input parameters and output variables among the RF, MLP and XGBoost models. Subsequently, the MLP model is chosen as the preferred prediction model by comparing it with petrophysical laws, showing much better balancing performance and interpretability. To verify the accuracy of the SHAP



(a) (b) (c)

(d) (e) (f)

Fig. 10. Sensitivity analysis results of the MLP model for various input parameters. (a) T_R, (b) r, (c) x_{vol}, (d) x_{int}, (e) x_{vol}/x_{int}, and (f) $MW_{C_{7+}}$.

interpretability results of the MLP model, a sensitivity analysis is conducted to establish the crucial relationship between inputs and outputs.

For the above purpose, a single-factor sensitivity analysis method (Chen et al., 2022; Shen et al., 2023) is employed. This involves altering only one input feature at a time (One-at-a-time) while keeping the values of the other five features constant. For instance, the T_R value in each sample is increased by 5% compared to its original value, while the values of the remaining five features remain unaltered. The new prediction values of MMP are recorded, and a kernel density plot is generated to visualize the relationship between the raw values of T_R and the discrepancy between the newly predicted values and their corresponding raw values, as shown in Fig. 10(a). The sensitivity analysis results of the other features are presented in order in Fig. 10.

When the change value of MMP exceeds zero, this suggests that an increase in the feature value corresponds to a rise in MMP, demonstrating a positive correlation, as depicted in Fig. 9. The transition of the color bar from blue to red represents the degree of data point aggregation in the region, facilitating the observation of the main relationship between the parameters of the coordinate axes. Accordingly, positive differences in MMP are observed as increases in the features T_R, r, x_{vol}, x_{vol}/x_{int} and $MW_{C_{7+}}$, as seen in Figs. 10(a)-10(c) and 10(e)-10(f), respectively, indicating a robust positive relationship between these features and MMP. Most points in the region with a difference greater than zero indicate that x_{int} experiences two changes and maintains a significantly negative relationship with MMP, as demonstrated in Fig. 10(d). In summary, the sensitivity analysis of all input features reveals a consistent correlation with the SHAP interpretation results, as presented in Fig. 7, excluding the features T_R and x_{int}. This discrepancy can be attributed to the fact that single-factor sensitivity analysis examines one factor at a time; therefore, it cannot fully capture the interaction between input factors. In contrast, the results explained by SHAP encompass the interaction between model inputs.

5. Conclusions

In this study, SHAP-based interpretability machine learning models are developed to evaluate the impact of nanoconfinement on MMP in CO_2-EOR processes. Four models are established with inputs comprising three sets of parameters, including reservoir temperature, pore size, and crude oil composition. Furthermore, SHAP is utilized to enhance the interpretability and rationality of prediction models, followed by validation by sensitivity analysis. As the main contributions of this work, the following conclusions can be generally drawn:

1) RF, MLP, SVR and XGBoost models are established to predict the MMP based on a dataset consisting of 348 cases. Various metrics, such as MAE, MSE, MAPE and R^2, are employed to evaluate the performance of these models, yielding a performance ranking of XGBoost > MLP > RF > SVR.
2) The SHAP interpretability results reveal significant disparities among the XGBoost, MLP and RF models in terms of feature importance ranking and the impact pattern of each feature on MMP. The MLP model exhibits

superior suitability for the dataset. This model accurately captures the correlation between features and MMP, with a feature importance ranking of $MW_{\mathrm{C}_{7+}} > r > x_{vol} > x_{vol}/x_{int} > x_{int} > T_R$.

3) In nano-confined spaces, the influence of reservoir temperature on MMP is decreased. This diminishing impact is reflected in the average absolute SHAP value, which is measured to be 0.05.
4) A distinct relationship between MMP and r is observed, indicating that MMP undergoes a significant decrease as r falls below a critical threshold. The critical values of r for the RF, MLP and XGBoost models are determined to be 75, 75 and 100 nm, respectively. These findings suggest that the threshold of r is around 75 nm, signifying notable changes in the MMP of the CO_2-oil system below this pore size.

Acknowledgements

The authors gratefully acknowledge the financial support of the National Key Research and Development Program of China (No. 2023YFE0120700), the National Natural Science Foundation of China (Nos. 52274041 and 52304023), the Distinguished Young Sichuan Science Scholars (No. 2023NSFSC1954).

Conflict of interest

The authors declare no competing interest.

References

Alharthy, N., Teklu, T., Kazermi, H., et al. Enhanced oil recovery in liquid-rich shale reservoirs: Laboratory to field. SPE Reservoir Evalation & Engineering, 2018, 21(1): 137-159.

Bo, B., Feng, J., Qiu, J., et al. Direct measurement of minimum miscibility pressure of decane and CO_2 in nanoconfined channels. ACS Omega, 2021, 6(1): 943-953.

Breiman, L. Random forests. Machine Learning, 2001, 45: 5-32.

Cai, J., Xu, K., Zhu, Y., et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Applied Energy, 2020, 262: 114566.

Chemmakh, A., Merzoug, A., Ouadi, H., et al. Machine learning predictive models to estimate the minimum miscibility pressure of CO_2-oil system. Paper SPE 207865 Presented at Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 15-18, November, 2021.

Chen, G., Fu, K., Liang, Z., et al. The genetic algorithm based back propagation neural network for MMP prediction in CO_2-EOR process. Fuel, 2014, 126: 202-212.

Chen, H., Zhang, C., Yu, H., et al. Application of machine learning to evaluating and remediating models for energy and environmental engineering. Applied Energy, 2022, 320: 119286.

Dargahi-Zarandi, A., Hemmati-Sarapardeh, A., Shateri, M., et al. Modeling minimum miscibility pressure of pure/impure CO_2-crude oil systems using adaptive boosting support vector regression: Application to gas injection processes. Journal of Petroleum Science and Engineering, 2020, 184: 106499.

Dehghani, S. A. M., Sefti, M. V., Ameri, A., et al. Minimum miscibility pressure prediction based on a hybrid neural genetic algorithm. Chemical Engineering Research and Design, 2008, 86(2): 173-185.

Feng, Q., Xu, S., Xing, X., et al. Advances and challenges in shale oil development: A critical review. Advances in Geo-Energy Research, 2020, 4(4): 406-418.

Ge, D., Cheng, H., Cai, M., et al. A new predictive method for CO_2-oil minimum miscibility pressure. Geofluids, 2021, 2021: 8868592.

Hamada, Y., Koga, K., Tanaka, H. Phase equilibria and interfacial tension of fluids confined in narrow pores. The Journal of Chemical Physics, 2007, 127: 084908.

Hao, M., Liao, S., Yu, G., et al. Performance optimization of CO_2 huff-n-puff for multifractured horizontal wells in tight oil reservoirs. Geofluids, 2020, 2020: 8840384.

Hawthorne, S. B., Miller, D. J., Jin, L., et al. Rapid and simple capillary-rise/vanishing interfacial tension method to determine crude oil minimum miscibility pressure: pure and mixed CO_2, methane and ethane. Energy & Fuels, 2016, 30(8): 6365-6372.

Huang, C., Tian, L., Wu, J., et al. Prediction of minimum miscibility pressure (MMP) of the crude oil-CO_2 systems within a unified and consistent machine learning framework. Fuel, 2023, 337: 127194.

Li, L., Qiao, J., Yu, G., et al. Interpretable tree-based ensemble model for predicting beach water quality. Water Research, 2022, 211: 118078.

Li, S., Luo, P. Experimental and simulation determination of minimum miscibility pressure for a Bakken tight oil and different injection gases. Petroleum, 2017, 3(1): 79-86.

Liang, M., Yuan, H., Yang, Y., et al. Research progress on miscible gas displacement and determination of minimum miscibility pressure. Journal of Southwest Petroleum University (Science & Technology Edition), 2017, 39(5): 101-112. (in Chinese)

Liao, C., Liao, X., Chen, J., et al. Correlations of minimum miscibility pressure for pure and impure CO_2 in low permeability oil reservoir. Journal of the Energy Institute, 2014, 87(3): 208-214.

Lu, H., Xu, Y., Duan, C., et al. Experimental study on capillary imbibition of shale oil in nanochannels. Energy & Fuels, 2022, 36(10): 5267-5275.

Lundberg, S. M., Lee, S. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 2017, 30: 4768-4777.

Lv, Q., Zheng, R., Guo, X., et al. Modelling minimum miscibility pressure of CO_2-crude oil systems using deep learning, tree-based, and thermodynamic models: Application to CO_2 sequestration and enhanced oil recovery. Separation and Purification Technology, 2023, 310: 123086.

Ma, X., Hou, M., Zhan, J., et al. Interpretable predictive modeling of tight gas well productivity with SHAP and LIME techniques. Energies, 2023, 16: 3653.

Min, C., Wen, G., Gou, L., et al. Interpretability and causal discovery of the machine learning models to predict the production of CBM wells after hydraulic fracturing. Energy, 2023, 285: 129211.

Molnar, C. Interpretable machine learning, 2020.

Nguyen, P., Mohaddes, D., Riordon, J., et al. Fast fluorescence-based microfluidic method for measuring minimum miscibility pressure of CO_2 in crude oils. Analytical Chemistry, 2015, 87(6): 3160-3164.

Novosad, Z., Sibbald, L. R., Costain, T. G. Design of miscible solvents for a rich gas drive-comparison of slim tube and rising bubble tests. Journal of Canadian Petroleum Technology, 1990, 29(1): 37-42.

Safaei, A., Riazi, M., Shariat, S. A novel experimental-theoretical method to improve MMP estimation using VIT technique. Journal of Petroleum Science and Engineering, 2023, 220: 111182.

Sambo, C., Liu, N., Shaibu, R., et al. A technical review of CO_2 for enhanced oil recovery in unconventional oil reservoirs. Geoenergy Science and Engineering, 2023, 221: 111185.

Shen, B., Yang, S., Gao, X., et al. Interpretable knowledge-guided framework for modeling minimum miscible pressure of CO_2-oil system in CO_2-EOR projects. Engineering Applications of Artificial Intelligence, 2023, 118: 105687.

Shokir, E. M. E. CO_2-oil minimum miscibility pressure model for impure and pure CO_2 streams. Journal of Petroleum Science and Engineering, 2007, 58(1-2): 173-185.

Sun, H., Li, H. Minimum miscibility pressure determination in confined nanopores considering pore size distribution of tight/shale formations. Fuel, 2021, 286: 119450.

Teklu, T. W., Alharthy, N., Kazemi, H., et al. Vanishing interfacial tension algorithm for MMP determination in unconventional reservoirs. Paper SPE 169517 presented at SPE Western North American and Rocky Mountain Joint Meeting, Denver, Colorado, 17-18 April, 2014a.

Teklu, T. W., Alharthy, N., Kazemi, H., et al. Phase behavior and minimum miscibility pressure in nanopores. SPE Reservior Evalation & Engineering, 2014b, 17(3): 396-403.

Tovar, F. D., Barrufet, M. A., Schechter, D. S. Enhanced oil recovery in the Wolfcamp shale by carbon dioxide or nitrogen injection: An experimental investigation. SPE Journal, 2021, 26(1): 515-537.

Wang, S., Ma, M., Chen, S. Application of PC-SAFT equation of state for CO_2 minimum miscibility pressure prediction in nanopores. Paper SPE 179535 Presented at SPE Improved Oil Recovery Conference, Tulsa, Oklahoma, USA, 11-13 April, 2016.

Wei, B., Liu, J., Zhang, X., et al. Advances of enhanced oil recovery method and theory in tight reservoirs. Journal of Southwest Petroleum University (Science & Technology Edition), 2021, 43(1): 91-102. (in Chinese)

Wei, B., Zhong, M., Wang, L., et al. Oil recovery dynamics of natural gas huff 'n' puff in unconventional oil reservoirs considering the effects of nanopore confinement and its proportion: A mechanistic study. SPE Reservoir Evalation & Engineering, 2022a, 25(4): 667-683.

Wei, B., Zhong, M., Zhao, J., et al. Prediction method for the minimum miscibility pressure of crude oil and natural gas in micro-nano confined space. Acta Petrolei Sinica, 2022b, 43(11): 1604-1613. (in Chinese)

Yang, L., Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 2020, 415: 295-316.

You, J., Ampomah, W., Sun, Q., et al. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO_2-EOR project. Journal of Cleaner Production, 2020, 260: 120866.

ZareNezhad, B. A new correlation for predicting the minimum miscibility pressure regarding the enhanced oil recovery processes in the petroleum industry. Petroleum Science and Technology, 2016, 34 (1): 56-62.

Zhang, K., Jia, N., Zeng, F., et al. A new diminishing interface method for determining the minimum miscibility pressures of light oil–CO_2 systems in bulk phase and nanopores. Energy & Fuels, 2017, 31(11): 12021-12034.

Zhang, K., Jia, N., Zeng, F., et al. A review of experimental methods for determining the oil-gas minimum miscibility pressures. Journal of Petroleum Science and Engineering, 2019, 183: 106366.

Zhao, H., Fang, Z. Improved multiple-mixing-cell method for accelerating minimum miscibility pressure calculations. SPE Journal. 2020, 25(4): 1681-1696.

Zhao, J., Jin, L., Azzolina N. A., et al. Investigating enhanced oil recovery in unconventional reservoirs based on field case review, laboratory, and simulation studies. Energy & Fuels, 2022, 36(24): 14771-14788.