ARTICLE

Models for Predicting the Minimum Miscibility Pressure (MMP) of CO_2-Oil in Ultra-Deep Oil Reservoirs Based on Machine Learning

Kun Li[1], Tianfu Li[2,*], Xiuwei Wang[1], Qingchun Meng[1], Zhenjie Wang[1], Jinyang Luo[1,2], Zhaohui Wang[1] and Yuedong Yao[2]

[1]Huabei Oilfield Company, PetroChina, Renqiu, 062552, China
[2]College of Petroleum Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
*Corresponding Author: Tianfu Li. Email: 2022210392@student.cup.edu.cn

ABSTRACT: CO_2 flooding for enhanced oil recovery (EOR) not only enables underground carbon storage but also plays a critical role in tertiary oil recovery. However, its displacement efficiency is constrained by whether CO_2 and crude oil achieve miscibility, necessitating precise prediction of the minimum miscibility pressure (MMP) for CO_2-oil systems. Traditional methods, such as experimental measurements and empirical correlations, face challenges including time-consuming procedures and limited applicability. In contrast, artificial intelligence (AI) algorithms have emerged as superior alternatives due to their efficiency, broad applicability, and high prediction accuracy. This study employs four AI algorithms—Random Forest Regression (RFR), Genetic Algorithm Based Back Propagation Artificial Neural Network (GA-BPNN), Support Vector Regression (SVR), and Gaussian Process Regression (GPR)—to establish predictive models for CO_2-oil MMP. A comprehensive database comprising 151 data entries was utilized for model development. The performance of these models was rigorously evaluated using five distinct statistical metrics and visualized comparisons. Validation results confirm their accuracy. Field applications demonstrate that all four models are effective for predicting MMP in ultra-deep reservoirs (burial depth >5000 m) with complex crude oil compositions. Among them, the RFR and GA-BPNN models outperform SVR and GPR, achieving root mean square errors (RMSE) of 0.33% and 2.23%, and average absolute percentage relative errors (AAPRE) of 0.01% and 0.04%, respectively. Sensitivity analysis of MMP-influencing factors reveals that reservoir temperature (T_R) exerts the most significant impact on MMP, while Xint (mole fraction of intermediate oil components, including C_2-C_4, CO_2, and H_2S) exhibits the least influence.

KEYWORDS: MMP; random forest regression; genetic algorithm based back propagation artificial neural network; support vector regression; gaussian process regression

1 Introduction

In recent years, the rapid growth of the global economy has driven a steady rise in demand for oil resources across nations [1]. However, conventional oil recovery methods often exhibit low extraction efficiency, prompting heightened interest in technologies capable of enhancing hydrocarbon recovery [2]. Among these, CO_2 flooding stands out as a highly effective and economically viable enhanced oil recovery (EOR) technique. Its significance stems not only from its ability to achieve carbon capture, utilization, and storage (CCUS)—thereby contributing to the mitigation of the greenhouse effect—but also from its cost-effectiveness and operational efficiency [3–5]. The mechanisms by which CO_2 enhances oil recovery primarily include viscosity reduction through dissolution into crude oil, interfacial tension reduction, and most critically, miscible or immiscible displacement achieved via compositional interactions with the crude

oil. When CO_2 is miscible with crude oil, the oil recovery rate will be greatly improved, and theoretically, the displacement efficiency can reach 100% [6].

Related experimental studies and field practices [7,8] have demonstrated that oil recovery rates under miscible and immiscible conditions exhibit marked disparities. Therefore, accurate determination of the minimum miscible pressure (MMP)—the critical threshold delineating miscible and immiscible regimes—is crucial for improving oil recovery efficiency. At present, the main methods for determining the minimum mixed phase pressure include experimental methods, empirical formulas, specific correlations, and intelligent algorithms.

The experimental determination of the MMP is widely adopted in CO_2 flooding projects due to its high accuracy and precision. Currently, commonly used experimental methods for determining the MMP include capillary tests, the bubble apparatus method, and the interfacial tension elimination method [9–11]. Although these experimental techniques exhibit high reliability, they are associated with limitations such as time-consuming procedures, high operational costs, and complex operating procedures [12]. Given these constraints, developing rapid and inexpensive method methods for CO_2-oil MMP estimation is imperative. Table 1 summarizes prior studies investigating MMP under diverse conditions using experimental and empirical approaches.

Due to the limitations of experimental methods and the need for rapid prediction of the MMP, various empirical approaches have been proposed, demonstrating significant applicability and gradually emerging as alternatives to traditional experimental techniques. For instance, Holm et al. [13] introduced the plate method for MMP estimation, which relies on analyzing changes in crude oil density within reservoir formations. However, this method is restricted to reservoirs with temperatures between 32.2°C and 82.2°C and formation pressures ranging from 9.65 to 22 MPa. Mansour et al. [14] developed an empirical formula that integrates T_C, API, MWC_{5+}, and MWC_{7+} to predict MMP for live oil systems. Similarly, Chen et al. [15] proposed a multi-factor empirical model accounting for ten variables influencing MMP, which exhibits superior accuracy compared to earlier formulations. Despite these advancements, empirical formulas often suffer from narrow applicability, being valid only for reservoirs with specific characteristics and failing to generalize across diverse geological settings. Consequently, there is a pressing need to develop more accurate and universally adaptable methods for CO_2-oil MMP prediction.

Table 1: Previous experimental and empirical studies on MMP

Author	Method	Independent parameters	Temperature (°C)	Remarks
Holm et al. (1974) [13]	Experimental/ plot method	T_R, MWC_{5+}	32.2–82.2	-Temperature ramg limit (32.2°C to 82.2°C) -Pressure range limit (9.65 MPa to 22 MPa) $-180 < MWC_{5+} < 240$
Yellig and Metcalfc (1980) [16]	Empirical methods	T_R	35–88.9	Temperature rang limit (35°C to 88.9°C)
Glasø (1985) [17]	Empirical methods	T_R, MWC_{7+}, C_2-C_6	54.4–105.5	Influenced by C_2-C_6 components
Alston et al. (1985) [18]	Empirical methods	T_R, MWC_{5+}, X_{vol}, X_{int}	32.2–117.2	Limited by bubble point pressure

(Continued)

Table 1 (continued)

Author	Method	Independent parameters	Temperature (°C)	Remarks
Abdurrahman et al. (2015) [19]	Experimental	T_R, P_r,	60–66	Experimental operation is complex
Mansour et al. (2018) [14]	Empirical methods	Tc, API, MWC_{5+}, MWC_{7+}	32.2–118.9	Limited by oil samples
Chen et al. (2020) [15]	Improved empirical methods	T_R, MWC_{7+}, dataset of pure and impure CO_2	30.5–147.8	Empirical corelation
Ge et al. (2021) [20]	Empirical methods	T_R, reservoir oil composition	71.6–108.4	Limited by oil samples
Yu et al. (2024) [7]	Empirical methods	T_R, composition of injected gas, reservoir oil composition	40.6–122.5	Limited by crude oil composition and data sets

Recent advancements in artificial intelligence (AI), particularly its superior machine learning (ML) capabilities and robust data analytics proficiency, have catalyzed the increasing adoption of ML techniques to address complex challenges in oilfield production. Key applications spanning well-test interpretation, log analysis, reservoir classification, production performance forecasting, and gas solubility prediction in crude oil have demonstrated the efficacy of ML-driven approaches, showcasing their adaptability and precision in handling datasets [21–24]. Given these successes, researchers are now actively exploring ML-based frameworks to predict the CO_2-oil MMP. Huang et al. [25] developed a three-layer back propagation artificial neural network(BPNN) model and comprehensively considered the effects of Molecular weight of C_{5+} fraction, reservoir temperature, and volatility to predict pure and implied CO_2-oil MMP. Shokrollahi et al. [26] developed a Least Squares Support Vector Machine model (LSSVM) that comprehensively considers the effects of reservoir temperature, MWC_{5+}, and the contents of y_{CO2}, y_{H2S}, y_{C1}, and N_2 in the injected gas to predict pure and impure MMP. Kamari et al. [27] developed a Genetic expression Programming (GEP) model to predict pure and impure CO_2-oil MMP in live reservoir oil systems during CO_2 flooding. Table 2 provides a comparative summary of predictive models for the CO_2-oil MMP developed in previous studies, detailing their input variables and methodological constraints.

Table 2: A summary of Previous studies on models for prediction CO_2-oil MMP

Author	Method	Independent parameters	Temperature (°C)	Remarks
Huang et al. (2003) [25]	BPNN	TR, MW_{C5+}, X_{vol}, X_{int}	16.7–86.7	Limited by data rang, possibility of overfitting
Emera et al. (2005) [28]	GA	TR, MW_{C5+}, X_{vol}, X_{int}	Not mentioned	Limited by bubble point pressure, narrow range of application

Table 2 (continued)

Author	Method	Independent parameters	Temperature (°C)	Remarks
Birang et al. (2007) [29]	ANN	TR, MW_{C2-C5} and MW_{C7+} component, X_{vol}, X_{C1-C5}, composition of injected gas	55.5–106.5	Limited by hidden neurons, missing sample data points
Shokrollahi et al. (2013) [26]	LSSVM	T_R, y_{CO2}, y_{H2S}, y_{C1}, MW_{C5+}, N_2	31.7–117.8	Difficulty in selecting hyper-parameters of LSSVM method
Kamari et al. (2015) [27]	GEP	T_R, MW_{C5+}, X_{vol}/X_{int}, composition of injected gas	31.7–117.8	The model has a low R^2 (0.63) on the validation set
Bian et al. (2016) [30]	SVR-GA	T_R, X_{int}, X_{vol} MWC_{5+}	31.7–117.8	The sample data size is not mention
Saeedi Dehaghani et al. (2020) [31]	ANN, POS-ANN, ANN-TLBO	T_R, MW_{C5+}, X_{vol}, X_{int}, composition of injected gas	66.27–118.3	Lack of discussion on sensitivity analysis
Nait Amar et al. (2020) [32]	SVM	T_R, C_2-C_6, MW_{C5+}, X_{vol}/X_{int}	32.2–136.7	The sample datapoints are not mention
Ghiasi et al. (2021) [33]	hybrid-ANFIS and AdaBoost-CART	T_R, T_c, composition of injected gas, MW_{C5+}, X_{vol}/X_{int}	32.2–136.7	The sample datapoints are not mention and limited by membership functions setting
Al-Khafaji et al. (2023) [34]	DT, SVR, KNN, RF, MLR	MW_{C5+}, T_R, y_{H2S}	31.7–136.7	Limited by model training time

The results of these machine learning models for predicting MMP indicate that reservoir temperature and crude oil composition have significantly impact on MMP [21,23,24,26,27]. For ultra-deep oil reservoirs buried deeper than 5000 m, characteristics such as high reservoir temperature and complex crude oil composition are present. Existing prediction models exhibit low accuracy and local overfitting problems when predicting MMP in ultra-deep oil reservoirs. In order to quickly and accurately predict MMP in ultra-deep oil reservoirs, we developed multiple models with ten input parameters (i.e., T_R, X_{vol}, X_{int}, X_{C5-C6}, MWC_{7+}, mole fraction of CO_2 in solution (y_{CO2}), mole fraction of C_1 in the CO_2 injection gas (y_{C1}), mole fraction of N_2 in the injection gas (y_{N2}), mole fraction of H_2S in the injection gas (y_{H2S}), and mole fraction of C_2-C_4 in the injection gas (y_{C2-C4})) to predict MMP in both pure and impure CO_2 injection cases in ultra-deep oil reservoirs. Based on the above description, multiple models were established and optimized, and suitable model for CO_2 MMP in ultra-deep oil reservoirs was selected by comparing the prediction results of different models. Grid search was employed to tune and optimize the hyperparameters of these models, which predict the CO_2-oil MMP in ultra-deep oil reservoirs, thereby ensuring their predictive accuracy. In addition, a sensitivity analysis was systematically conducted to quantify the relative

contributions of individual factors affecting MMP and to elucidate the functional dependencies of MMP on these critical variables.

2 Minimum Miscibility Pressure (MMP)

During the process of injecting CO_2 into oil reservoirs for enhanced oil recovery, interactions between gas, oil, and water phases occur in the rock layers, resulting in interphase component transfer, phase transitions, and other complex phase behaviors. CO_2 undergoes multiple contact displacements with crude oil under reservoir temperature and pressure, achieving either miscible or immiscible flooding. The CO_2 mixed-phase drive refers to the formation of a stable mixed-phase zone at the leading edge between CO_2 and crude oil when the injection pressure is higher than a certain value and the disappearance of the oil-gas interface, resulting in zero interfacial tension in porous media. This pressure is called the minimum miscibility pressure.

It is widely recognized that temperature, crude oil components, and the composition of injected gas significantly influence the MMP. To date, researchers [27,30,31,33] have conducted comprehensive studies on the influencing factors affecting MMP. Reservoir temperature plays a crucial role in the MMP during CO_2 flooding. As the reservoir temperature increases, the amount of CO_2 dissolved in crude oil decreases, and the supercritical nature of CO_2 diminishes, particularly in terms of density. This reduction seriously affects the ability of CO_2 extraction to extract light hydrocarbons. Currently, most research indicates a linear positive correlation between CO_2-oil MMP and reservoir temperature. Alston et al. [18] studied the crude oil components, dividing them into three parts—X_{vol}, X_{int}, and MWC_{5+}—have an impact on MMP, but MW_{C5+} has a more significant effect on CO_2-oil MMP. Additionally, injecting gas components also has a significant impact on MMP. Sebastian et al. [35] found that injecting CO_2 gas containing C_1, N_2, H_2S, and light component hydrocarbons (C_2-C_4) can have an impact on CO_2-oil MMP.

Consequently, ten factors (T_R, x_{int}, x_{vol}, x_{C5-C6}, MW_{C7+}, y_{C1}, y_{N2}, y_{CO2}, y_{C2-C4}, and y_{H2S}) that significantly affect MMP were selected as the input variables and imported to design multiple models for MMP simulation. To evaluate the correlation between these characteristic variables and their influence on the MMP, this study employs the partial correlation analysis to detect any redundant information among the variables and ensure their independence. The results are shown in Fig. 1. It can be seen from the Fig. 1 that the r values between any two feature variables and MMP differ, indicating that the ten features are independent of each other and there is no significant collinearity problem.

3 Basic Description of Models and Experimental Data

3.1 GA-BPNN Model

Artificial Neural Networks (ANNs) are computational models designed to simulate the structure and function of neurons in the human brain. Similar to the way the human brain works, artificial neural networks can automatically process multiple input variables in parallel, passing information from the input layer to the output layer through the learning process [36,37]. In addition, ANNs also have the ability to reveal potential linear or nonlinear relationships between input and output data. A complete neural network consists of an input layer, an output layer, and at least one hidden layer. The basic structural unit of a neural network is the neuron, which can receive, process, and output signals. It is a nonlinear element with multiple inputs and a single output. A large number of neurons are interconnected via connection weights, bias values, and transfer functions between neurons in each layer, forming a neural network with robust information processing capabilities. A back propagation neural network (BPNN) is a type of feedforward neural network that incorporates a backpropagation learning algorithm. Due to its adjustable architecture, capability for

forward signal propagation, and mechanism of error backpropagation, the BPNN has emerged as the most widely implemented neural network paradigm. The schematic architecture of a standard three-layer BPNN is depicted in Fig. 2. However, ANN models are prone to challenges, including overfitting, suboptimal parameter selection, and computational inefficiency during training. To mitigate these limitations, a hybrid framework integrating a Genetic Algorithm (GA) and BPNN is employed to optimize network weights and biases, thereby enhancing prediction accuracy [38].
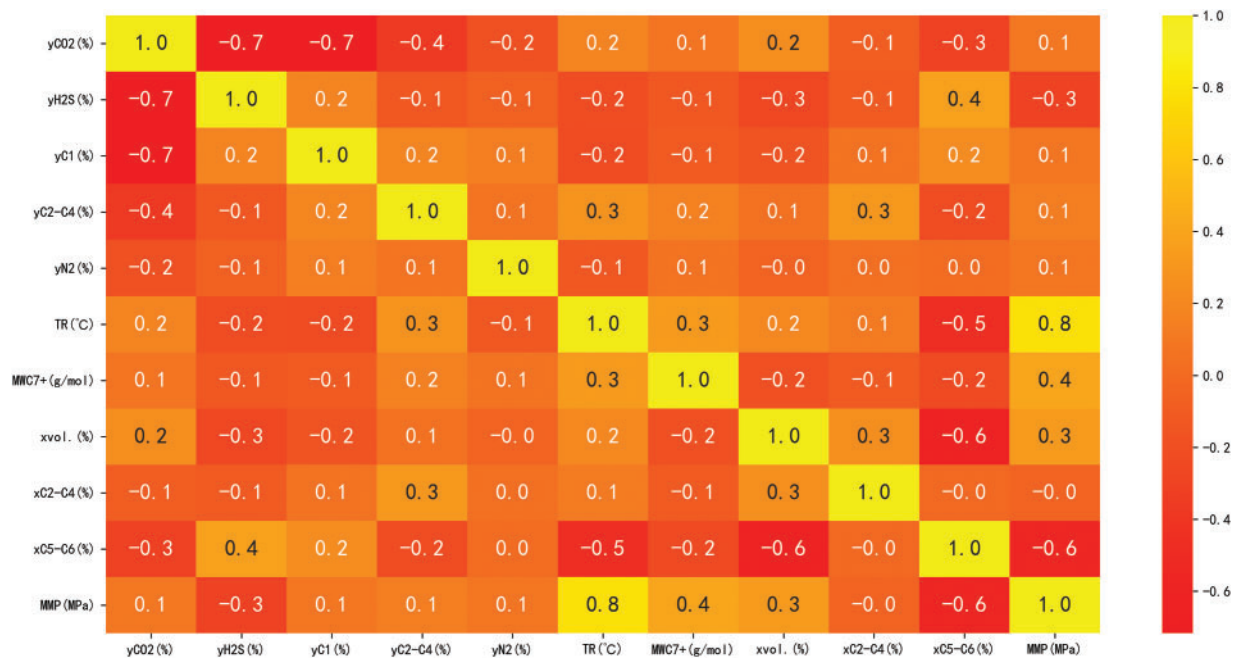

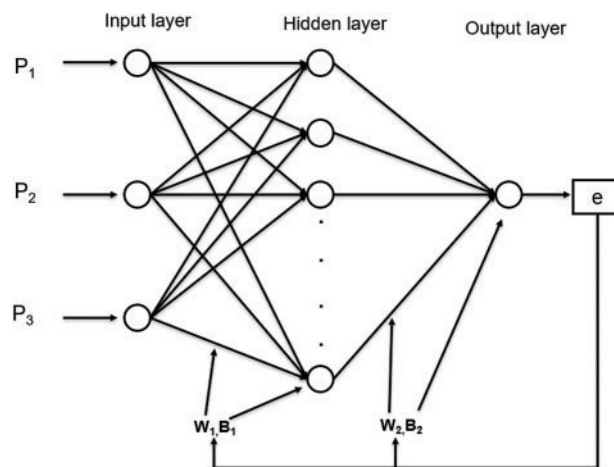**Figure 1:** Partial correlation analysis results heatmap

Figure 2: Typical three-layer backpropagation neural network model structure diagram

The GA, first proposed by American scholar John Holland, it a global optimization algorithm that simulates the process of biological evolution.GA mimics principles of natural selection, crossover, and mutation

to enhance individual fitness within a population. Its workflow comprises four stages: parameter encoding, population initialization, fitness evaluation, and iterative genetic operations [39,40]. Genetic operations include selection, crossover, and mutation. The selection operation can inherit excellent chromosomes to the next generation or generate new individuals through crossover operation and pass them on to the next generation. By adopting an optimal preservation method, the fitness values of individuals within the same population are sorted in descending order, with priority given to selecting individuals with higher fitness values for inheritance to the next generation. The crossover operation is the process of selecting chromosomes from a population and recombining them into new chromosomes through crossover operations. This operation can effectively inherit excellent genes from the population. The mutation operation involves using other genes in an individual to replace specific genes, which can enhance local search ability and effectively preserve population diversity. In general, selection operations ensure that the population evolves toward higher fitness, crossover operations accelerate the speed of genetic algorithm searching for optimal solutions, and mutation operations help prevent the algorithm from falling into local optima too early. Together, the three genetic operations of selection, crossover, and mutation enable GA to efficiently search the solution space and gradually approach the optimal solution.

Optimizing the BPNN model through GA and developing the GA-BPNN hybrid model. The specific steps are as follows:

(1) Initialization: Define the BPNN architecture and initialize GA parameters and population.

(2) Fitness Evaluation: Train the BPNN on the dataset and calculate individual fitness based on prediction accuracy.

(3) GA Execution: Perform selection, crossover, and mutation operations within the GA.

(4) Population Update: Replace old individuals with newly generated individuals to form a new population.

(5) Termination Check: Determine whether the termination condition is satisfied. If the termination condition is not satisfied, return to step 2 to continue the iteration.

(6) Model Finalization: Deploy the BPNN with optimized weights and biases from the highest-fitness individual.

3.2 Random Forest Regression

RFR is an ensemble algorithm composed of multiple decision trees based on the bootstrap aggregating method; it was initially designed for classification tasks. As research progressed, the RFR model was further applied to predict CO_2-oil MMP. For instance, Al Khafaji et al. [34] used the RFR model to predict MMP. The basic principles of RFR can be summarized as four steps: sample selection, feature selection, decision tree construction, and prediction. The number of trees in a random forest is represented as B, and the predicted result of the b-th tree is represented as $F_b(x)$. Therefore, the prediction formula for RFR can be expressed as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} f_b(x) \tag{1}$$

The steps for establishing a random forest regression model are as follows:

(1) Data Standardization: Standardize all data to ensure that their values are between –1 and 1, in order to obtain accurate prediction results and accelerate training speed.

(2) Import Data and Select Parameters: Import all data and output the dataset to the model for model training and testing.

(3) Bootstrap Sampling: The bootstrap resampling method is employed to randomly sample the original dataset with replacement, generating multiple subsets. Each subset trains a decision tree, while the unsampled instances (termed out-of-bag (OOB) data) are reserved for validation.

(4) Tree Growth: Each decision tree randomly selects several features from the input variable x for the current node's split subset. The optimal feature for splitting is determined by maximizing variance reduction, which enables maximal tree growth.

(5) OOB Validation: The OOB data, which are excluded from the training process, serve to evaluate model performance and accuracy. The optimal number of decision trees, T, is determined by iteratively refining the model based on OOB prediction errors until convergence.

(6) Prediction: For the input variable x_i (i = 1, 2, ... , k) that needs to be predicted. Each decision tree will output a corresponding predicted value y_i (i = 1, 2, ... , k).The predicted value of RFR is the average of the predicted values of all decision trees.

3.3 Support Vector Regression

SVR is a regression analysis method derived from Support Vector Machines (SVM). While SVR shares conceptual similarities with traditional SVM classification algorithms, its objective diverges by focusing on identifying an optimal fitting function for continuous prediction problems, as opposed to delineating classification boundaries [41]. The basic idea of the SVR model is to find a function that can predict the output value as accurately as possible while keeping the complexity of the model low to avoid overfitting. To achieve this goal, SVR uses a loss function called "ε-insensitive loss function", which means that the loss is only calculated when the prediction error exceeds a certain threshold, ε [42]. Its basic definition is as follows:

$$L\left(y, f\left(x\right)\right) = \begin{cases} 0 & if \left|y - f\left(x\right)\right| \le \varepsilon \\ \left|y - f\left(x\right)\right| - \varepsilon & otherwise \end{cases} \tag{2}$$

In the nonlinear case, by introducing relaxation variables ε_i and ε_i^*, as well as a penalty coefficient C, the penalty coefficient is used to control the trade-off between model complexity and data fitting. This transforms the optimization problem of SVR into solving the following optimization problem:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \varepsilon_i + \varepsilon_i^* \tag{3}$$

The model is subject to the following constraints:

$$\begin{cases} y_i - \left(w \cdot x + b\right) \le \varepsilon + \varepsilon_i \\ \left(w \cdot x + b\right) - y_i \le \varepsilon + \varepsilon_i^* \\ \varepsilon \ge 0, \varepsilon_i^* \ge 0 \end{cases} \tag{4}$$

The solution to this problem can be achieved by introducing Lagrange functions $L(w, b, \varepsilon_i, \varepsilon_i)$, transforming the original problem into a dual problem, and using KKT conditions to find the optimal solution. The specific process is as follows:

$$L\left(w, b, \varepsilon_i, \varepsilon_i\right) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \varepsilon_i + \varepsilon_i^* - \sum_{i=1}^{m} a_i \left(\varepsilon + \varepsilon_i - y_i + w \times x_i + b\right)$$

$$- \sum_{i=1}^{m} a_i^* \left(\varepsilon + \varepsilon_i^* + y_i - w \times x_i - b\right) - \sum_{i=1}^{m} \left(\eta_i \varepsilon_i + \eta_i^* \varepsilon_i^*\right) \tag{5}$$

where, $a_i, a_i^*, \eta_i, \eta_i^*$ are Lagrange operators, partial derivatives are obtained for $w, b, \varepsilon_i, \varepsilon_i$, respectively, to obtain saddle points. Then, the results are substituted into the original optimization problem, which can be transformed into the following dual problem:

$$max\,W(a) = -\frac{1}{2}\sum_{i,j=1}^{m}(a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) - \varepsilon\sum_{i=1}^{m}(a_i + a_i^*) + \sum_{i=1}^{m}y_i(a_i - a_i^*) \tag{6}$$

The constraints imposed on the SVR model are as follows:

$$\begin{cases} 0 \le a_i, a_i^* \le C \\ \sum_{i=1}^{m}(a_i - a_i^*) = 0 \end{cases} \tag{7}$$

where, $K(x_i, x_j)$ is a kernel function used to calculate the inner product of data points, x_i and x_j in high-dimensional space. The commonly used kernel functions include linear kernel, polynomial kernel, Gaussian kernel, etc. Among them, Gaussian kernel has the best applicability and is also known as radial basis function kernel (RBF).

Based on the above analysis, obtain the optimal solution and derive the final regression function:

$$f(x) = \sum_{i=1}^{m}(a_i - a_i^*)K(x_i, x_j) + \overline{b} \tag{8}$$

where, $\overline{b}$ is the bias term determined through training data.

The steps for establishing a support vector machine regression model are as follows:

(1) Read and preprocess the dataset.

(2) Build a model and define a hyperparameter search space.

(3) Use grid search cross-validation to find the optimal hyperparameters.

(4) Train the model using the optimal hyperparameters and conduct model evaluation.

(5) Use the trained model to predict new input data and output the prediction results.

3.4 Gaussian Process Regression

GPR is a data-driven method proposed by Williams and Rasmussen. Its essence lies in learning the kernel function with probabilistic significance [43]. The optimal hyperparameters are obtained by analyzing historical sample data, after which a prediction model is developed to forecast subsequent samples. The prediction generated by the GPR method carry probabilistic significance and can describe associated with these predictions. Compared with ANN, SVR and other methods, the hyperparameters required for prediction are easier to obtain and are suitable for high-dimensional nonlinear systems. Fig. 3 illustrates the specific process of GPR. The GPR algorithm is described below.

Given the training set D:$\{X_i, t_i\}_{i=1}^{N}$, N is the number of training samples, X_i is the vector defining time t_i, and the joint probability distribution function of the random variables in the Gaussian process is:

$$P(t|C(X_m, X_n; \Theta), \{X_n\}) = \frac{1}{Z}\exp\left(-\frac{1}{2}(t - \mu)^{\mathrm{T}} \cdot C(X_m, X_n; \Theta)^{-1}(t - \mu)\right) \tag{9}$$

where, X_m and X_n represent the m-th and n-th vectors, $C(X_m, X_n; \Theta)$ is parameter covariance function, μ represents mean vector.
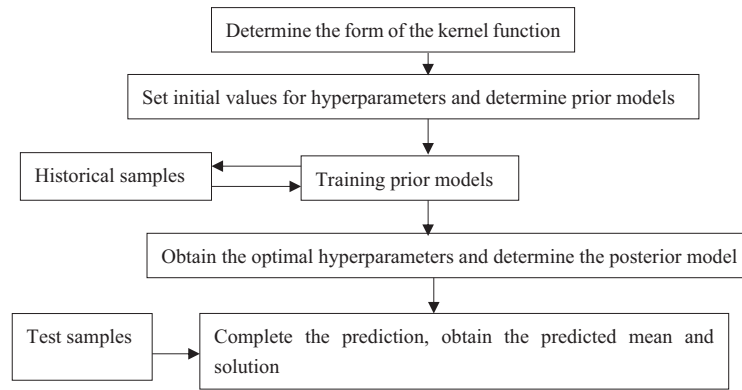
Figure 3: Process of GPR

The most widely used covariance function is:

$$C(X_m, X_n; \Theta) = \theta_1 \exp\left(-\frac{1}{2}\sum_{i=1}^{L} \frac{\left(X_m^{(l)} - X_n^{(l)}\right)}{\delta^2}\right) \tag{10}$$

where, L represents the number of elements in vector X_i, δ is variance, θ_1 and θ_2 are parameters.

For a new input vector X_{N+1}, use a predictive model to estimate its t_{N+1} value rate distribution, expectation, and variance are used for prediction, and the model function is represented as:

$$t_n = y(X_n) + v_n \tag{11}$$

where, $y(X_n)$ represents modeling functions, v_n is the prediction error.

So the model probability is:

$$P(T_N | \{X_n\}, A, B) = \int P(T_N | \{X_n\}, y, v) P(y|A) \cdot P(v|B) \, dy dv \tag{12}$$

where, $P(y|A)$ is the pre probability distribution of y(x), A is a set of hyperparameters of $P(y|A)$, $P(v|B)$ is the pre probability distribution of prediction error, B is a parameter representing the error v.

Let $T_N = (t_1, t_2, \ldots, t_n)$, $T_{N+1} = (t_1, t_2, \ldots, t_n, t_{n+1})$, therefore the conditional distribution of t_{N+1} can be represented by Eq. (13), and it can be used to predict t_{N+1}:

$$P(t_{N+1}|D, A, B, X_{N+1}) = \frac{P(T_{N+1} | \{X_n\}, A, B, X_{N+1})}{P(T_N | \{X_n\}, A, B)} \tag{13}$$

According to Bayesian method, it can be inferred that the distribution function of t_{N+1} is:

$$P(t_{N+1}|D, C(X_n, X_m; \Theta), X_{N+1}, \Theta) = \frac{P(T_{N+1}|C(X_n, X_m; \Theta), \Theta, X_{n+1}, \{X_n\})}{P(T_N|C(X_n, X_m; \Theta), \Theta, \{X_n\})} \tag{14}$$

By substituting the Gaussian distribution formula, the preliminary calculation can be simplified as:

$$P(t_{N+1}|D, C(X_n, X_m; \Theta), X_{N+1}, \Theta) = \frac{Z_N}{Z_{N+1}} \exp\left[-\frac{1}{2}\left(t_{N+1}^T C_{N+1}^{-1} t_N - t_N^T C_N^{-1} t_N\right)\right] \tag{15}$$

where, Z_N and Z_{N+1} are two normalization constants.

After further simplification, the following equation can be obtained:

$$P\left(t_{N+1}|D, C(X_n, X_m; \Theta), X_{N+1}, \Theta\right) = \frac{1}{z} \exp\left(-\frac{\left(t_{N+1} - \hat{t}_{N+1}\right)^2}{2\sigma^2_{\hat{t}_{N+1}}}\right) \tag{16}$$

where, $\hat{t}_{N+1}$ represented expected value. Therefore, predictions can be made based on their expectations and variances.

3.5 Experimental Database

The predictive reliability and accuracy of machine learning models are critically contingent upon the representativeness and robustness of the training dataset. In the frameworks of ANN, RFR, GPR, and SVR, input data are conventionally partitioned into training and testing subsets. The former serves to calibrate model parameters and optimize predictive performance, whereas the latter is employed to assess the accuracy and stability of the models.

The experimental dataset employed in this investigation comprises 151 data entries, partitioned into two categories: 89 records of pure CO_2 injection and 62 records of impure CO_2 injection. The dataset was divided into training (90%, $n = 136$) and testing (10%, $n = 15$) subsets to ensure unbiased model validation. These data were collected from various mature experimental works in the literature [13,15,44,45]. Data sets include experimental values for reservoir temperature, composition of drive gas (y_{CO2}, y_{H2S}, y_{N2}, and y_{C2-C4}), molecular weight of the C_{7+} fraction in crude oil (MW_{C7+}), the ratio of volatile (X_{vol}) to intermediate (X_{int}) components in crude oil, pure and impure CO_2-oil MMP. Table 3 presents a statistical description of the dataset used. A detailed analysis of the experimental data highlights that the reservoir temperature ranges from 31.00°C to 148.90°C, the experimental values for MMP range from 6.55 to 40.22 MPa.

Table 3: Statistical description of the dataset used in this studies

Parameters	Unit	Min	Max	Mean	STD	25%	50%	75%
T_R	°C	31.00	148.90	80.49	81.11	57.20	81.66	112.10
y_{CO2}	%	30.00	100.00	88.78	100.00	83.00	100.00	100.00
y_{H2S}	%	0	50.00	2.81	0	0	0	0
y_{C1}	%	0	54.00	6.27	0	0	0	8.05
y_{N2}	%	0	19.00	1.83	0	0	0	0
X_{int}	%	0	42.44	16.16	14.28	11.37	14.28	20.59
X_{vol}	%	0	65.61	24.77	25.01	10.50	25.56	38.80
y_{C2-C4}	%	0	25.00	1.83	0	0	0	0
X_{C5-C6}	%	0.01	11.79	6.19	6.52	4.22	6.47	8.21
MW_{C7+}	g/mol	153.90	402.70	227.26	220.00	200.00	220.00	240.00
MMP	MPa	6.55	40.22	20.31	20.24	12.96	20.43	25.31

4 Results and Discussions

In this study, we have researched and implemented four intelligent algorithm technologies, including GA-BPNN, RFR, GPR and SVR, and established an accurate and stable MMP prediction model. The optimal configuration of the GA-BPNN model is obtained through trial and error. Table 4 shows the optimal network structure of the GA-BPNN model for CO_2-oil MMP prediction. All models are written and implemented in Python.

Table 4: The optimal network structure of GA-BPNN model for CO_2-oil MMP prediction

Parameters	Value
Number of hidden layer	2
Number of neurons in hidden layer 1	128
Number of neurons in hidden layer 2	128
Transition function of input-hidden layer 1	Relu
Transition function of hidden layer 1-hidden	Relu
Transition function of hidden-output layer	Relu
Input data	[−1, 1]
Learning rate	0.001
Training algorithm	IM
Cycle index	9000
Population size of GA	60
Selectivity factor	1.0
Crossover factor	0.75
Mutation factor	0.2

To significantly demonstrate the predictive performance of each model, four statistical error functions were selected to evaluate the performance of the models: the coefficient of variation (CV), absolute percent relative error (APRE), average absolute percent relative error (AAPRE), and root mean square error (RMSE). The closer the value is to 0, the higher the accuracy of the model. Additionally, the coefficient of determination (R^2) is also used as the regression evaluation index, and the closer R^2 is to 1, the better the fitting effect. The specific calculation methods for the four statistical error functions are as follows:

$$CV = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{MMP_i^{mea}-MMP_i^{pre}}{MMP_i^{mea}}\right)^2}}{\overline{MMP_i^{mea}}} \tag{17}$$

$$APRE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{MMP_i^{mea}-MMP_i^{pre}}{MMP_i^{mea}}\right)\times 100\% \tag{18}$$

$$AAPRE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{MMP_i^{mea}-MMP_i^{pre}}{MMP_i^{mea}}\right|\times 100\% \tag{19}$$

$$RMSE = \sqrt{\frac{\sum_i^N\left(MMP_i^{mea}-MMP_i^{pre}\right)^2}{N}} \tag{20}$$

$$R^2 = 1 - \frac{\sum_i^N\left(MMP_i^{mea}-MMP_i^{pre}\right)^2}{\sum_i^N\left(MMP_i^{mea}-\overline{MMP_i^{pre}}\right)^2} \tag{21}$$

where, MMP_i^{mea} is MMP value of experimental, MMP_i^{pre} is the MMP predicted value, $\overline{MMP_i^{pre}}$ is the MMP predicted average value.

4.1 Comparison between Different Models

In this section, we will statistically analyze the values of four statistical error functions, with Table 5 displaying the implementation performance of each model. In this table, the performance of the training and the testing data is listed separately. From Table 5, it can be observed that the RFR model has R^2 values close to 1 compared to the other three models on both the training and testing sets, specifically are 0.98 and 0.95, respectively. The MAPE and RMSE values on the testing set are also quite low, at 4.31% and 1.00%, respectively. These indicators suggest that the RFR model exhibits the highest prediction accuracy and strong stability. Besides, the GA-BPNN model demonstrates good prediction performance, ranking second to RFR. In comparison, it is evident that GPR has the lowest performance among the four models.

Table 5: Evaluation index values of the four MMP prediction models

Data base	Evaluating indicator	GA-BPNN	RFR	SVR	GPR
	R^2	0.95	0.98	0.93	0.91
	APRE	5.10	4.31	5.45	8.05
Training data	AAPRE	2.91	2.51	3.28	4.76
	RMSE	1.25	1.00	1.47	1.54
	CV	11.35	5.98	7.17	10.06
	R^2	0.93	0.95	0.90	0.88
	APRE	6.55	5.98	7.26	10.64
Test data	AAPRE	5.54	5.05	6.56	8.51
	RMSE	2.45	2.13	3.01	4.02
	CV	8.65	9.60	7.19	8.38

Among the four models evaluated, the RFR operates under the ensemble learning paradigm, which aggregates predictions from multiple decision trees to mitigate overfitting and enhance generalization. This method has obvious advantages over neural networks in processing small and medium-sized data sets (<10,000 samples).The GPR model is distinguished by kernel flexibility. By selecting the appropriate covariance function, GPR can adapt to different data structures (periodic/non-periodic), and provide probability output to quantify the prediction uncertainty, which is a very valuable feature for risk perception decision-making in reservoir engineering. However, GPR's dependence on computationally intensive hyperparameter optimization often leads to local optimum, especially in high-dimensional parameter space, so advanced global optimization techniques such as Bayesian optimization are needed. Due to its ε-insensitive loss function, SVR is theoretically robust to outliers and noise, and can ignore the error within the tolerance threshold. Its performance depends heavily on the choice of kernel functions (e.g., linear, polynomial) and hyperparameter tuning, but it has interpretability defects and sensitivity to parameter initialization. The predictive performance of the neural network model in this study was not outstanding, and the BPNN algorithm optimized with GA did not demonstrate its outstanding predictive ability. It is speculated that this may be due to the lack of massive data in related research fields, and the big data fitting advantage of the GA-BPNN model has not been fully utilized.

To better illustrate the model performance, Fig. 4 displays the MMP data predicted by RFR, GA-BPNN, SVR, and GPR models. From Fig. 4, it can be seen that the RFR, GA-BPNN, and SVR models have achieved good alignment near the unit slope compared to the GPR model. Among the four models, the error distribution of the RFR and the GA-BPNN models is relatively balanced, with fewer outliers, thus ensuring

good stability. The error visualization of the GPR model around the unit slope is quite evident, but it remains within an acceptable range.
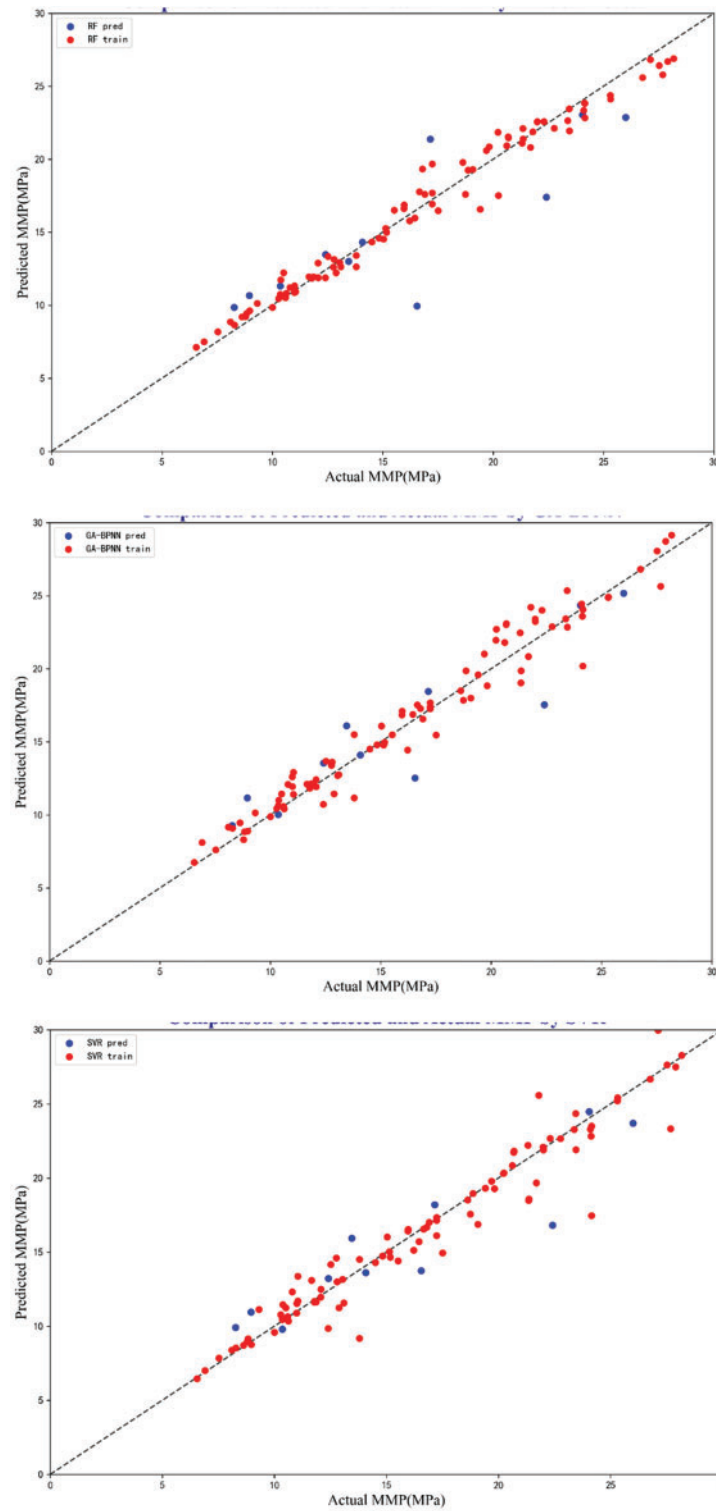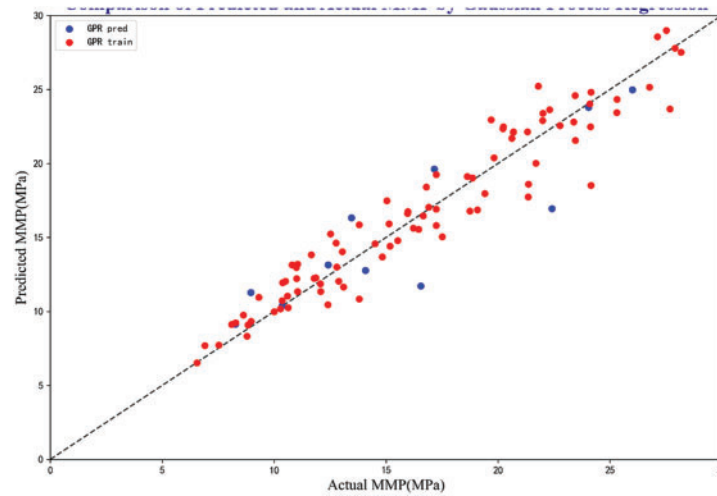

Figure 4: (Continued)

Figure 4: Prediction performance of four models

4.2 Comparison Models with Correlations from Literature

To further validate the predictive performance of the model, a statistical error analysis will be conducted based on previous models. The prediction results of the RFR model and GA-BPNN model developed in this paper will be compared with six forecasting MMP methods proposed by Lee [46], Cronquist [47], Yellig and Metcalfc [16], Alston et al. [18], Emera et al. [28], and Yuan et al. [48]. The six comparative models selected in this article simultaneously consider the influence of CO_2 injection gas impurities on the prediction results of CO_2-oil MMP, as well as the influencing factors of MMP prediction under pure CO_2 injection. Therefore, they were used for comparison with RFR and GA-BPNN models.

Fig. 5 visually displays the comparison results between the simulated MMP values and literature methods. The unit slope line in Fig. 5 serves as the evaluation criterion; the closer the data points are to the unit slope line, the higher the consistency between the simulated and experimental values. Table 6 shows the comparison of various indicators of the model. These results indicate that the RFR model and GA-BPNN model have strong predictive abilities.
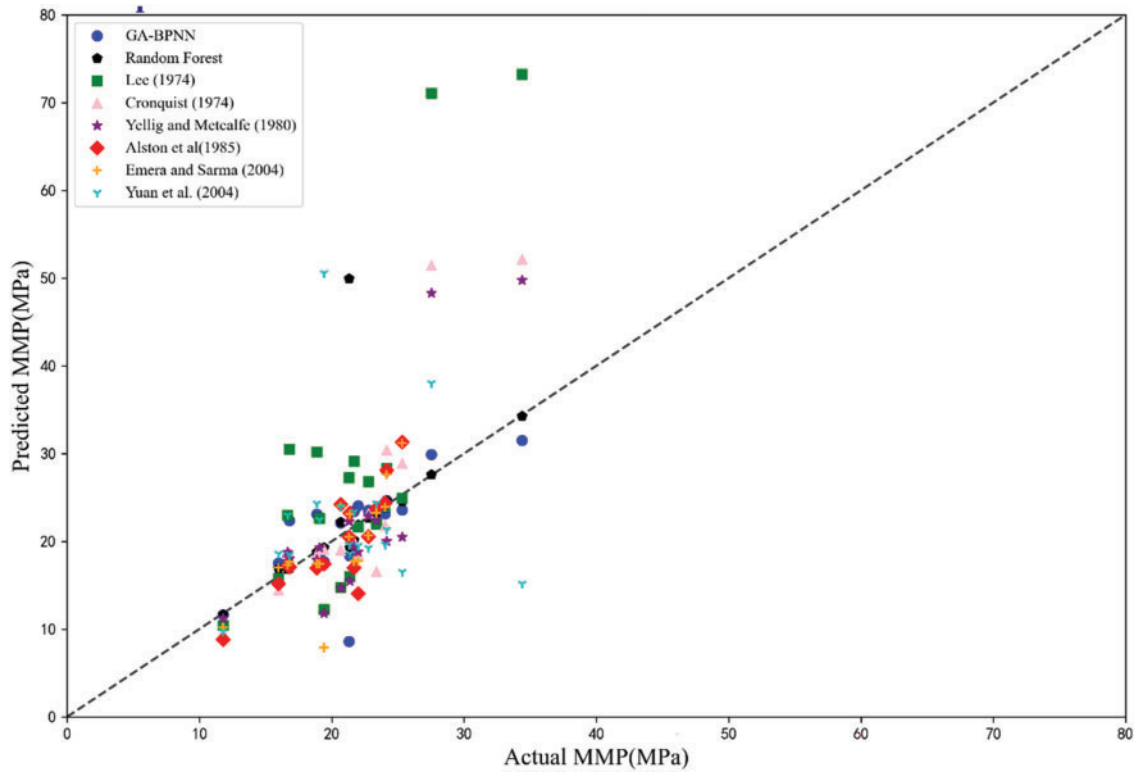
Figure 5: Comparison of the simulated MMP values with that calculated from literature models [16,18,28,46–48]

Table 6: Statistical error analysis of different literature models and the model presented in this studies

Model name	AAPRE	APRE	RMSE
RFR	4.55	1.01	0.89
GA-BPNN	6.45	2.86	1.66
Emera et al. [28]	15.46	5.12	3.84
Yuan et al. [48]	17.12	6.46	4.76
Alston et al. [18]	18.32	7.04	5.12
Yellig and Metcalfc [16]	19.65	7.85	5.85
Cronquist [47]	20.12	8.64	7.12
Lee [46]	22.54	9.54	8.86

4.3 Sensitivity Analysis

Analyze the impact of the characteristic parameters of the four models used in this article on predicting MMP in order to deepen our understanding of the CO_2-oil interaction mechanism and its intrinsic correlation. In this study, the relevance factor (R) was utilized to assess the extent of influence that each parameter has on MMP. For input variables, the R is a value ranging from −1 to 1. When R is negative, it represents that the input variable has a negative effect on the output variable, and when R is positive, it indicates that the input variable has a positive effect on the output variable. The absolute value of R represents the impact of the input variable on the output variable, and the larger the absolute value of R, the greater the

impact of the input variable on the output variable. The calculation formula for R is as follows:

$$R\left(X_j, MMP\right) = \frac{\sum_{i=1}^{n}\left(X_{j,i} - \overline{X_J}\right)\sum_{i=1}^{N}\left(MMP_i - \overline{MMP}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{j,i} - \overline{X_J}\right)^2 \sum_{i=1}^{N}\left(MMP_i - \overline{MMP}\right)^2}} \tag{22}$$

where, $X_{j,i}$ and $\overline{X_j}$ is the i-th and average number of input j, MMP_i indicates the i-th, $\overline{MMP}$ represents the average value of the MMP output.

Fig. 6 visually displays the R value of each input variable. From Fig. 6, it can be seen that X_{int}, $X_{C5\text{-}C6}$, and y_{H2S} can reduce the value of MMP. The observed negative correlations between X_{int}, $X_{C5\text{-}C6}$, y_{H2S}, and MMP can be attributed to their thermodynamic roles in reducing CO_2-oil interfacial tension. The ability of y_{H2S} to reduce MMP is due to the polarity of H_2S molecules, which can effectively reduce the interfacial tension between CO_2 and oil. The C_2-C_6 components in crude oil have similar molecular weights and properties to CO_2, which helps to increase the solubility of CO_2 in crude oil and reduce CO_2-oil MMP.
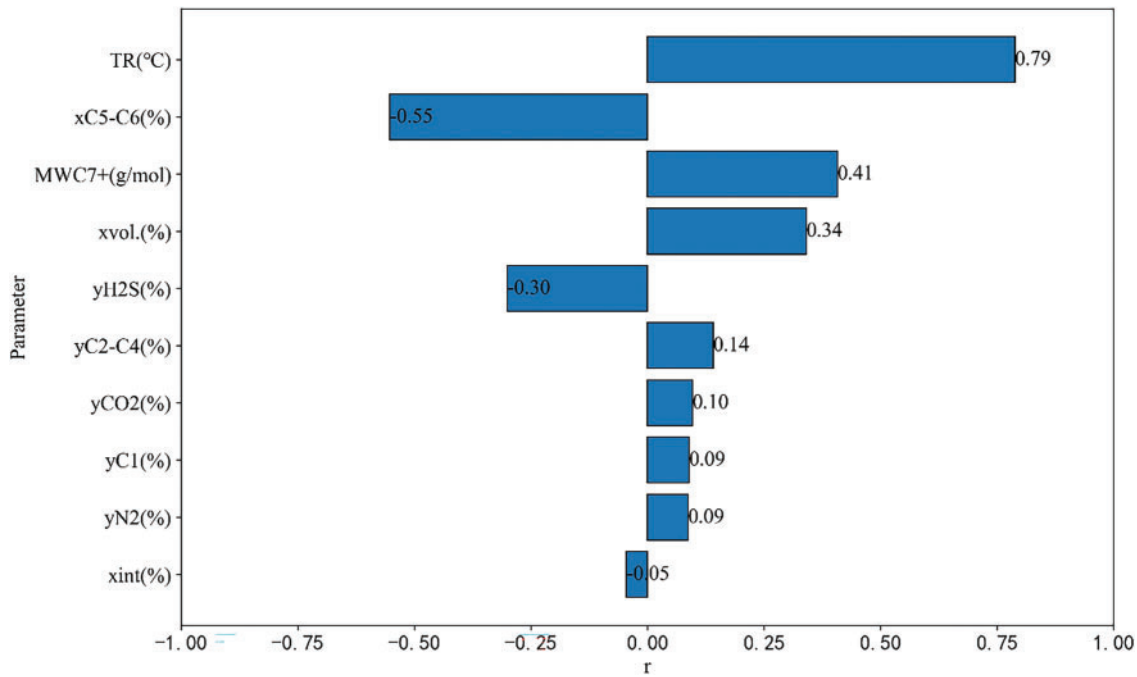

Figure 6: Relevancy factor of input variables on CO_2-oil MMP

The increase of the other seven factors (T_R, MW_{C7+}, X_{vol}, $y_{C2\text{-}C4}$, y_{CO2}, y_{C1}, y_{N2}) will lead to an increase in CO_2-oil MMP. The increase of T_R will result in a decrease in the solubility of CO_2 in crude oil, which seriously affects the ability of CO_2 extraction to extract light hydrocarbons, ultimately leading to an increase in CO_2-oil MMP. An increase of MW_{C7+} may create a greater difference between crude oil and CO_2 molecules, resulting in an increase in interfacial tension and thus affecting MMP. For the light components (X_{vol}) in crude oil, due to their easy volatilization, they are prone to entering the CO_2 gas phase from the crude oil during the CO_2 oil multi-stage mixed phase contact process, which results in a decrease in CO_2 purity and an increase in MMP. The increase in CO_2 concentration enhances the frequency of contact between CO_2 and crude oil. As CO_2 continuously extracts lighter components from the crude oil, the proportion of heavier components increases. Since these heavier components exhibit poor miscibility with CO_2, a higher pressure is required to

achieve miscibility. Within the parameter curve range, the influence of the ten influencing factors on MMP is ranked in descending order: $T_R > X_{C5\text{-}C6} > MW_{C7+} > X_{vol} > y_{H2S} > y_{C2\text{-}C4} > y_{CO2} > y_{C1}, y_{N2} > X_{int}$.

4.4 Models Application to Ultra-Deep Oil Reservoir

The Xinghua Block of Bayan oil field is a typical ultra-deep oil reservoir characterized by a large burial depth (>5000 m), high formation temperature and pressure, complex crude oil composition, and intricate fluid phases [49,50]. To verify the applicability of the established CO_2-oil MMP prediction models for ultra-deep oil reservoirs, experimental measurement data of CO_2 flooding MMP from two wells in Xinghua block of Bayan oil field were selected. The thin tube experiment refers to a simulated displacement test conducted in a thin tube model, and the experimental flowchart is shown in Fig. 7. The criterion for determining whether the thin tube experiment qualifies as a mixed-phase displacement is that the crude oil recovery rate, when injected with 1.2 pore volume(PV) exceeds 90%. Additionally, as the displacement pressure increases, there is no significant rise in displacement efficiency, and mixed-phase fluids can be observed in the observation window. The method for determining MMP is to plot the relationship curve between the recovery degree and displacement pressure when injecting 1.2 PV in each thin tube experiment, while ensuring that the thin tube experiment achieves three cycles of miscible flooding and three cycles of immiscible flooding [51].The pressure corresponding to the intersection of the non-mixed phase and mixed phase curves is the minimum mixed phase pressure (MMP), as shown in Fig. 8. The results of the slim tube test showed that injecting pure CO_2 at formation temperatures of 139°C and 133.9°C resulted in MMP values of 49.79 and 48.48 MPa for the reservoir, respectively.
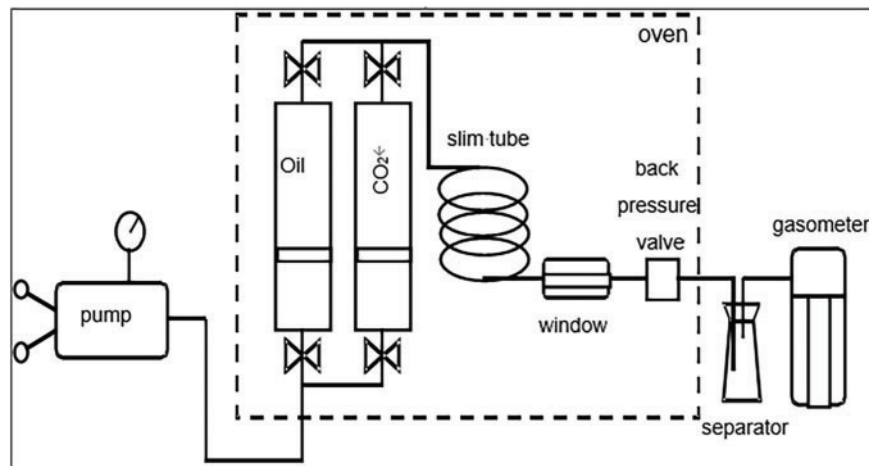

Figure 7: Flow chart of thin tube test

Calculate the MMP using the development prediction models and compare the calculated value with the experimental value measured by the thin tube experiment, Table 7 shows the comparison results between the MMP values predicted by four models and the experimental values measured by capillary experiments. As summarized in Table 7, both the RFR and GA-BPNN models demonstrate superior predictive accuracy, with RMSE values below 3% and AAPRE less than 0.1%. Furthermore, all four models exhibit minimal deviations between predicted and experimentally measured MMP, with errors consistently confined within engineering-acceptable thresholds (relative error < 5%). These results validate the robustness and applicability of the proposed models for MMP prediction in ultra-deep reservoirs characterized by high-temperature, high-pressure conditions.
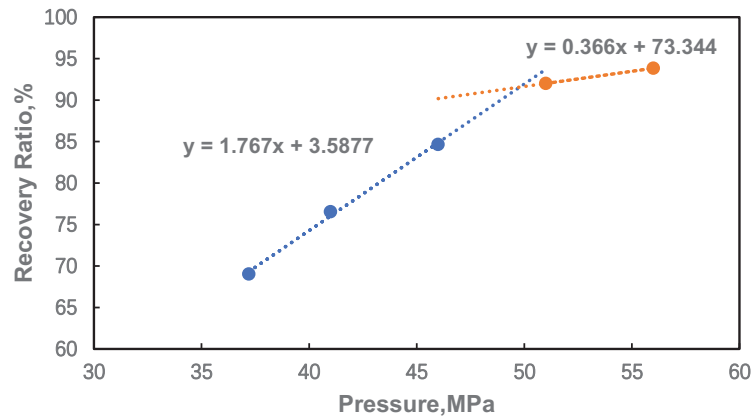
Figure 8: Relationship curve between CO_2 displacement recovery degree and displacement pressure in the formation oil slim tube experiment of Bayan oil field

Table 7: Comparison of experimental and predicted values of CO_2-oil MMP

Model	Temp (°C)	MW_{C7+} (g/mol)	X_{vol} (%)	X_{int} (%)	X_{C5-C6} (%)	Experimental value (MPa)	Prediction value (MPa)	RMSE (%)	AAPRE (%)
RFR	139.00	296.92	8.95	3.95	4.40	49.79	50.12	0.33	0.01
	133.90	279.88	9.95	4.99	4.14	48.18	49.00	0.82	0.02
GA-BPNN	139.00	296.92	8.95	3.95	4.40	49.79	47.56	2.23	0.04
	133.90	279.88	9.95	4.99	4.14	48.18	45.65	2.53	0.05
SVR	139.00	296.92	8.95	3.95	4.40	49.79	53.85	4.06	0.08
	133.90	279.88	9.95	4.99	4.14	48.18	52.45	4.27	0.09
GPR	139.00	296.92	8.95	3.95	4.40	49.79	43.56	6.23	0.13
	133.90	279.88	9.95	4.99	4.14	48.18	43.12	5.06	0.11

5 Conclusion

In this study, we developed four MMP prediction models based on machine learning, namely RFR,GA-BPNN,SVR and GPR. Additionally, we separately investigated pure CO_2 gas injection and impure CO_2 injection as described in the literature. Across these four models, total of ten impact factors (e.g., T_R, x_{int}, x_{vol}, x_{C5-C6}, MW_{C7+}, y_{C1}, y_{N2}, y_{CO2}, y_{C2-C4}, y_{H2S}) were selected, with MMP serving as the output variable. Based on a comprehensive understanding of four models and discussions of the simulation results, the study's conclusions can be briefly described as follows:

(1) The developed fours models for predicting MMP can effectively predict MMP for both pure CO_2 and impure CO_2 gas injections, and they can all be utilized for MMP prediction in ultra-deep oil reservoirs.

(2) Among the four models studied in this article, the RFR model exhibits the best predictive performance, while the GPR model shows the worst predictive performance. However, the BPNN algorithm optimized with GA did not demonstrate its outstanding predictive ability, which may be due to its incomplete utilization of the advantages of big data fitting.

(3) In the comparison with experimental results with those from six forecasting MMP methods, the RFR model and GA-BPNN model showed excellent performance.

(4) Among the ten influencing factors studied, X_{int}, X_{C5-C6}, and y_{H2S} are negatively correlated with MMP, while T_R, MW_{C7+}, X_{vol}, y_{C2-C4}, y_{CO2}, y_{C1} and y_{N2} are positively correlated with MMP. Among those, T_R has the greatest impact on MMP, followed by X_{C5-C6} and MW_{C7+}. The degree of influence of the ten influencing

factors on MMP, in descending order, is as follows: $T_R > X_{C5\text{-}C6} > MW_{C7+} > X_{vol} > y_{H2S} > y_{C2\text{-}C4} > y_{CO2} > y_{C1}, y_{N2} > X_{int}$.

(5) The difference between the properties of crude oil components and those of injected gas components impacts on MMP. The closer the properties of the component are to those of gas, the smaller the interfacial tension between oil and gas, and the lower the MMP.

(6) Although the four developed models have demonstrated promising performance in predicting the MMP of ultra-deep reservoirs, the limited sample size prevents a comprehensive analysis of interaction mechanisms among different influencing factors. Future research will focus on expanding the dataset and integrating experimental studies to systematically investigate the interaction mechanisms between various parameters affecting MMP, as well as to further improve and establish a high-performance MMP prediction model for ultra-deep reservoirs.

Availability of Data and Materials: The data that support the findings of this study are available from author, Tianfu Li, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report in the present study.

Nomenclature

CO_2	Carbon dioxide
EOR	Enhanced Oil Recovery
MMP	Minimum Miscible Pressure
AI	Artificial Intelligence
API	American Petroleum Institute
CCUS	Carbon Capture, Utilization and Storage
GA	Genetic Algorithm
BPNN	Back Propagation Neural Network
RFR	Random Forest Regression
P_r	Reservoir Pressure
SVR	Support Vector Regression
GPR	Gaussian Process Regression
LSSVM	Least Squares Support Vector Machine
GEP	Genetic Expression Programming
GA-BPNN	Genetic Algorithm Based Back Propagation Artificial Neural Network
ANN	Artificial Neural Network
PV	Pore Volume
ML	Machine Learning
DT	Decision Trees

y_{CO2}	Mole fraction of CO_2 in solution
STD	Standard Deviation
CV	Coefficient of Variation
APRE	Average Relative Error Percent
SVM	Support Vector Machines
AAPRE	Average Absolute Percent Relative Error
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
R	Relevance Factor
T_R	Reservoir Temperature
T_c	Critical temperature
MW_{C5+}	Molecular weight of C_{5+} oil fraction, g/mol
MW_{C7+}	Molecular weight of C_{7+} oil fraction, g/mol
X_{vol}	Mole fraction of volatile oil components, including (C_1 and N_2)
X_{int}	Mole fraction of intermediate oil components, including (C_2-C_4, CO_2, and H_2S)
X_{C5-C6}	Mole fraction of oil components, including(C_5 and C_6)
y_{C1}	Mole fraction of C_1 in the CO_2 injection gas
y_{N2}	Mole fraction of N_2 in the injection gas
y_{H2S}	Mole fraction of H_2S in the injection gas
y_{C2-C4}	Mole fraction of C_2-C_4 in the injection gas
MLR	Multiple Linear Regression

References

1. Kumar N, Sampaio MA, Ojha K, Hoteit H, Mandal A. Fundamental aspects, mechanisms and emerging possibilities of CO_2 miscible flooding in enhanced oil recovery: a review. Fuel. 2022;330(3):125633. doi:10.1016/j.fuel.2022.125633.

2. Al-Obaidi DA, Al-Mudhafar WJ, Hussein HH, Rao DN. Experimental influence assessments of water drive and gas breakthrough through the CO_2-assisted gravity drainage process in reservoirs with strong aquifers. Fuel. 2024;370:131873. doi:10.1016/j.fuel.2024.131873.

3. Ali M, Jha NK, Pal N, Keshavarz A, Hoteit H, Sarmadivaleh M. Recent advances in carbon dioxide geological storage, experimental procedures, influencing parameters, and future outlook. Earth-Sci Rev. 2022;225(14):103895. doi:10.1016/j.earscirev.2021.103895.

4. Liang B, Chen C, Jia C, Wang C, Wang X, Zha Y. Carbon capture, utilization and storage (CCUS) in oil and gas reservoirs in China: status, opportunities and challenges. Fuel. 2024;375(2):132353. doi:10.1016/j.fuel.2024.132353.

5. Huang T, Li X, Cheng L. Modeling of the flow and heat transfer of supercritical CO_2 flowing in serpentine tubes. Front Heat Mass Trans. 2020;15(1):1–8. doi:10.5098/hmt.15.15.

6. Kang X, Kang W, Li Z, Yang H, Xie A, Li M, et al. Stability influence factors and mechanism of produced emulsion from CO_2 flooding. J Molecular Liq. 2021;333:115974. doi:10.1016/j.molliq.2021.115974.

7. Yu H, Feng J, Zeng H, Xie Q, Wang J, Song J, et al. A new empirical correlation of MMP prediction for oil—impure CO_2 systems. Fuel. 2024;371:132043. doi:10.1016/j.fuel.2024.132043.

8. Sun J, Sun L, Bao B. Recovery mechanism of supercritical CO_2 miscible flooding in 30 nm nanomatrices with multi-scale and fracture heterogeneity. Fuel. 2025;392:134835. doi:10.1016/j.fuel.2025.134835.

9. Flock DL, Nouar A. Parametric analysis on the determination of the minimum miscibility pressure in slim tube displacements. J Can Pet Technol. 1984;23(5). doi:10.2118/84-05-12.

10. Song Y, Song Z, Mo Y, Zhou Q, Jing Y, Chen F, et al. Determination of minimum miscibility and near-miscibility pressures for CO_2-oil mixtures in shale reservoirs. Fuel. 2025;388:134531. doi:10.1016/j.fuel.2025.134531.

11. Chen Z, Chen J, Zhang X. A comprehensive review of minimum miscibility pressure determination and reduction strategies between CO_2 and crude oil in CCUS processes. Fuel. 2025;384:134053. doi:10.1016/j.fuel.2024.134053.

12. Xian B, Hao H, Deng S, Wu H, Sun T, Cheng L, et al. Laboratory experiments of CO_2 flooding and its influencing factors on oil recovery in a low permeability reservoir with medium viscous oil. Fuel. 2024;371(4):131871. doi:10.1016/j.fuel.2024.131871.

13. Holm LW, Josendal VA. Mechanisms of oil displacement by carbon dioxide. SPE J Pet Technol. 1974;26:1427–38. doi:10.2118/4736-PA.

14. Mansour EM, Al-Sabagh AM, Desouky SM, Zawawy FM, Ramzi M. A new estimating method of minimum miscibility pressure as a key parameter in designing CO_2 gas injection process. Egyptian J Petrol. 2018;27(4):801–10. doi:10.1016/j.ejpe.2017.12.002.

15. Chen G, Gao H, Fu K, Zhang H, Liang Z, Tontiwachwuthikul P. An improved correlation to determine minimum miscibility pressure of CO_2–oil system. Green Ene Environ. 2020;5(1):97–104. doi:10.1016/j.gee.2018.12.003.

16. Yellig WF, Metcalfe RS. Determination and prediction of CO_2 minimum miscibility pressures (includes associated paper 8876). J Petrol Techol. 1980;32(1):160–8. doi:10.2118/7477-PA.

17. Glasø Ø. Generalized minimum miscibility pressure correlation. Soc Petrol Eng J. 1985;25(6):927–34. doi:10.2118/12893-PA.

18. Alston RB, Kokolis GP, James CF. CO_2 minimum miscibility pressure: a correlation for impure CO_2 streams and live oil systems. Soc Petrol Eng J. 1985;25(2):268–74. doi:10.2118/11959-PA.

19. Abdurrahman M, Permadi AK, Bae WS. An improved method for estimating minimum miscibility pressure through condensation-extraction process under swelling tests. J Petrol Sci Eng. 2015;131:165–71. doi:10.1016/j.petrol.2015.04.033.

20. Ge D, Cheng H, Cai M, Zhang Y, Dong P. A new predictive method for CO_2-oil minimum miscibility pressure. Geofluids. 2021;1:1–8. doi:10.1155/2021/8868592.

21. Chen Z, Li D, Zhang S, Liao X, Zhou B, Chen D. A well-test model for gas hydrate dissociation considering a dynamic interface. Fuel. 2022;314(1–2):123053. doi:10.1016/j.fuel.2021.123053.

22. Bagheri H, Mohebian R, Moradzadeh A, Olya BAM. Pore size classification and prediction based on distribution of reservoir fluid volumes utilizing well logs and deep learning algorithm in a complex lithology. Artif Intell Geosci. 2024;5:100094. doi:10.1016/j.aiig.2024.100094.

23. Bhattacherjee R, Botchway K, Pashin JC, Chakraborty G, Bikkina P. Developing statistical and machine learning models for predicting CO_2 solubility in live crude oils. Fuel. 2024;368:131577. doi:10.1016/j.fuel.2024.131577.

24. Wu B, Xie R, Xiao L, Guo J, Jin G, Fu J. Integrated classification method of tight sandstone reservoir based on principal component analysis—simulated annealing genetic algorithm-fuzzy cluster means. Petrol Sci. 2023;20(5):2747–58. doi:10.1016/j.petsci.2023.04.014.

25. Huang YF, Huang GH, Dong MZ, Feng GM. Development of an artificial neural network model for predicting minimum miscibility pressure in CO_2 flooding. J Petrol Sci Engg. 2003;37(1):83–95. doi:10.1016/S0920-4105(02)00312-1.

26. Shokrollahi A, Arabloo M, Gharagheizi F, Mohammadi AH. Intelligent model for prediction of CO_2—reservoir oil minimum miscibility pressure. Fuel. 2013;112:375–84. doi:10.1016/j.fuel.2013.04.036.

27. Kamari A, Arabloo M, Shokrollahi A, Gharagheizi F, Mohammadi AH. Rapid method to estimate the minimum miscibility pressure (MMP) in live reservoir oil systems during CO_2 flooding. Fuel. 2015;153:310–9. doi:10.1016/j.fuel.2015.02.087.

28. Emera MK, Sarma HK. Use of genetic algorithm to estimate CO_2–oil minimum miscibility pressure—a key parameter in design of CO_2 miscible flood. J Petrol Sci Eng. 2005;46(1):37–52. doi:10.1016/j.petrol.2004.10.001.

29. Birang Y, Dinarvand N, Shariatpanahi SF, Edalat M. Development of a new artificial-neural-network model for predicting minimum miscibility pressure in hydrocarbon gas injection. Paper presented at: The SPE Middle East Oil and Gas Show and Conference; 2007 Mar; Manama, Bahrain. doi:10.2118/105407-MS.

30. Bian X, Han B, Du Z, Jaubert JN, Li M. Integrating support vector regression with genetic algorithm for CO_2-oil minimum miscibility pressure (MMP) in pure and impure CO_2 streams. Fuel. 2016;182:550–7. doi:10.1016/j.fuel.2016.05.124.

31. Dehaghani AHS, Soleimani R. Prediction of CO_2-oil minimum miscibility pressure using soft computing methods. Chem Eng Technol. 2020;43(7):1361–71. doi:10.1002/ceat.201900411.

32. Amar MN, Zeraibi N. Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO_2-EOR process. Petroleum. 2020;6(4):415–22. doi:10.1016/j.petlm.2018.08.001.

33. Ghiasi MM, Mohammadi AH, Zendehboudi S. Use of hybrid-ANFIS and ensemble methods to calculate minimum miscibility pressure of CO_2—reservoir oil system in miscible flooding process. J Molecul Liquids. 2021;331:115369. doi:10.1016/j.molliq.2021.115369.

34. Al-Khafaji HF, Meng Q, Hussain W, Mohammed RK, Harash F, AlFakey SA. Predicting minimum miscible pressure in pure CO_2 flooding using machine learning: method comparison and sensitivity analysis. Fuel. 2023;354:129263. doi:10.1016/j.fuel.2023.129263.

35. Sebastian HM, Wenger RS, Renner TA. Correlation of minimum miscibility pressure for impure CO_2 streams. J Pet Technol. 1985;37:2076–82. doi:10.2118/12648-PA.

36. Fathaddin MT, Irawan S, Setiati R, Rakhmanto PA, Prakoso S, Mardiana DA. Optimized artificial neural network application for estimating oil recovery factor of solution gas drive sandstone reservoirs. Heliyon. 2024;10(13):e33824. doi:10.1016/j.heliyon.2024.e33824.

37. Zhang R, Tong W, Xu S, Qiu Q, Zhu X. ANN model with feature selection to predict turbulent heat transfer characteristics of supercritical fluids: take CO_2 and H_2O as examples. Int J Therm Sci. 2023;188(1):108247. doi:10.1016/j.ijthermalsci.2023.108247.

38. Bayazitova G, Anastasiadou M, Santos VDD. Oil and gas flow anomaly detection on offshore naturally flowing wells using deep neural networks. Geoene Sci Eng. 2024;242(4):213240. doi:10.1016/j.geoen.2024.213240.

39. Yang L, Li B, Zhang K, Huang C, Jing F. Multi-objective shape optimization of offshore swirl-vane oil-water separators using genetic algorithm network coupled computational fluid dynamics. Ocean Eng. 2025;325:120827. doi:10.1016/j.oceaneng.2025.120827.

40. Guo D, Kang Y, Wang Z, Zhao Y, Li S. Optimization of fracturing parameters for tight oil production based on genetic algorithm. Petroleum. 2022;8(2):252–63. doi:10.1016/j.petlm.2021.11.006.

41. Wen S, Wei B, You J, He Y, Xin J, Varfolomeev MA. Forecasting oil production in unconventional reservoirs using long short term memory network coupled support vector regression method: a case study. Petroleum. 2023;9(4):647–57. doi:10.1016/j.petlm.2023.05.004.

42. Dargahi-Zarandi A, Hemmati-Sarapardeh A, Shateri M, Menad NA, Ahmadi M. Modeling minimum miscibility pressure of pure/impure CO_2-crude oil systems using adaptive boosting support vector regression: Application to gas injection processes. J Petrol Sci Eng. 2020;184:106499. doi:10.1016/j.petrol.2019.106499.

43. Lv Q, Rashidi-Khaniabadi A, Zheng R, Zhou T, Mohammadi MR, Hemmati-Sarapardeh A. Modelling CO_2 diffusion coefficient in heavy crude oils and bitumen using extreme gradient boosting and Gaussian process regression. Energy. 2023;275:127396. doi:10.1016/j.energy.2023.127396.

44. Cui X, Zheng L, Liu Z, Cui P, Du D. Determination of the minimum miscibility pressure of the CO_2/oil system based on quantification of the oil droplet volume reduction behavior. Coll Surf A: Physicochem Eng Asp. 2022;653:130058. doi:10.1016/j.colsurfa.2022.130058.

45. Zang H, Yan S, Liu Z, Li Y, Du D. Determination of the minimum miscibility pressures of CO_2 and multicomponent oil system with a novel diffusion coefficient fit line intersection methodology. Fuel. 2025;385(43):134106. doi:10.1016/j.fuel.2024.134106.

46. Lee IJ. Effectiveness of carbon dioxide displacement under miscible and immiscible conditions. Calgary, AB, Canada: Petroleum Recovery Inst.; 1979. PRI-7910; CE-02239.

47. Cronquist C. Carbon dioxide dynamic displacement with light reservoir oils. In: Fourth Annual U.S. DOE Symposium; 1978; Tulsa, OK, USA.

48. Yuan H, Johns RT, Egwuenu AM, Dindoruk B. Improved MMP correlations for CO_2 floods using analytical gas flooding theory. Paper presented at: The SPE/DOE Symposium on Improved Oil Recovery; 2004 Apr; Tulsa, Oklahoma. doi:10.2118/89359-MS.

49. Ma Y, Cai X, Li M, Li H, Zhu D, Qiu N, et al. Research advances on the mechanisms of reservoir formation and hydrocarbon accumulation and the oil and gas development methods of deep and ultra-deep marine carbonates. Petrol Explorat Develop. 2024;51(4):795–812. doi:10.1016/S1876-3804(24)60507-0.

50. Lei Q, Xu Y, Yang Z, Cai B, Wang X, Zhou L, et al. Progress and development directions of stimulation techniques for ultra-deep oil and gas reservoirs. Petrol Explorat Develop. 2021;48:221–31. doi:10.1016/S1876-3804(21)60018-6.

51. Yu H, Lu X, Fu W, Wang Y, Xu H, Xie Q, et al. Determination of minimum near miscible pressure region during CO_2 and associated gas injection for tight oil reservoir in Ordos Basin, China. Fuel. 2020;263:116737. doi:10.1016/j.fuel.2019.116737.