

Research Article

Permeability Predictions for Tight Sandstone Reservoir Using Explainable Machine Learning and Particle Swarm Optimization

Jing-Jing Liu  and Jian-Chao Liu

School of Earth Science and Resources, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Jing-Jing Liu; liujingjing@chd.edu.cn

Received 15 September 2021; Accepted 14 December 2021; Published 6 January 2022

Academic Editor: Hongqing Song

Copyright © 2022 Jing-Jing Liu and Jian-Chao Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High-precision permeability prediction is of great significance to tight sandstone reservoirs. However, while considerable progress has recently been made in the machine learning based prediction of reservoir permeability, the generalization of this approach is limited by weak interpretability. Hence, an interpretable XGBoost model is proposed herein based on particle swarm optimization to predict the permeability of tight sandstone reservoirs with higher accuracy and robust interpretability. The porosity and permeability of 202 core plugs and 6 logging curves (namely, the gamma-ray (GR) curve, the acoustic curve (AC), the spontaneous potential (SP) curve, the caliper (CAL) curve, the deep lateral resistivity (RILD) curve, and eight lateral resistivity (RFOL) curve) are extracted along with three derived variables (i.e., the shale content, the AC slope, and the GR slope) as data sets. Based on the data preprocessing, global and local interpretations are performed according to the Shapley additive explanations (SHAP) analysis, and the redundant features in the data set are screened to identify the porosity, AC, CAL, and GR slope as the four most important features. The particle swarm optimization algorithm is then used to optimize the hyperparameters of the XGBoost model. The prediction results of the PSO-XGBoost model indicate a superior performance compared with that of the benchmark XGBoost model. In addition, the reliable application of the interpretable PSO-XGBoost model in the prediction of tight sandstone reservoir permeability is examined by comparing the results with those of two traditional mathematical regression models, five machine learning models, and three deep learning models. Thus, the interpretable PSO-XGBoost model is shown to have more advantages in permeability prediction along with the lowest root mean square error, thereby confirming the effectiveness and practicability of this method.

1. Introduction

Permeability is an important parameter in tight sandstone reservoir evaluation and oil and gas field development and is the basis for establishing geological models, accurately estimating oil and gas reserves, and determining reasonable development plans [1–4]. However, because tight sandstone reservoirs have experienced complex diagenesis and exhibit strong heterogeneity, it is difficult to predict their permeability with high accuracy [5, 6]. At present, the main methods for the high-precision prediction of permeability are (i) mathematical regression methods such as porosity and permeability regression based on petrophysical data or the response of logging curves [7–10]; (ii) theoretical modeling based on high-pressure mercury intrusion data, e.g., Win-

land, Purcell, Swanson, and Katz-Thompson; [11–13] and (iii) prediction based on special logging series, e.g., nuclear magnetic resonance logging and dipole sonic logging [14–16]. These methods have certain differences in terms of data requirements and parameter selection and are challenged by two main problems: (1) the accuracy of the mathematical regression method is poor. Whether it is based on porosity and permeability or the relationship between various logging curves and permeability, it is a complex nonlinear relationship. Hence, it is impossible to obtain high-precision permeability prediction results by using simple linear analytical expressions; (2) petrophysical experiments and special logging series each incur prohibitive costs and rather long evaluation periods. For petrophysical experiments, only a limited number of core plugs can be collected, and these

are unable to provide the same resolution as logging curves. The arbitrary selection of core plugs with irregular intervals also restricts the accurate evaluation of the heterogeneity of the permeability of the tight sandstone reservoir. Therefore, a convenient, cheap, and reliable method that can examine data and express complex nonlinear relationships to generate a reliable conclusion is desirable as an effective technical approach to characterizing the permeability of tight sandstone reservoirs [17].

Machine learning is the core method of research and application in the field of artificial intelligence and can quickly find the mapping relationship between large amounts of data [18]. In recent years, the application of machine learning to solve the task of predicting the petrophysical properties of the reservoir has attracted increasing attention from geologists [19–25]. Logging curves reflect the formation characteristics of different depths and contain various information on the formation parameters. Moreover, the data sets used for training are all easily-obtained, high-resolution logging curve data, thus, solving the cost problem. The studies of Al-Anazi and Gates [19] and Gholami et al. [20] both applied support vector machines (SVM) to construct the relationship between logging data and core permeability to achieve good results. Ao et al. [21] introduced the application of the random forest (RF) algorithm to predict the permeability and other formation characteristics and compared the reliability of the results with those of eight other algorithms for predicting the petrophysical properties of the reservoir. Although the above research has had good predictive effect, the accuracy of prediction is still limited because the SVM has no universal solution to nonlinear problems and because it is difficult to find a suitable kernel function. Moreover, the RF algorithm is prone to overfitting when solving the regression problem and is not at all well generalizable. In order to solve these problems, the abovementioned researchers proposed a powerful machine learning framework termed the extreme gradient boosting decision tree (XGBoost) [22]. The XGBoost framework has the following advantages: (i) it draws on the random forest algorithm to sample features; (ii) it establishes a new model by continuously reducing residuals, thus, decreasing any overfitting while reducing the amount of calculation; (iii) it uses standardized regularization terms that further mitigate against overfit in the trained model; and (iv) it enables the performance of parallel calculations to achieve high predictive efficiency and precision [23]. For example, Otchere et al. [24] explored the reservoir characterization method based on the XGBoost framework and carried out feature selection based on the integrated random forest algorithm and lasso regularization. The results show superior to the benchmark XGBoost and PCA-XGBoost models for predicting the petrophysical properties of the reservoir in the high permeability zone. In addition, Gu et al. [25] proposed an application of the particle swarm optimization (PSO) algorithm to improve the hyperparameter selection of the XGBoost framework to more accurately predict the permeability of tight sandstone reservoirs and favorably compared the performance with that of the stepwise, SVR, and GBDT algorithms.

Although the above methods show superiority in permeability prediction, the majority of machine learning methods are equivalent to black box processes in which a decision result is obtained from a batch of input data, but the basis of the intermediate decision is unknown and, hence, lacks interpretability [26]. While some studies have used feature importance methods to improve the interpretability of machine learning algorithms, this only reveals the key variables that affect the overall prediction result; it does not reveal the specific influence of the positive or negative aspects of each variable upon the final prediction, nor can it explain the specific details [27–29]. In 2017, Lundberg and Lee proposed an interpretable method based on the game theory approach, known as Shapley additive explanations (SHAP) [30]. This is a method of model interpretation after the fact, in which calculation of the Shapley value of each feature enables both a local and global interpretation (the latter being obtained by averaging the absolute Shapley values over all features). Hence, the SHAP method has attracted widespread attention in recent years for enabling the capture of meaningful features in the application of machine learning and for improving the interpretability and transparency of the algorithm [31–33].

In view of the above discussion, the present study attempts to use the SHAP method to introduce the interpretable XGBoost framework into the permeability prediction field. In addition, the particle swarm optimization algorithm is used to search for hyperparameters of the model and find the best combination in order to optimize the performance of the XGBoost. Finally, the performance of the proposed SHAP-PSO-XGBoost model in predicting the permeability of the tight sandstone reservoir is compared with that of traditional mathematical regression methods and other commonly-used machine learning algorithms, including the SVM, decision tree (DT), RF, gradient boosting decision tree (GBDT), and LightGBM. In addition, the performance of the proposed SHAP-PSO-XGBoost model is compared with that of traditional deep learning technologies such as convolutional neural networks (CNNs), long short-term memory (LSTM), and the gate recurrent unit (GRU). The experimental results indicate that the SHAP-PSO-XGBoost model has the highest accuracy, thus, demonstrating the feasibility and superiority of the proposed model and laying the foundation for the fine evaluation of heterogeneous tight sandstone reservoirs based on permeability.

The remainder of this article is organized as follows: Section 2 describes the data set and data preprocessing methods used in the study; Section 3 introduces the theoretical basis of the proposed method; Section 4 presents and discusses the results of the computational experiments, and Section 5 presents a brief summary.

2. Data Collecting and Preprocessing

2.1. Geological Setting and Data Acquisition. The Ordos Basin is located in central China and is the second largest sedimentary basin in China (Figure 1). The study area is the Yishan slope belt, which is the main body of the Ordos Basin and the main area of oil and gas enrichment. The

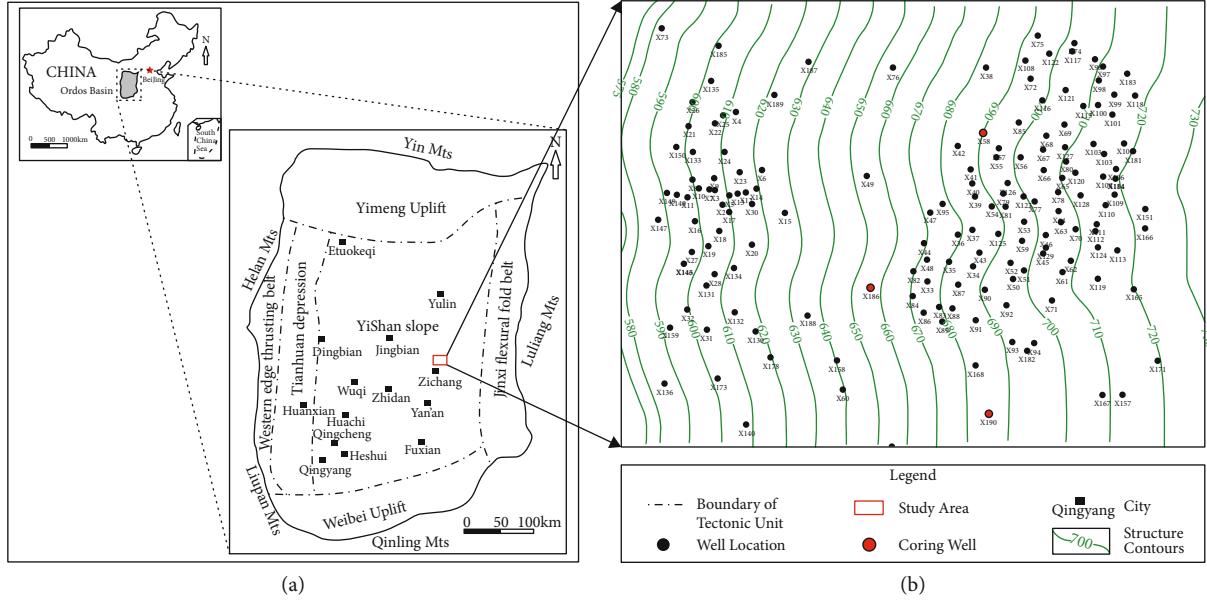


FIGURE 1: Location of the study area and coring wells. (a) Location of the study area. (b) Location of the coring wells.

morphology of the area is gentle and monoclinic in the west, with an inclination angle of less than 1°. During the Yanchang period of the Late Triassic, the basin formed a typical continental clastic sedimentary system characterized by river-lake facies under the background of overall subsidence. From bottom to top, it can be divided into ten oil-bearing formations. Among these, the Chang 6 and Chang 4 + 5 oil-bearing formations were formed in the early stage of contraction of the lacustrine basin. The lithology is dominated by fine-grained feldspar sandstone. These are the important oil and gas exploration target layers in the study area [34–36].

Overlying pressure porosity and permeability analyses were performed on 202 core plugs of the Chang 4 + 5 and Chang 6 tight sandstone reservoirs from two coring wells in the study area. The porosity of the core samples was found to be between 2.0 and 13.2%, with an average porosity of 10.3%, and the permeability was between 0.025×10^{-3} and $2.600 \times 10^{-3} \mu\text{m}^2$, with an average of $0.745 \times 10^{-3} \mu\text{m}^2$. The exponential regression relationship between the porosity and permeability of the core samples is presented in Figure 2, where the poor correlation is indicated by a correlation coefficient of only 0.441. This also indicates that the pore-throat structure of tight sandstone reservoirs is more complex than that of conventional oil reservoirs and, hence, it is more difficult to predict the permeability with high accuracy.

The core porosity (Por), core permeability, and sample points of the nine logging curves corresponding to the coring depths were selected as input data on the basis of core homing, giving a total of 202 data points. The six logging curves included the gamma-ray (GR) curve, acoustic curve (AC), spontaneous potential (SP) curve, caliper (CAL) curve, deep lateral resistivity (RILD) curve, and eight lateral resistivity (RFOC) curves. Among these, the RILD and RFOC used logarithmic data. In addition, the following three derived variables were used: the shale volume, the AC slope, and the GR slope.

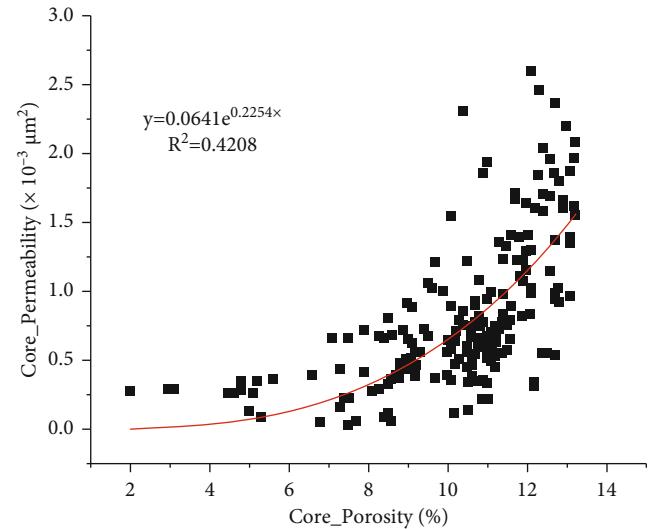


FIGURE 2: The relationship between core porosity and core permeability.

2.2. Data Normalization. To eliminate the systematic errors in the various measurement tools and the dimensional influence between the various logging curves [37], the input data were normalized according to the following equation:

$$x_{id} = \frac{x_i - x_{i \min}}{x_{i \max} - x_{i \min}}, \quad (1)$$

where x_{id} and x_i are, respectively, the actual and normalized values of each index, and $x_{i \max}$ and $x_{i \min}$ are the maximum and minimum values of each index. This process also facilitates the comparison of indicators having diverse units or magnitudes, as the quantified indicators will be distributed in the interval [0, 1]. In addition, the data normalization helps to accelerate the training process.

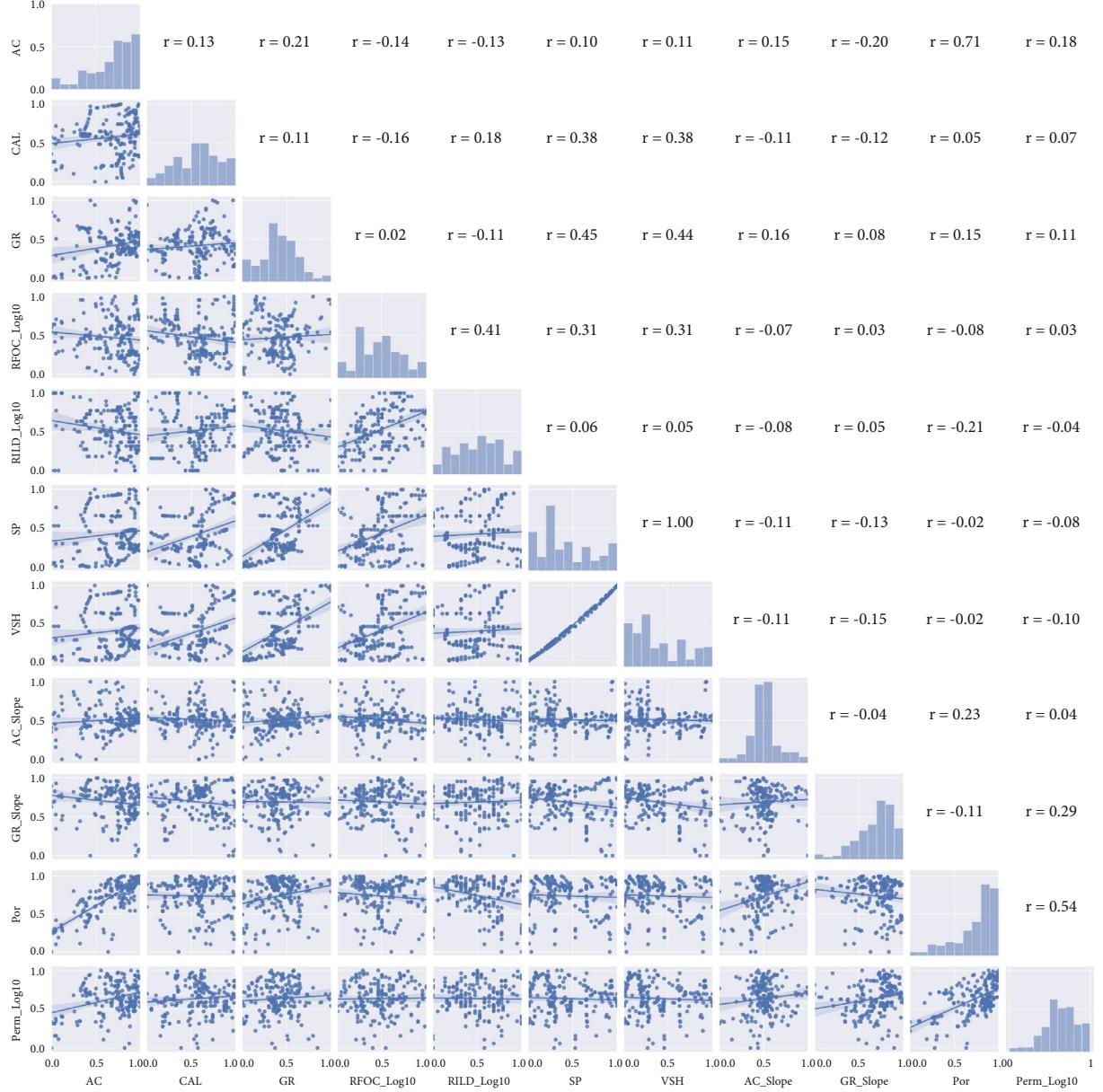


FIGURE 3: Graphical view of the normalized dataset (r represents the Spearman correlation coefficient).

The normalized core porosity and permeability are presented in Figure 3, along with the statistical characteristics of the wireline logging curves. Here, the diagonal direction shows the distribution of the 11 features, all of which are discrete data, and none (except for the GR, AC slope, and GR slope) conform to the normal distribution. Therefore, the Spearman correlation coefficient was used to measure the dependence between two variables and is presented on the upper half of the figure, while the lower half shows the linear fitting relationship between the two features. It can be seen that, except for the slightly higher correlation with the porosity ($r = 0.57$), the correlation between the permeability and the other features is very poor. This indicates that the use of any of these features to predict the permeability directly will be inefficient and unreliable.

3. Methodology

3.1. Explainable Machine Learning

3.1.1. XGBoost. The XGBoost is an efficient implementation of the GBDT algorithm proposed by Chen and Guestrin [22]. Compared with the GBDT, the XGBoost not only provides parallel computing but also significantly improves the accuracy of the algorithm; hence, it is widely used in various industries [38–40]. The XGBoost framework uses decision trees as base learners and combines them to form a strong learner. The algorithm flow is shown schematically in Figure 4.

The principle of the XGBoost algorithm is as follows.

Given a data set $D = \{(x_i, y_i) | x_i \in R^m, y_i \in R, i = 1, \dots, n\}$, the number of features in the data set is m , the number of samples is n , and the predicted value (\hat{y}_i) is the cumulative

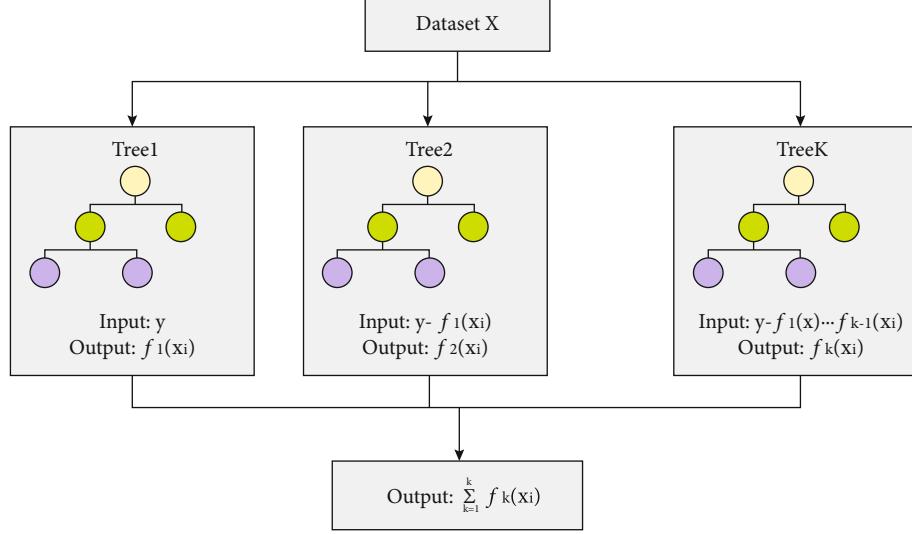


FIGURE 4: Flow chart of XGBoost framework.

value of the output results of k decision trees based on the input x_i . The formula is given by the following equation:

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F, \quad (2)$$

where $F = \{f(x) = W_{q(x)}\} (q : Rm \rightarrow T, w \in R^T)$ is the set of regression trees, CART is a common regression tree type, q represents the structure of the regression tree, T is the number of leaves in the regression tree, f_k is the input-output function of the k^{th} regression tree (with each f_k corresponding to the structure q of the k^{th} regression tree) and the leaf node weight w , and w_i is the score of the i^{th} leaf node.

In addition, the XGBoost adds a regularization term to the loss function to control the complexity of the model. The loss function is defined by the following equation:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad (3)$$

where y_i is the true value of the i -th sample; $\sum_i l(y_i, \hat{y}_i)$ is the sum of the errors of the model's predicted values for all samples; Ω is the regularization term. The formula is given by the following equation:

$$\Omega(f_k) = \gamma T_k + \frac{1}{2} \lambda \|w_k\|^2, \quad (4)$$

where T_k is the number of leaf nodes in the tree f ; $\|w_k\|$ is the normal form of the weight of the leaf nodes in the tree; γ and λ are hyperparameters, γ is the penalty coefficient, and the value range is $[0,1]$; λ is regular term coefficient.

Next, we need to find the f_k that minimizes the loss function. Different from the traditional GBDT, XGBoost optimizes the loss function and carries out the second-

order Taylor expansion of the loss function, so Eq. (3) can be approximated as the following equation:

$$L^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \quad (5)$$

where g_i represents the first derivative of $L(y_i, \hat{y}^{(t-1)})$ with respect to $\hat{y}^{(t-1)}$, and h_i represents the second derivative of $L(y_i, \hat{y}^{(t-1)})$ with respect to $\hat{y}^{(t-1)}$.

Since the prediction score of the first $t-1$ tree and the residual of y are constant, it does not affect the optimization of the objective function and can be directly removed. The loss function can be simplified as the following equation:

$$L^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t). \quad (6)$$

Define $I_j = \{i \mid q(x_i) = j\}$ as the sample set in leaf node j . Since XGBoost splits all leaf nodes in the same layer indiscriminately when splitting a leaf node, the weight value on the same leaf node is the same and can be merged, so the loss function can be transferred from the sample dimension to the leaf node dimension. The formula is defined by the following equation:

$$\begin{aligned} L^{(t)} &= \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T. \end{aligned} \quad (7)$$

Derivation of w_j based on the loss function and set the derivative to 0, the optimal weight w_j expression of the leaf node area is obtained as the following equation:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}. \quad (8)$$

The corresponding loss function is Eq. (9), which can be used as the value to evaluate the quality of a tree.

$$L^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \quad (9)$$

Let $G_i = \sum_{i \in I_j} g_i$, $H_i = \sum_{i \in I_j} h_i$, Eq. (9) can be simplified to the following equation.

$$L^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T. \quad (10)$$

3.1.2. Shapley Additive Explanations. Although many methods of machine learning, including the XGBoost framework, can generate high-precision prediction results, they are essentially black box models. As noted above, this makes the prediction process difficult for humans to intuitively understand and, hence, the credibility and practical application of the machine learning models are limited by poor interpretability [26]. To address this issue, the SHAP is employed as an interpretable algorithm based on the Shapley value or characteristic attribution value that was proposed by Lloyd Shapley in the field of game theory [30]. In this approach, all features are considered as “contributors.” For each permeability prediction sample, the model produces a prediction value, i.e., the Shapley value. The sum of contributions of all the features is the final prediction of the model [31, 32], which can be expressed as the following equation:

$$f(x) = g(z') = \varphi_0 + \sum_{i=1}^K \varphi_i z'_i, \quad (11)$$

where $f(x)$ is the predicted value of the machine learning model, g is the explanatory model, and $z' \in \{0, 1\}^k$. K is the dimension of the input feature. When the input feature x_i is observed, $z'_i = 1$, otherwise, it is 0. φ_0 is the average predicted value (baseline value) of the machine learning model on the data set, and φ_i is the contribution value of the i -th feature (Shapley value). For a single sample, the feature with the larger absolute value of the Shapley value has a greater impact on the sample’s prediction result. The positive or negative of the Shapley value reflects whether the feature makes the predicted value of the model increase or decrease, and it has a positive or negative effect on the model. The φ_i is given by the following equation:

$$\varphi_i = \sum_s \frac{|S|!(n-|S|-1)!}{n!} [f_x(S \cup \{x_i\}) - f_x(S)], \quad (12)$$

where n is the number of features. S is the feature subset that does not contain x_i , $|S|$ represents the number of elements in the set S ; $f_x(S \cup \{x_i\})$ and $f_x(S)$ are the predicted values

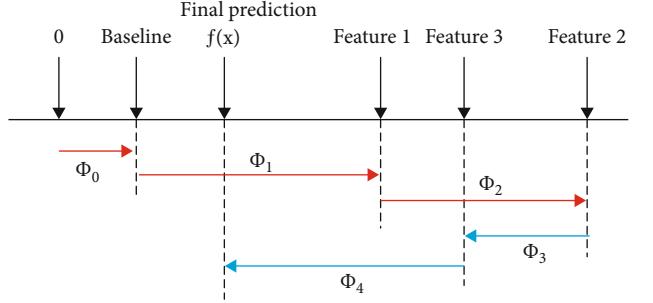


FIGURE 5: Schematic diagram of a four-feature SHAP model.

that contain feature x_i and that do not contain feature x_i , respectively.

As shown schematically in Figure 5, red indicates that the feature contribution is positive, and blue indicates that the feature contribution is negative. The SHAP method can not only rank the importance of each feature but also reveal the positive and negative influence of each feature upon the prediction result, thereby effectively combining global and local interpretability.

3.2. Particle Swarm Optimisation. The particle swarm algorithm is an intelligent search algorithm that simulates the foraging behavior of birds. Each individual is called a particle. The potential solution of the optimization problem is searched in D-dimensional space, initialized as a group of random particles, and the optimal solution is found by iteration. Each particle has 2 parameters, which can record its own speed and position. In each iteration, the particle tracks its own local optimal position or individual extreme value (P_{best}) and the group’s global optimal position or global extreme value (p_{best}) in order to update itself. This process is repeated until the global optimal solution is searched or the maximum number of iterations is reached [41, 42]. The particles update their speed and position through the following equations.

$$v_{i,j}^{t+1} = w v_{i,j}^t + c_1 r_1 (p_{i,j}^t - x_{i,j}^t) + c_2 r_2 (p_{g,j}^t - x_{i,j}^t), \quad (13)$$

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1}, \quad (14)$$

where $i = 1, 2, \dots, N$; $j = 1, 2, \dots, D$, which represents the dimension of the particle; t represents the number of iterations; $v_{i,j}$ is the velocity of the k -th iteration of the i -th particle on the j -th dimension; $x_{i,j}$ is the position of k -th iteration of the i -th particle on the j -th dimension; the individual extreme value found by the particle i is represented by $p_i = (p_{i,1}, p_{i,2}, \dots, p_{i,D})$, and the global extreme value found by the particle swarm is represented by $p_g = (p_{g,1}, p_{g,2}, \dots, p_{g,D})$; w is the inertia weight coefficient, which reflects the influence of the original speed of the particle on the current moving speed; c_1 and c_2 are the learning factors, which reflect the learning ability of the particle in its best state and the global best state, respectively; r_1 and r_2 are the random number between [0,1].

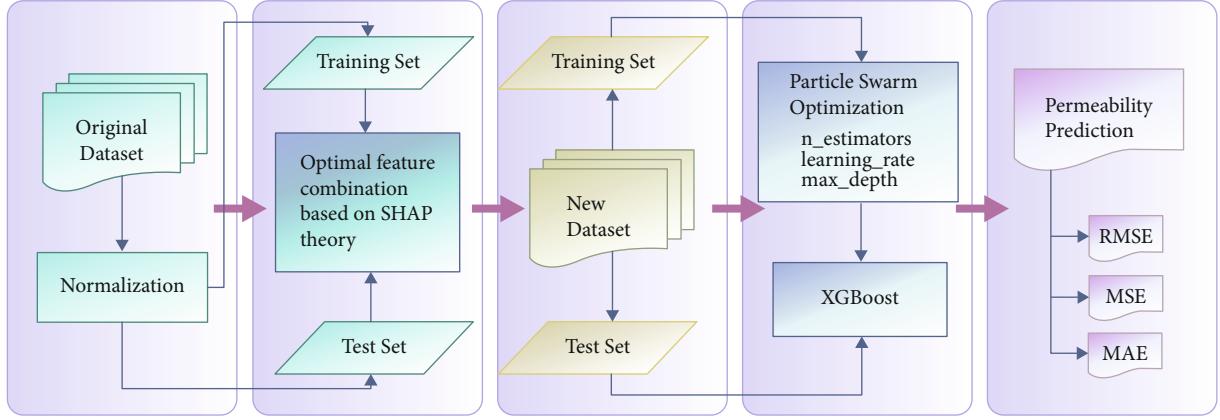


FIGURE 6: Modeling flowchart for predicting permeability.

3.3. Using the SHAP-PSO-XGBoost Framework for Estimating Permeability. The proposed modeling process of permeability prediction based on the SHAP-PSO-XGBoost is shown schematically in Figure 6. In brief, the following six primary steps are involved:

- (1) The sample data set is normalized, and the range of each feature is transformed to [0, 1]
- (2) The 202 samples in the data set are randomly divided into a training set and a test set. The training set accounts for 80%, and the test set accounts for 20% of the data set. The training set is used to train the model, and the test set is used to evaluate the performance of the model
- (3) The XGBoost model is constructed for the sample data, the grid parameters are initialized according to the grid search method, and determine the hyperparameters
- (4) Based on the established model, the SHAP attribution theory is used to analyze the factors that influence the evaluation results in terms of both the overall and individual aspects. Then, according to the global importance result, the feature combination is optimized to construct a new data set
- (5) The XGBoost model is reapplied to the new data set, and the particle swarm algorithm is used to search for the best hyperparameters of the model, including the n_estimators, learning_rate, and max_depth. The ten-fold crossover method is used for verification during the search process. That is to say, the training set is divided into ten parts, including nine training parts and one verification part, which are run in sequence. The average value is then used to express the general performance of each set of parameters, thus, avoiding any deviation due to randomly dividing the training set
- (6) After determining the optimal hyperparameters, the test set is brought into the model to determine the predicted permeability. Then, the true performance

of the model is evaluated according to the evaluation index

The experimental environment was as follows: the CPU configuration was an Intel (R) Core™ i5-3210M with 16 GB RAM and a frequency of 2.5 GHz. The programming language was Python3.7.6.

3.4. Model Evaluation Indicators. The root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) were used to evaluate the prediction performance of the proposed model [43–45]. These were calculated using Eqs. (5)–(7):

$$y_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (15)$$

$$y_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (16)$$

$$y_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (17)$$

where y_i is the actual value, \hat{y}_i is the predicted value, and N is the total number of samples.

4. Results and Discussion

4.1. Feature Selection and Interpretation. The results obtained according to the Shapley values of the features at the global level are presented in Figure 7. In Figure 7(a), the sample points for each feature are plotted against the corresponding Shapley values, with high values being indicated by the degree of redness, and low values indicated by the degree of blueness. Here, the numerical distribution variance of the Shapley value of porosity (Por) is seen to be significantly higher than that of the other features, especially for different types of sample, thus, enabling the model to distinguish between samples. Hence, this feature makes the largest contribution to the prediction of permeability. This

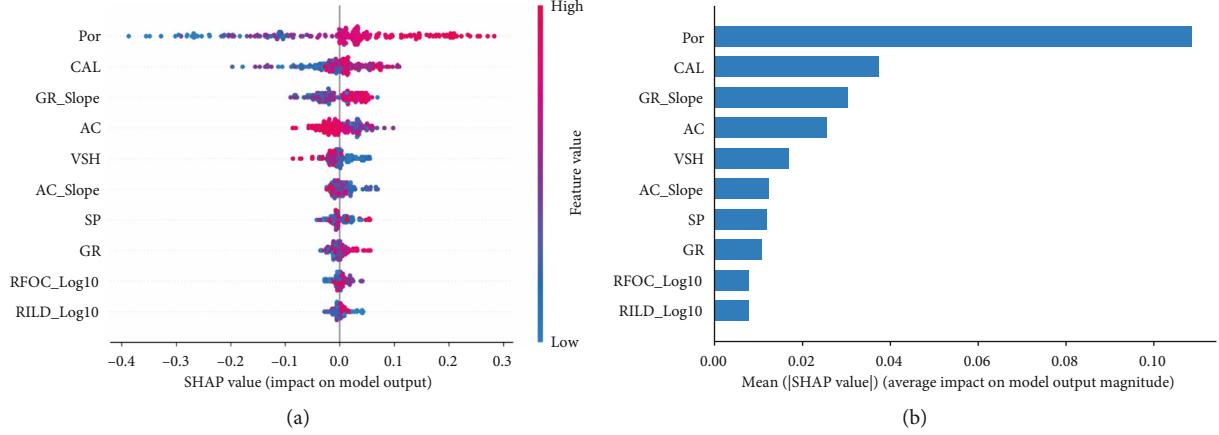


FIGURE 7: Global model explanation results by SHAP. (a) SHAP value distribution. (b) Feature importance analysis.

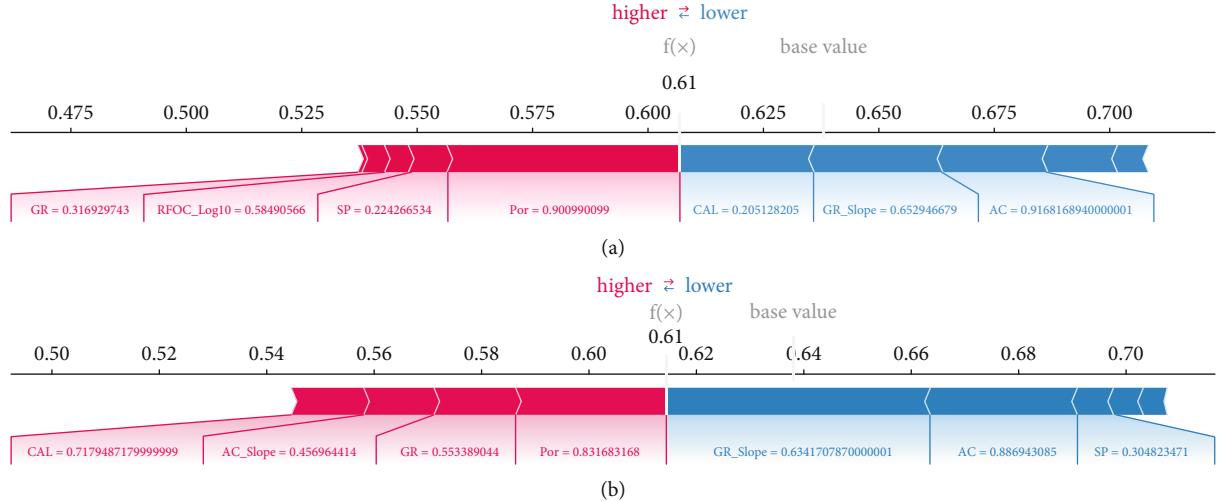


FIGURE 8: Single sample model explanation results by SHAP. (a) Sample no. 20. (b) Sample no. 25.

understanding also conforms to the laws of petrophysics. Further, the features are ranked in order of importance and are plotted against the mean of the absolute Shapley value in Figure 7(b). Here, the Por, AC, CAL, and GR slope are seen to be the four most important features, and their average absolute Shapley values are all greater than 0.02. The other features have little effect on the results and can thus be considered to have little effect on the prediction of permeability. Consequently, only the Por, AC, CAL, and GR slope values are considered hereafter as suitable input variables.

At the local level, the use of the SHAP interpretation model to predict the permeability is demonstrated for two typical samples in Figure 8, where the red part represents features with positive Shapley values, and the blue part represents features with negative values. Moreover, the quantitative influence of each feature is revealed, as a larger area indicates a larger absolute Shapley value and, hence, a greater impact upon the predicted permeability. Finally, the sum of the Shapley value of each feature plus the baseline value of the sample is the output result of the model, i.e., the predicted permeability of the sample. Taken together, the

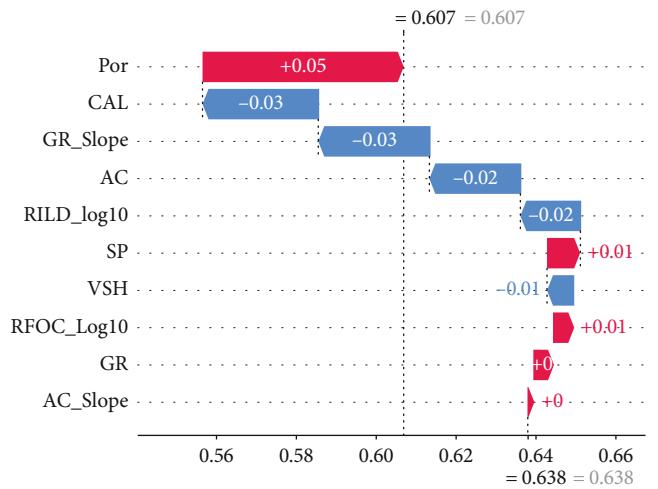


FIGURE 9: The waterfall plot for an individual case of sample no. 20.

results in Figures 7 and 8 indicate that although different features have the greatest impact on permeability prediction for different samples, and the feature with the highest global

TABLE 1: XGBoost algorithm main hyperparameter default values, search interval, and PSO optimization results.

No.	Hyperparameter	Meaning	Defaults	Search interval	Optimization results
1	n_estimators	Number of iterations	100	[50, 500]	300
2	Learning_rate	Learning rate	0.3	[0.01, 0.3]	0.1
3	Max_depth	Maximum height of the tree	6	[3, 8]	3

importance (i.e., the porosity) consistently has an important impact on the prediction results. Thus, when the positive and negative correlation of each feature in Figure 8(a) is considered for sample number 20, the Por feature is seen to increase the permeability prediction results of the model by 0.05, while the CAL, GR slope, and AC features decrease the permeability prediction by 0.03, 0.03, and 0.02, respectively. This gives a final prediction of 0.607 (Figure 9), which is close to the experimental value of 0.653.

4.2. Optimization Results. As the XGBoost framework contains multiple hyperparameters, the particle swarm optimization search method was used to automatically determine the optimal value of each of the three parameters with the greatest impact on the performance of the algorithm. The search interval of the parameters was based on the results of previous studies, the population size (N) was 25, the initial weight (w_1) and termination weight (w_2) were both 0.5, the learning factor was $c1 = c2 = 1.5$, and the maximum number of iterations (T) was 200. As detailed in Section 3.3, a 10-fold cross-validation process was used to verify the optimization results. As indicated in Table 1, the three optimized hyperparameters are quite different from the default values, thus, demonstrating the necessity of using the particle swarm optimization process. Further, the prediction results obtained using the XGBoost models before and after the hyperparameter optimization are presented in Table 2. Here, the significantly improved accuracy of the particle-swarm optimized (PSO) XGBoost model is clearly evidenced. Compared with the benchmark XGBoost framework, the PSO-XGBoost model reduces the RMSE, MSE, and MAE by 22.2%, 43.8%, and 22.3%, respectively.

4.3. Performance of the Proposed Model. To verify the validity and superiority of the interpretable PSO-XGBoost method for predicting the permeability of tight sandstone reservoirs, the results are compared with those obtained using the SVM, DT, RF, GBDT, and LightGBM approaches, each of which also uses the PSO method. The results presented in Table 3 and Figure 10 clearly confirm the accuracy and superiority of the proposed machine learning technology. Thus, the PSO-XGBoost framework has an average RMSE of 0.117, which is 25.6% lower than that of the SVR, 19.7% lower than that of the PSO-DecisionTree, 7.7% lower than that of the PSO-RandomForest, 6.8% lower than that of the PSO-GBDT, and 29.1% lower than that of PSO-LightGBM. Similarly, the average MSE of the PSO-XGBoost framework (0.016) is 50% lower than that of the SVR, 43.8% lower than that of the PSO-DecisionTree, 12.5% lower than that of both the PSO-RandomForest and the PSO-GBDT, and 62.5% lower than that of the PSO-

TABLE 2: XGBoost model error index comparison before and after PSO optimization.

Model	Mean RMSE	Mean MSE	Mean MAE
PSO-XGBoost	0.117	0.016	0.094
Benchmark XGBoost	0.143	0.023	0.115

TABLE 3: Prediction error comparison between experiment models.

Model	Mean RMSE	Mean MSE	Mean MAE
SVM	0.147	0.024	0.126
PSO_DecisionTree	0.140	0.023	0.120
PSO_RandomForest	0.126	0.018	0.101
PSO_GBDT	0.125	0.018	0.102
PSO_XGBoost	0.117	0.016	0.094
PSO_LightGBM	0.151	0.026	0.125

LightGBM. Finally, the average MAE of the PSO-XGBoost framework (0.094) is 34.0% lower than that of the SVR, 27.7% lower than that of the PSO-DecisionTree, 7.4% lower than that of the PSO-RandomForest, 8.5% lower than that of the PSO-GBDT, and 33.0% lower than that of the PSO-LightGBM. These results fully demonstrate that the proposed interpretable PSO-XGBoost exhibits the smallest permeability prediction error rate.

4.4. Performance Comparison

4.4.1. Comparison with Traditional Methods. The linear fitting relationship between the core porosity and core permeability after data normalization is presented in Figure 11 and is given by the following equation:

$$\text{Perm} = 0.5062 * \text{Por} + 0.2638. \quad (18)$$

After normalization, a multivariate regression analysis was performed using the following equation:

$$\begin{aligned} \text{Perm} = & -0.1866 * \text{AC} + 0.1376 * \text{CAL} + 0.2900 * \text{GR}_{\text{Slope}} \\ & + 0.7196 * \text{Por}. \end{aligned} \quad (19)$$

Using the results of Eqs. (8) and (9) and the test data, a 10-fold cross-validation was performed for comparative analysis, and the results are presented in Table 4. Here, the PSO-XGBoost model is seen to decrease the RMSE, MSE, and MAE by 25.6%, 50.0%, and 39.4%, respectively, compared to those of the linear regression model. Moreover, the PSO-XGBoost model decreases the RMSE, MSE, and

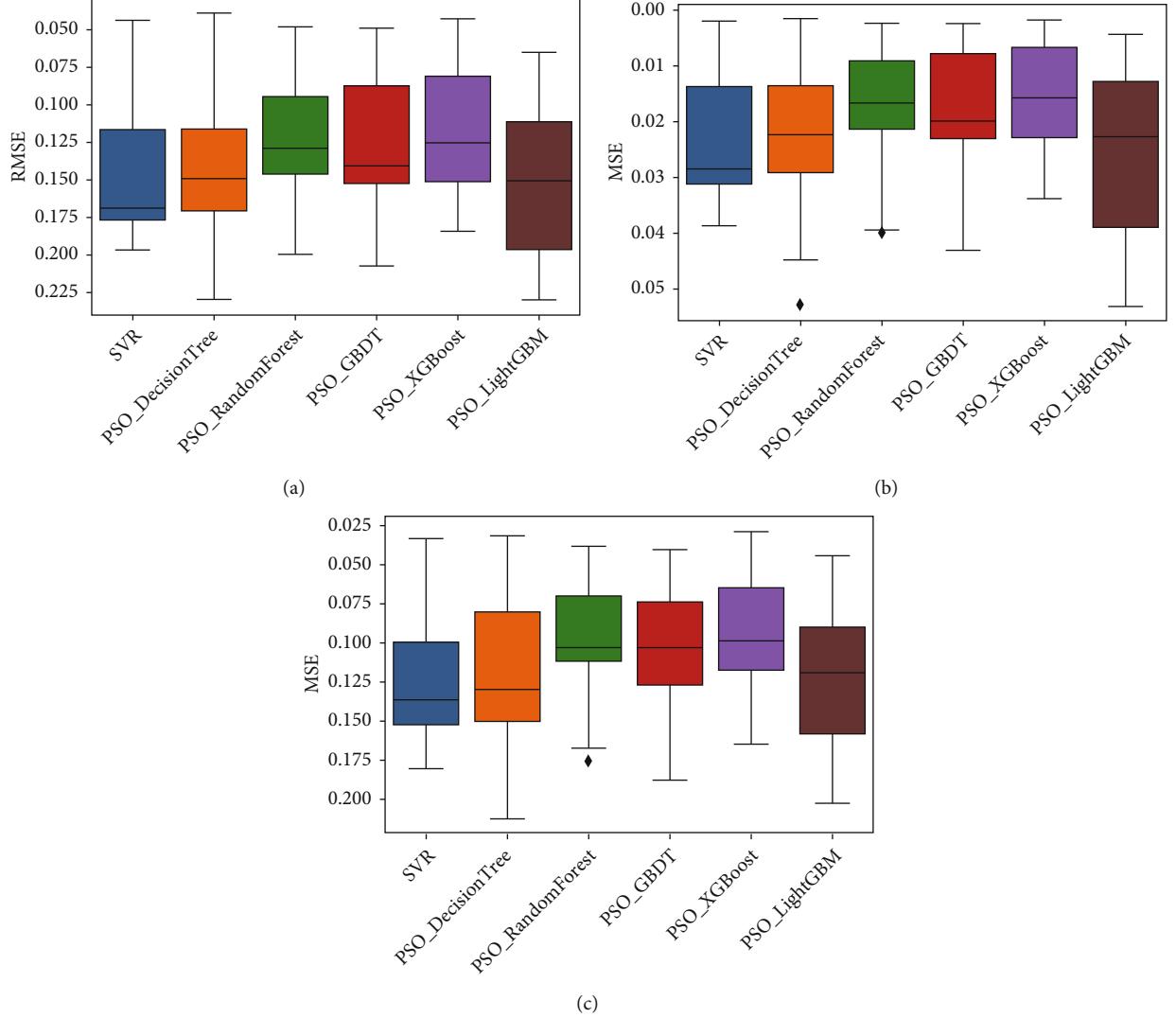


FIGURE 10: Comparison of the indicators of each machine learning model to predict permeability: (a) RMSE, (b) MSE, and (c) MAE.

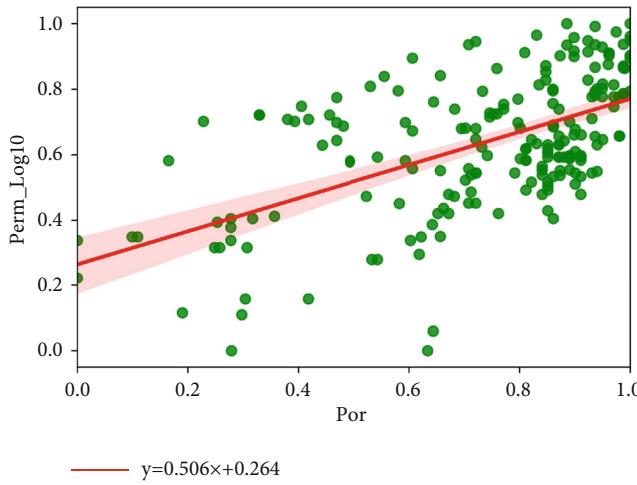


FIGURE 11: Cross-plot of the normalized core porosity and core permeability.

TABLE 4: PSO-XGBoost model and traditional permeability prediction model error index comparison.

Model	Mean RMSE	Mean MSE	Mean MAE
Linear regression	0.147	0.024	0.131
Multivariate regression	0.134	0.020	0.116
PSO-XGBoost	0.117	0.016	0.094

MAE by 14.5%, 25.0%, and 23.4%, respectively, compared to those of the multivariate regression model.

In addition, a cross-plot of the actual and predicted core permeabilities obtained by the traditional mathematical regression method and the proposed PSO-XGBoost model is presented in Figure 12. Here, it can be seen that both the training set and the test set have a good match with the 45-degree line. Therefore, it can be concluded that the interpretable PSO-XGBoost model has better prediction accuracy with respect to the permeability of tight sandstone reservoirs than do the traditional mathematical regression methods.

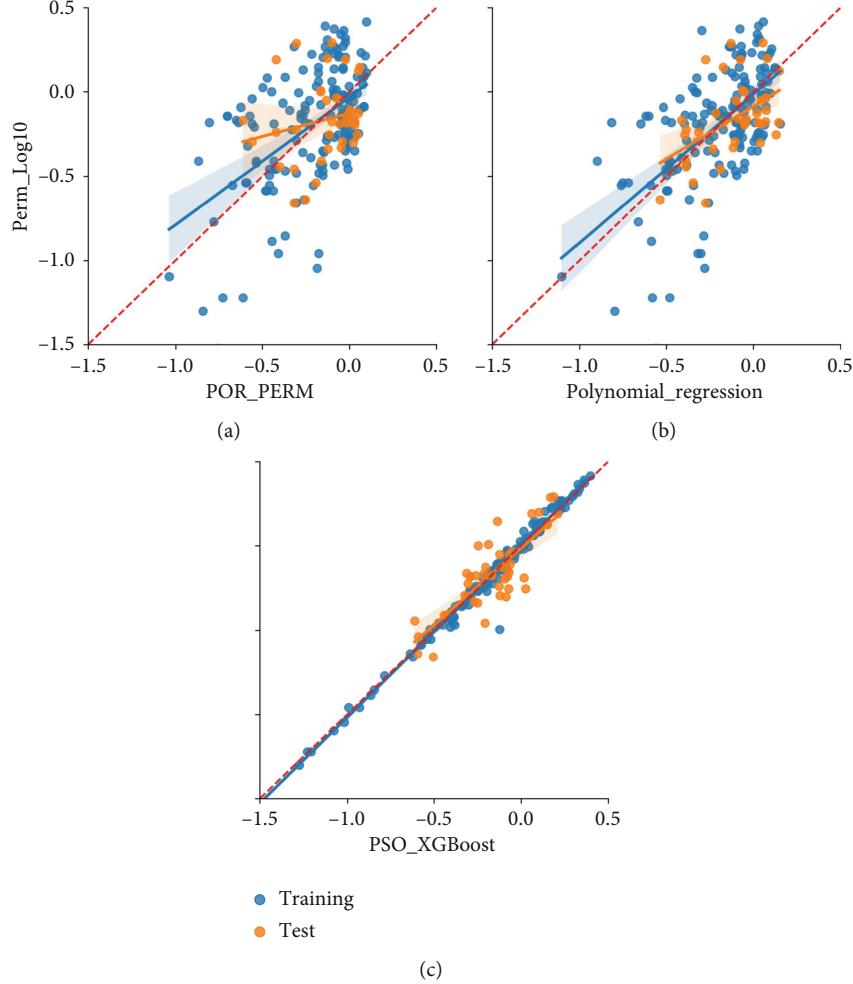


FIGURE 12: Cross-plot of the core permeability and predicted permeability by the various models: (a) linear regression, (b) multivariate regression, and (c) PSO-XGBoost.

TABLE 5: Configuration parameters of all models.

Model	Parameters
CNN	Convolutional layer with 32 filters of size (2, 2) Max pooling layer with size (2, 2) Fully connected layer with 64 neurons
LSTM	LSTM layer with 64 units Fully connected layer with 32 neurons
GRU	GRU layer with 16 units Fully connected layer with 64 neurons

4.4.2. Comparison with Deep Learning Technologies. In recent years, with the rapid development of deep learning, more and more network models have been proposed and widely used in the field of oil exploration and development. Hence, the permeability prediction performance of the proposed interpretable PSO-XGBoost model is compared with that the three main deep learning models, namely, the CNN, LSTM, and GRU [46–48]. The accuracy of each deep learning model was increased via the grid search for hyper-parameters, and the configuration parameters for each model are presented in Table 5. The 10-fold cross-

TABLE 6: Prediction error comparison between experiment models.

Model	Mean RMSE	Mean MSE	Mean MAE
CNN	0.256	0.138	0.316
LSTM	0.439	0.249	0.445
GRU	0.424	0.174	0.414
PSO_XGBoost	0.117	0.016	0.094

validation method was then used to compare and analyze the various models using the test data set, and the results are presented in Table 6. Here, the CNN, LSTM, and GRU are all seen to perform poorly in predicting the permeability of tight sandstone reservoirs, with average RMSEs that are, respectively, 2.18, 3.75, and 3.62 times that of the PSO_XGBoost model. Similarly, the average MSEs of the CNN, LSTM, and GRU are, respectively, 8.63, 15.56, and 10.88 times that of the PSO_XGBoost model, and the average MAEs are 3.36, 4.73, and 4.40 times that of the PSO_XGBoost model.

From the perspective of error indicators, there is a big gap between the three deep learning models and the various

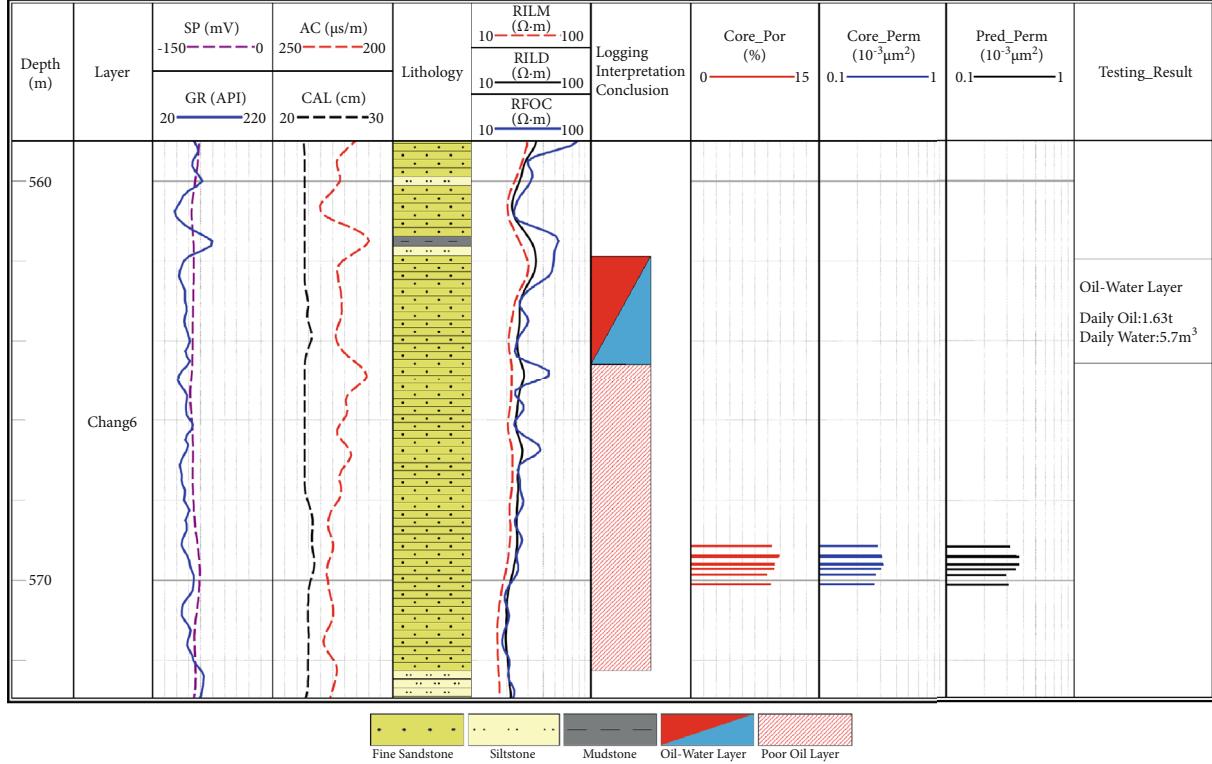


FIGURE 13: Permeability prediction of tight sandstone reservoir in Well X190.

types of machine learning model examined in previous sections (Table 3 and Figure 10), again indicating that the PSO-XGBoost gives the best performance. It is speculated that this is because the data set used in the present study was small, whereas the popular deep learning models are more suitable for larger data sets. It is difficult to obtain a good performance when there are only a few hundred data samples, and overfitting is likely. Therefore, it is suggested that machine learning algorithms should be preferred for small data sets, and that the interpretable machine learning method with the most obvious relationship between the features and the results should be selected. Thus, among all kinds of ensemble learning algorithms based on decision tree models (including the RF, GBDT, XGBoost, and LightGBM), the particle-swarm optimized XGBoost framework should be selected first.

4.5. Practical Application. Based on the high-precision permeability prediction of tight sandstone reservoirs in the study area, the interpretable PSO-XGBoost model was used to process Well X190. Well X190 is also a coring well and has been used in petrophysical experiments, but was not included in the sample data set. Hence, it can be used to test the prediction results of the PSO-XGBoost model. The results are presented in Figure 13, where Core_Por is the core porosity, Core_Permit is the core permeability, and Pred_Permit is the permeability predicted by the interpretable PSO-XGBoost model. The perforation section was 562–564.5 m, and its logging interpretation indicated an oil-water layer. The first-month oil production was 1.63 t/d,

which meets the industrial oil flow standard of 0.85 t/d in the study area. The core plugs were located in the tight sandstone reservoir 569.1–570.1 m below the perforation section, which the logging interpretation indicates as a poor oil layer, i.e., an oil layer whose petrophysical properties are worse than those of the other conventional oil layers in the tight sandstone reservoir and which therefore has lower production. The upper limit of the petrophysical properties of the poor oil layer should be the lower limit of the conventional oil layers in the tight sandstone reservoir. According to the statistical analysis, the evaluation criteria for poor oil layers in the study area are a porosity of 9–10%, a permeability of $0.3 \times 10^{-3} - 0.4 \times 10^{-3} \mu\text{m}^2$, an RILD of 17–22 Ω, an AC of at least $217 \mu\text{s}/\text{m}$, and a GR of no more than 90 API. The core permeability of the sample at 569.1–570.1 m is between 0.269×10^{-3} and $0.318 \times 10^{-3} \mu\text{m}^2$, and the predicted permeability is between 0.294×10^{-3} and $0.376 \times 10^{-3} \mu\text{m}^2$. These prediction results are consistent with the electrical characteristics of poor oil layers, thus, demonstrating the importance of the high-precision PSO-XGBoost model as an indicator for evaluating the permeability of tight sandstone reservoir types.

5. Conclusions

To improve the accuracy and interpretability of the permeability prediction of tight sandstone reservoirs, an improved XGBoost model based on the PSO algorithm and attributable interpretation was proposed herein. The following conclusions were drawn:

- (1) The SHAP can not only explain the importance of input features from a global perspective, mine the key features that affect the permeability prediction, and reduce the dimensionality of the samples according to the key features, but can also clarify the quantitative contribution of each feature towards the permeability prediction based on the evaluation results of each individual sample. This unifies the global-local interpretability of the model and improves the credibility of the prediction results
- (2) Particle swarm optimization was used to realize the automatic optimization of the hyperparameters of the XGBoost framework. The results of the computational experiments based on sample data showed that, compared with the baseline XGBoost model, the prediction error of the PSO-XGBoost model is significantly reduced, and the accuracy of the permeability prediction is higher
- (3) The interpretable PSO-XGBoost model demonstrates powerful predictive capabilities. By comparing the predictions of five types of machine learning models, two types of mathematical regression methods, and three types of deep learning models, along with the RMSE, MSE, and MAE data of the measured and predicted values, it was concluded that the PSO-XGBoost model is better in predicting the permeability of tight sandstone reservoirs. The proposed method is clearly more accurate and practical for use with small data sets and lays a solid foundation for the evaluation of tight sandstone reservoirs
- (4) In future, the proposed improved model could be further studied from the following three aspects: (i) the petrophysical data and wireline logging data from multiple areas could be selected to further test the performance and explore the generalizability of the proposed model with respect to different data sets, (ii) strategies for adding more geological constraint information, such as the sedimentary characteristics of the reservoir, lithological changes, and the relative position of the sample wells, etc., could be developed in order to further improve the accuracy of the model, and (iii) it is possible to explore the reasonable predictions of reservoir parameters and evaluation methods of high-quality reservoirs in the areas far away from the wellbore by establishing a fine three-dimensional geological model of the study area

Data Availability

The data are available in this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to thank Zichang Oil Production Plant for providing samples and data. This research was supported by the Fundamental Research Funds for the Central Universities (no. 300102278402).

References

- [1] W. Dou, L. Liu, L. Jia, Z. Xu, M. Wang, and C. Du, "Pore structure, fractal characteristics and permeability prediction of tight sandstones: a case study from Yanchang formation, Ordos Basin, China," *Marine and Petroleum Geology*, vol. 123, article 104737, 2021.
- [2] J. Shi, J. Zhang, C. Zhang, T. Jiang, and G. Huang, "Experimental investigation of permeability evolution on sandstone in triaxial and long-term dissolution experiment," *Geofluids*, vol. 2021, Article ID 5580185, 12 pages, 2021.
- [3] Y. Zhou, X. Zhao, C. Jiang, S. Liu, Z. Han, and G. Wang, "Permeability prediction of multi-stage tight gas sandstones based on Bayesian regularization neural network," *Marine and Petroleum Geology*, vol. 133, article 105320, 2021.
- [4] R. Baouche and B. S. Nabawy, "Permeability prediction in argillaceous sandstone reservoirs using fuzzy logic analysis: a case study of triassic sequences, southern Hassi R'Mel gas field, Algeria," *Journal of African Earth Sciences*, vol. 173, article 104049, 2021.
- [5] P. Yang, L. Zhang, K. Liu, B. Cao, J. Gao, and G. Qiu, "Diagenetic history and reservoir evolution of tight sandstones in the second member of the Upper Triassic Xujiahe Formation, western Sichuan Basin, China," *Journal of Petroleum Science and Engineering*, vol. 201, article 108451, 2021.
- [6] Y. Wu, H. Tang, Y. Wang et al., "Sedimentary facies of the Longmaxi formation shale gas reservoir in the Weiyuan area based on elemental characteristics," *Geofluids*, vol. 2021, Article ID 5021298, 17 pages, 2021.
- [7] L. Gan, Y. Wang, X. Luo et al., "A permeability prediction method based on pore structure and lithofacies," *Petroleum Exploration and Development*, vol. 46, no. 5, pp. 935–942, 2019.
- [8] H. P. Menke, J. Maes, and S. Geiger, "Upscaling the porosity-permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning," *Scientific Reports*, vol. 11, no. 1, p. 2625, 2021.
- [9] A. Ahangar-Asr, A. Faramarzi, N. Mottaghifard, and A. A. Javadi, "Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression," *Computers & Geosciences*, vol. 37, no. 11, pp. 1860–1869, 2011.
- [10] C. Qian, S. Yang, Y. Wang, C. Wu, and Y. Zhang, "Prediction and modeling of petrophysical parameters of deep-buried, low permeability glutinite reservoirs in Yubei area, Turpan-Hami Basin, China," *Journal of Petroleum Science and Engineering*, vol. 207, article 109154, 2021.
- [11] H. A. Nooruddin, M. E. Hossain, H. Al-Yousef, and T. Okasha, "Comparison of permeability models using mercury injection capillary pressure data on carbonate rock samples," *Journal of Petroleum Science and Engineering*, vol. 121, pp. 9–22, 2014.
- [12] A. Soleiman-zadeh, S. Parvin, and S. Kord, "Effect of overburden pressure on determination of reservoir rock types using RQI/FZI, FZI* and Winland methods in carbonate rocks," *Petroleum Science*, vol. 16, no. 6, pp. 1403–1416, 2019.

- [13] Y. Song, G. Dai, L. Zhao, Z. Bian, P. Li, and L. Song, "Permeability prediction of hydrated cement paste based on its 3D image analysis," *Construction and Building Materials*, vol. 247, article 118527, 2020.
- [14] E. H. Rios, I. Figueiredo, A. K. Moss et al., "NMR permeability estimators in 'chalk' carbonate rocks obtained under different relaxation times and MICP size scalings," *Geophysical Journal International*, vol. 206, no. 1, pp. 260–274, 2016.
- [15] I. Zahmatkesh, B. Soleimani, A. Kadkhodaie, A. Golalzadeh, and A.-M. Abdollahi, "Estimation of DSI log parameters from conventional well log data using a hybrid particle swarm optimization-adaptive neuro-fuzzy inference system," *Journal of Petroleum Science and Engineering*, vol. 157, pp. 842–859, 2017.
- [16] Q. Qi, L.-Y. Fu, J. Deng, and J. Cao, "Attenuation methods for quantifying gas saturation in organic-rich shale and tight gas formations," *Geophysics*, vol. 86, no. 2, pp. D65–D75, 2021.
- [17] H. Al Khalifah, P. Glover, and P. Lorinczi, "Permeability prediction and diagenesis in tight carbonates using machine learning techniques," *Marine and Petroleum Geology*, vol. 112, article 104096, 2020.
- [18] M. Alizamir, S. Kim, O. Kisi, and M. Zounemat-Kermani, "A comparative study of several machine learning based nonlinear regression methods in estimating solar radiation: case studies of the USA and Turkey regions," *Energy*, vol. 197, article 117239, 2020.
- [19] A. Al-Anazi and I. D. Gates, "A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs," *Engineering Geology*, vol. 114, no. 3–4, pp. 267–277, 2010.
- [20] R. Gholami, A. R. Shahraki, and M. Jamali Paghaleh, "Prediction of hydrocarbon reservoirs permeability using support vector machine," *Mathematical Problems in Engineering*, vol. 2012, Article ID 670723, 18 pages, 2012.
- [21] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," *Journal of Petroleum Science and Engineering*, vol. 174, pp. 776–789, 2019.
- [22] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, New York, United States, August 2016.
- [23] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, 2017.
- [24] D. A. Otchere, T. O. A. Ganat, R. Gholami, and M. Lawal, "A novel custom ensemble learning model for an improved reservoir permeability and water saturation prediction," *Journal of Natural Gas Science and Engineering*, vol. 91, article 103962, 2021.
- [25] Y. Gu, D. Zhang, and Z. Bao, "A new data-driven predictor, PSO-XGBoost, used for permeability of tight sandstone reservoirs: a case study of member of chang 4+5, western Jiyuan oilfield, Ordos Basin," *Journal of Petroleum Science and Engineering*, vol. 199, article 108350, 2021.
- [26] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.
- [27] R. Rodriguez-Perez and J. Bajorath, "Feature importance correlation from machine learning indicates functional relationships between proteins and similar compound binding characteristics," *Scientific Reports*, vol. 11, no. 1, p. 14245, 2021.
- [28] A. Razmjoo, P. Xanthopoulos, and Q. P. Zheng, "Online feature importance ranking based on sensitivity analysis," *Expert Systems with Applications*, vol. 85, pp. 397–406, 2017.
- [29] J.-J. Liu and J.-C. Liu, "An intelligent approach for reservoir quality evaluation in tight sandstone reservoir using gradient boosting decision tree algorithm - a case study of the Yanchang formation, mid-eastern Ordos Basin, China," *Marine and Petroleum Geology*, vol. 126, article 104939, 2021.
- [30] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777, Red Hook, United States, 2017.
- [31] M. V. García and J. L. Aznarte, "Shapley additive explanations for NO₂ forecasting," *Ecological Informatics*, vol. 56, article 101039, 2020.
- [32] S. Mangalathu, S.-H. Hwang, and J.-S. Jeon, "Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach," *Engineering Structures*, vol. 219, article 110927, 2020.
- [33] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using Shapley Additive Explanations," *Expert Systems with Applications*, vol. 186, article 115736, 2021.
- [34] K. Xi, Y. Cao, K. Liu et al., "Diagenesis of tight sandstone reservoirs in the upper Triassic Yanchang formation, southwestern Ordos Basin, China," *Marine and Petroleum Geology*, vol. 99, pp. 548–562, 2019.
- [35] H. Guo, R. He, W. Jia et al., "Pore characteristics of lacustrine shale within the oil window in the upper Triassic Yanchang formation, southeastern Ordos Basin, China," *Marine and Petroleum Geology*, vol. 91, pp. 279–296, 2018.
- [36] K. Zhang, R. Liu, Z. Liu, B. Li, J. Han, and K. Zhao, "Influence of volcanic and hydrothermal activity on organic matter enrichment in the upper Triassic Yanchang formation, southern Ordos Basin, Central China," *Marine and Petroleum Geology*, vol. 112, article 104059, 2020.
- [37] M. Sabah, M. Talebkeikhah, D. A. Wood, R. Khosravianian, M. Anemangely, and A. Younesi, "A machine learning approach to predict drilling rate using petrophysical and mud logging data," *Earth Science Informatics*, vol. 12, no. 3, pp. 319–339, 2019.
- [38] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018.
- [39] J. Nobre and R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," *Expert Systems with Applications*, vol. 125, pp. 181–194, 2019.
- [40] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access*, vol. 7, pp. 118310–118321, 2019.
- [41] H. Jiang, Z. He, G. Ye, and H. Zhang, "Network intrusion detection based on PSO-Xgboost model," *IEEE Access*, vol. 8, pp. 58392–58401, 2020.

- [42] M. A. Hossain, H. R. Pota, S. Squartini, F. Zaman, and J. M. Guerrero, “Energy scheduling of community microgrid with battery cost using particle swarm optimisation,” *Applied Energy*, vol. 254, article 113723, 2019.
- [43] A. K. Mulashani, C. Shen, B. M. Nkurlu, C. N. Mkono, and M. Kawamala, “Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data,” *Energy*, vol. 239, article 121915, 2021.
- [44] F. Male and I. J. Duncan, “Lessons for machine learning from the analysis of porosity-permeability transforms for carbonate reservoirs,” *Journal of Petroleum Science and Engineering*, vol. 187, article 106825, 2020.
- [45] M. Safaei-Farouji and A. Kadkhodaie, “Application of ensemble machine learning methods for kerogen type estimation from petrophysical well logs,” *Journal of Petroleum Science and Engineering*, vol. 208, article 109455, 2021.
- [46] Z. Zhong, T. R. Carr, X. Wu, and G. Wang, “Application of a convolutional neural network in permeability prediction: a case study in the Jacksonburg-Stringtown oil field, West Virginia, USA,” *Geophysics*, vol. 84, no. 6, pp. B363–B373, 2019.
- [47] W. Chen, L. Yang, B. Zha, M. Zhang, and Y. Chen, “Deep learning reservoir porosity prediction based on multilayer long short-term memory network,” *Geophysics*, vol. 85, no. 4, p. WA213, 2020.
- [48] Z. Zhang, Y. Wang, and P. Wang, “On a deep learning method of estimating reservoir porosity,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 6641678, 13 pages, 2021.