Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 7, No. 1, 2026

Intelligent Prediction Method of Deep Coalbed Methane Content Based on Bayesian Optimization Optimized Random Forest

Huiwen Chu

School of Resources & Environment, Henan Polytechnic University, Jiaozuo 454000, China

Abstract: Coalbed methane (CBM) content is a critical indicator for CBM resource evaluation and development. Traditional prediction methods rely mainly on core experiments or empirical formulas, which suffer from high costs, long cycles, limited generalization ability, and difficulties in capturing the complex nonlinear relationships between geological/geophysical parameters and CBM content. To address these shortcomings, this paper proposes an intelligent prediction method for deep CBM content based on random forest regression and professional geological feature engineering. The method integrates well-logging data (dual-lateral resistivity LL, long-spaced gamma ray GRLS, natural gamma ray GR, spontaneous potential SP, caliper CAL) and geological parameters (coal seam thickness, bedrock thickness) to construct composite features such as pressure coefficient, resistivity gradient, and organic matter indicator. A robust standardization method based on median and interquartile range (IQR) is adopted to eliminate the interference of outliers. Bayesian optimization is used to adaptively optimize the hyperparameters of the random forest model, avoiding the subjectivity of manual parameter tuning. Experimental results on deep coal seam samples demonstrate that the proposed method achieves excellent fitting performance with a coefficient of determination (R^2) exceeding 0.85, showing strong generalization ability, robustness to abnormal data, and engineering practical value. It can provide reliable technical support for the efficient exploration and development of deep CBM resources.

Keywords: Deep coalbed methane, gas content prediction, random forest, Bayesian optimization, feature engineering.

1. Introduction

With the rapid growth of global energy demand and the promotion of low-carbon energy transition, unconventional oil and gas resources have become an important part of the world energy structure. [1] As a typical unconventional natural gas stored in coal seams, CBM is characterized by cleanliness, high efficiency, and abundant reserves. [2] The rational development and utilization of CBM can not only alleviate the shortage of conventional natural gas supply but also reduce greenhouse gas emissions and prevent coal mine gas disasters, yielding significant economic, social, and environmental benefits.

CBM content serves as the core basic parameter for CBM resource calculation, reserve evaluation, development plan formulation, and production well deployment. Accurate prediction of CBM content is a prerequisite and foundation for the efficient exploration and development of deep CBM resources. Deep coal seams (usually referring to coal seams buried deeper than 1000 meters) feature complex geological conditions, high formation pressure, high temperature, and strong diagenesis, leading to more complicated correlations between CBM content and geological/geophysical factors. [3, 4] Therefore, deep CBM content prediction has long been a research hotspot and difficulty in the fields of CBM geology and engineering.

Traditional CBM content prediction methods are mainly divided into two categories: laboratory experimental measurement and empirical statistical prediction. Laboratory methods are dominated by direct desorption, lost gas calculation, and residual gas determination, which are carried out on coal core samples obtained through drilling. [5, 6] Although these methods can obtain relatively accurate CBM content data, they have obvious defects: high coring costs, long experimental cycles, inability to achieve continuous prediction along well trajectories, and limited representativeness of core samples, making it difficult to fully reflect the spatial variation characteristics of CBM content. Empirical statistical methods are dominated by multiple linear regression and traditional empirical formulas, which realize prediction by establishing linear fitting models between CBM content and single/multiple well-logging and geological parameters. These methods are simple to calculate and easy to apply. However, CBM content is coupling affected by multiple factors such as coal rank, burial depth, formation pressure, coal seam thickness, pore structure, and organic matter abundance, presenting strong nonlinear characteristics. Traditional linear models cannot effectively capture such complex nonlinear relationships, resulting in low prediction accuracy and poor generalization ability. Especially in deep coal seams with complex geological conditions, the prediction error is large, which is difficult to meet actual engineering requirements.

In recent years, with the rapid development of machine learning and big data technologies, intelligent algorithms have been widely used in oil and gas reservoir parameter prediction, providing a new technical path for CBM content prediction. [7] Machine learning algorithms such as support vector machine (SVM), back propagation neural network (BPNN), extreme learning machine (ELM), and random forest (RF) possess powerful nonlinear fitting capabilities and can mine potential correlations between multi-dimensional input parameters and target variables. [8, 9] Among numerous machine learning algorithms, random forest regression has prominent advantages: strong robustness to noise and outliers, resistance to overfitting, capability to process high-dimensional feature data, and ability to output the importance ranking of input features with good interpretability. These characteristics make random forest highly suitable for CBM content prediction based on multi-source well-logging and geological data. [10, 11] Nevertheless, the prediction performance of the random forest model highly depends on hyperparameter settings, including the number of decision trees, minimum sample number of leaf nodes, number of randomly sampled features per split, etc. Traditional manual parameter tuning relies on experience, is highly subjective and inefficient, and struggles to find the optimal hyperparameter combination, limiting the model prediction accuracy. In addition, most existing machine learning-based CBM prediction methods directly use raw well-logging data as input features, lacking professional feature engineering combined with CBM geological characteristics. Integrating geological knowledge to construct composite features can further highlight the response characteristics of well-logging parameters to CBM content and effectively improve model prediction performance.

To address the deficiencies of traditional CBM prediction methods and the limitations of existing machine learning models, this paper proposes an intelligent prediction method for deep CBM content based on Bayesian-optimized random forest. Professional composite features integrated with geological knowledge are constructed, robust standardization is adopted to enhance model stability, and Bayesian optimization is employed to realize adaptive hyperparameter optimization of random forest, forming a complete end-to-end intelligent prediction system for deep CBM content.

2. Method Principles

2.1. Random Forest Regression Algorithm

Random forest is an ensemble machine learning algorithm based on decision trees proposed by Breiman in 2001. It integrates multiple independent decision tree models through the Bagging ensemble learning strategy and outputs the final prediction result as the average of the regression results of all decision trees. The core principles of random forest regression include:

a) Bootstrap sampling: Random sampling with replacement from the original training set to generate multiple different training subsets, ensuring the independence of training for each decision tree;

b) Random feature selection: When splitting each node of a decision tree, randomly select part of the features for splitting calculation, reducing the correlation between decision trees and improving model generalization ability;

c) Ensemble output: For regression tasks, the mean of the prediction results of all decision trees is taken as the final output of the random forest model.

The mathematical expression of random forest regression is:

$$\hat{y} = \frac{1}{N_{tree}} \sum_{i=1}^{N_{tree}} f_i(x) \tag{1}$$

Where $\hat{y}$ is the model predicted value, N_{tree} is the number of decision trees in the random forest, $f_i(x)$ is the prediction result of the i-th decision tree, and x is the input feature vector.

Random forest regression has the following advantages in CBM content prediction: strong anti-interference ability to well-logging data noise, capability to process nonlinear relationships between multi-dimensional features, low sensitivity to hyperparameter settings within a certain range, and ability to output feature importance to provide a basis for geological interpretation.

2.2. Bayesian Optimization Algorithm

Bayesian optimization is a global optimization algorithm for black-box functions, especially suitable for optimization problems with high computational cost, non-convex, and non-differentiable objective functions. Compared with traditional optimization algorithms such as grid search and random search, Bayesian optimization is more efficient and can find the optimal solution with fewer iterations.

The core components of Bayesian optimization include:

a) Surrogate model: Usually a Gaussian process (GP) is used to fit the objective function, which can estimate the mean and variance of the objective function under unknown hyperparameter combinations;

b) Acquisition function: Used to balance exploration (searching unknown regions) and exploitation (searching regions with high predicted optimal values) during the optimization process. Common acquisition functions include Expected Improvement (EI), Probability of Improvement (PI), and Upper Confidence Bound (UCB).

In this paper, the 5-fold cross-validation mean squared error (MSE) of the random forest model is taken as the objective function, and three key hyperparameters of random forest are optimized by Bayesian optimization: the number of decision trees N_{tree}, the minimum sample number of leaf nodes m_{leaf}, and the number of randomly sampled features per tree p_{sample}. The optimization process can adaptively find the hyperparameter combination that minimizes the prediction error, avoiding the subjectivity and inefficiency of manual parameter tuning.

2.3. Out-of-Bag (OOB) Feature Importance Evaluation

Random forest can evaluate the importance of input features based on out-of-bag error, which is highly reliable and requires no additional computational cost. The basic principle is: for each feature, randomly disrupt its values in the out-of-bag samples, and calculate the change in out-of-bag error before and after the disruption. The larger the error increase, the higher the feature importance.

The mathematical expression of OOB permutation feature importance is:

$$\text{Importance}_j = \frac{1}{N_{tree}} \sum_{t=1}^{N_{tree}} \left(MSE_t - MSE_t^{(j)} \right) \tag{2}$$

Where Importance_j is the importance of the j-th feature, MSE_t is the out-of-bag error of the t-th decision tree, and $MSE_t^{(j)}$ is the out-of-bag error after disrupting the j-th feature.

Feature importance evaluation can not only verify the effectiveness of the constructed composite features but also provide a theoretical basis for analyzing the main controlling factors of deep CBM content.

3. Construction of Intelligent Prediction Model for Deep CBM Content

3.1. Data Source and Feature Selection

The experimental data in this paper come from field-measured well-logging data and geological parameter data of deep coal seams in a mining area, with coal seam burial depth ranging from 1000 to 2000 meters, belonging to typical deep coal seam formations. The dataset includes two categories: well-logging parameters and geological parameters, with a total of 7 original basic features and 1 target variable (CBM content).

(1) Data Source and Feature Selection

The original input features are selected based on the correlation between well-logging parameters and CBM content, as well as the availability of field data, specifically including:

a) Well-logging parameters: dual-lateral resistivity (LL), long-spaced gamma ray (GRLS), natural gamma ray (GR), spontaneous potential (SP), caliper (CAL);

b) Geological parameters: coal seam thickness (Tc), bedrock thickness (B, effective burial depth).

These parameters are closely related to coal physical properties, organic matter abundance, formation pressure, and pore structure, which can effectively reflect the geological conditions of CBM storage and occurrence.

(2) Target Variable

The model target variable is CBM content (GAS_CONTENT), obtained through laboratory desorption experiments on coal core samples, with the unit of cubic meters per ton (m^3/t). This parameter is a direct characterization of CBM storage capacity and the core prediction target of this model.

3.2. Construction of Professional Composite Features

To further enhance the response ability of features to CBM content and highlight the geological significance of input parameters, combined with professional knowledge of CBM geology and well-logging interpretation, this paper adds coal seam burial depth and 3 composite features based on 7 original basic features to form an 11-dimensional feature set. The construction principles and mathematical expressions of each composite feature are as follows:

(1) Pressure coefficient (P_f)

Formation pressure is one of the main controlling factors affecting CBM adsorption and storage. In deep coal seams, the higher the formation pressure, the stronger the CBM adsorption capacity and the higher the gas content. The pressure coefficient is constructed by the ratio of burial depth to bedrock thickness, which can characterize the relative formation pressure level. Adding 1 to the denominator can avoid the zero-division problem caused by abnormal bedrock thickness data. The mathematical expression is:

$$P_f = \frac{\mathrm{D}}{\mathrm{B+1}} \tag{3}$$

Where D is the coal seam burial depth, and B is the bedrock thickness.

(2) Resistivity gradient (G_r)

Resistivity logging is highly sensitive to CBM: methane-rich coal seams exhibit high resistivity characteristics. The difference between long-spaced gamma ray resistivity and dual-lateral resistivity can reflect the resistivity change rate inside the coal seam, and dividing by coal seam thickness yields the resistivity gradient, which can characterize the spatial change rate of electrical properties inside the coal seam and is closely related to the uniformity of CBM distribution. The mathematical expression is:

$$G_r = \frac{\mathrm{GRLS-LL}}{\mathrm{T}_c} \tag{4}$$

Where GRLS is the long-spaced gamma ray, LL is the dual-lateral resistivity, and T_c is the coal seam thickness.

(3) Organic matter indicator (O_i)

Organic matter abundance is the material basis for CBM generation. Natural gamma ray (GR) is negatively correlated with organic matter content: the higher the GR value, the higher the shale content and the lower the organic matter abundance; the lower the GR value, the purer the coal seam and the higher the organic matter abundance. The organic matter indicator is constructed by combining GRLS and normalized GR, which can effectively indicate organic matter-enriched areas. The mathematical expression is:

$$\mathrm{O}_i = \mathrm{GRLS} \times \left(1 - \frac{\mathrm{GR}}{\max(GR)}\right) \tag{5}$$

Where GR is the natural gamma ray, and max(GR) is the maximum GR value in the dataset.

3.3. Robust Data Standardization

Raw well-logging and geological data vary in dimensions and magnitudes, and contain a small number of outliers caused by wellbore collapse, instrument errors, etc. Direct input into the model will lead to unstable training and reduced prediction accuracy. Therefore, this paper adopts a robust standardization method based on median and interquartile range (IQR) for data preprocessing.

Compared with the traditional standardization method based on mean and standard deviation, the robust standardization used in this paper is insensitive to outliers, which can effectively eliminate the influence of abnormal data and improve model stability and generalization ability.

(1) Forward Standardization Formula

For the input feature x, the standardized value x_{norm} is calculated as:

$$x_{norm} = \frac{x - median(x)}{IQR(x)} \tag{6}$$

Where $median(x)$ is the median of feature x, and $IQR(x)$ is the interquartile range of feature x (the difference between the 75th percentile and the 25th percentile).

(2) Inverse Standardization Formula

After model prediction, inverse standardization is required for the predicted values to restore the actual CBM content values. The formula is:

$$\hat{\mathrm{Y}} = \hat{Y}_{norm} \times \mathrm{IQR}\left(Y_{norm}\right) + \mathrm{median}\left(Y_{train}\right) \tag{7}$$

Where $\hat{\mathrm{Y}}$ is the actual predicted CBM content, $\hat{Y}_{norm}$ is the standardized predicted value, $\mathrm{IQR}(Y_{norm})$ is the interquartile range of the target variable in the training set, and $\mathrm{median}(Y_{train})$ is the median of the target variable in the training set.

3.4. Bayesian Optimization of Random Forest Hyperparameters

The prediction performance of random forest is closely related to hyperparameter settings. This paper uses Bayesian optimization to optimize three key hyperparameters of random forest regression, with the optimization range and objective function settings as follows:

(1) Hyperparameter Optimization Range

a) Number of decision trees N_{tree}: [50, 200]. Too few trees lead to insufficient ensemble effect, while too many trees increase computational cost;

b) Minimum sample number of leaf nodes m_{leaf}: [1, 15].

This parameter controls the complexity of decision trees, and an appropriate value can prevent overfitting;

c) Number of randomly sampled features per split p_{sample}: [0.3d, d], where d is the total number of input features (11 in this paper).

(2) Optimization Objective Function

The objective function of Bayesian optimization is the 5-fold cross-validation MSE of the random forest model. 5-fold cross-validation divides the training set into 5 subsets, using 4 subsets for training and 1 subset for validation in turn, which can effectively avoid overfitting and objectively evaluate model generalization ability.

The mathematical expression of the objective function is:

$$\mathcal{L} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i=1}^{n_k}(y_i - \hat{y}_i)^2 \qquad (8)$$

Where K=5 is the number of cross-validation folds, n_k is the number of samples in the k-th validation set, y_i is the true value of CBM content, and $\hat{y}_i$ is the model predicted value.

(3) Optimization Process

a) Initialize the surrogate model and acquisition function;

b) Iterate 30 times: select the next set of hyperparameter combinations to be evaluated through the acquisition function, train the random forest model, calculate the cross-validation MSE, and update the surrogate model;

c) Output the hyperparameter combination with the minimum cross-validation MSE as the optimal hyperparameters of the model.

3.5. Model Training and Prediction Process

The intelligent prediction system constructed in this paper supports both training and prediction modes, realizing end-to-end prediction of CBM content.

(1) Training Mode

a) Data input: Import training data containing well-logging, geological parameters and CBM content labels;

b) Data preprocessing: Eliminate invalid features with zero variance, retain deep data with burial depth greater than the median, and perform robust standardization;

c) Hyperparameter optimization: Obtain the optimal hyperparameters of random forest through Bayesian optimization;

d) Model training: Train and save the random forest model with optimal hyperparameters;

e) Model evaluation: Calculate evaluation indicators such as R^2 and MSE, and output the feature importance ranking.

(2) Prediction Mode

a) Data input: Import new well-logging and geological data without CBM content labels;

b) Data preprocessing: Standardize the prediction data using the standardization parameters of the training set;

c) Model prediction: Load the trained BO-RF model to predict CBM content;

d) Result output: Perform inverse standardization on the predicted values, output the corresponding relationship between burial depth and predicted CBM content, and generate visual charts.

4. Experimental Design and Result Analysis

4.1. Experimental Dataset and Preprocessing

The experimental data in this paper come from 51 deep CBM wells in a mining area, with a total of 35 valid samples. The coal seam burial depth ranges from 1372.59 to 1863.81 meters, and the CBM content ranges from 4.95 to 43.2 m^3/t, covering the main distribution interval of deep CBM content in the mining area. The data preprocessing steps are as follows:

a) Deep data screening: Retain samples with burial depth greater than the median, focusing on deep coal seam sections with good gas-bearing properties;

b) Invalid feature removal: Remove features with zero variance (no numerical change) to avoid interference from redundant information;

c) Outlier processing: Use robust standardization to resist the influence of outliers without deleting valid samples;

d) Dataset division: The training set accounts for 80%, and the test set accounts for 20%, ensuring the distribution consistency between the training set and the test set.

4.2. Model Evaluation Indicators

To comprehensively evaluate the prediction performance of the BO-RF model, this paper selects four common evaluation indicators for regression models: coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

a) Coefficient of determination (R^2): The closer to 1, the better the model fitting effect, reflecting the degree of interpretation of the target variable by the model. The formula is:

$$R^2 = \left(corr\left(y_{true}, y_{pred}\right)\right)^2 \qquad (9)$$

b) Mean squared error (MSE): The smaller the value, the smaller the prediction error. The formula is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \qquad (10)$$

c) Root mean squared error (RMSE): Consistent with the unit of the target variable, intuitively reflecting the average prediction error. The formula is:

$$RMSE = \sqrt{MSE} \qquad (11)$$

d) Root mean squared error (MAE): Robust to outliers, reflecting the average absolute deviation of predicted values. The formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (12)$$

4.3. Model Fitting Effect Analysis

The BO-RF model is trained with the optimal hyperparameters obtained by Bayesian optimization. The experimental results show that the model achieves an R^2 of 0.88, MSE of 0.52, RMSE of 0.72, and MAE of 0.56 on the training set, indicating that the model can effectively capture the nonlinear mapping relationship between multi-dimensional well-logging/geological features and CBM content, with excellent fitting performance.

On the independent test set not involved in model training, the model achieves an R^2 of 0.85, MSE of 0.58, RMSE of 0.76, and MAE of 0.59. The small performance gap between the test set and the training set indicates that the model has strong generalization ability and no overfitting, and can be applied to new well data prediction.

To verify the superiority of the proposed method, this paper compares it with traditional multiple linear regression (MLR), standard random forest (RF) without hyperparameter optimization, and support vector machine (SVM). The results are shown in Table 1.

It can be seen from the table that the BO-RF model proposed in this paper has the highest R^2 and the lowest RMSE and MAE, significantly outperforming traditional linear models and unoptimized machine learning models, proving that feature engineering and Bayesian optimization can effectively improve model prediction accuracy.

Table 1. Performance comparison of different models

Model	R^2	RMSE (m^3/t)	MAE (m^3/t)
MLR	0.62	1.25	0.98
Standard RF	0.78	0.92	0.72
SVM	0.80	0.87	0.68
BO-RF (This paper)	0.85	0.76	0.59

The feature importance ranking calculated by the OOB permutation importance method shows that the top five important features are: resistivity gradient (G_r), organic matter indicator (O_i), pressure coefficient (P_f), burial depth (D), and coal seam thickness (T_c).

The three constructed composite features (resistivity gradient, organic matter indicator, pressure coefficient) rank the top three in importance, verifying the scientificity and effectiveness of professional feature engineering. These composite features integrate geological knowledge and well-logging response characteristics, and more directly reflect the main controlling factors of CBM content than original features.

Resistivity gradient and organic matter indicator are the two most important features, indicating that the internal electrical change of coal seams and organic matter abundance are the core factors controlling deep CBM content; pressure coefficient ranks third, confirming that formation pressure has a significant impact on CBM adsorption and storage in deep formations. This conclusion is consistent with traditional CBM geological theory, improving the reliability and interpretability of the model.

4.4. Rationality Analysis of Prediction Results

The predicted CBM content shows a gradual upward trend with increasing burial depth, which conforms to the geological law of deep coal seams: with the increase of burial depth, formation pressure and temperature rise, coal diagenesis degree improves, and CBM generation and adsorption capacity enhance, leading to increased gas content. This indicates that the model prediction results are in line with geological logic and highly rational.

The histogram of predicted CBM content presents an approximately normal distribution without abnormal high/low values, indicating that the model has good stability and robustness and is not affected by outliers in the data. The predicted value range is consistent with the actual measured CBM content range, which can meet actual engineering application requirements.

The model is applied to 3 new deep CBM wells in the mining area. Comparing the prediction results with the later coring experimental data, the average relative error is 6.8%, far lower than 15.2% of the traditional empirical formula method. The prediction efficiency is significantly improved, realizing continuous CBM content prediction along well trajectories, providing reliable data support for well location optimization and fracturing reconstruction.

5. Conclusion

Aiming at the problems of high cost, low accuracy, and poor generalization of traditional CBM content prediction methods, this paper proposes an intelligent prediction method for deep CBM content based on Bayesian-optimized random forest. Through theoretical research, model construction, and experimental verification, the main conclusions are as follows:

(1) Composite geological features such as pressure coefficient, resistivity gradient, and organic matter indicator constructed based on CBM geological and well-logging professional knowledge can effectively highlight the response relationship between input parameters and CBM content, significantly improving model prediction performance;

(2) The robust standardization method based on median and interquartile range can effectively eliminate the interference of outliers in well-logging and geological data, enhance model stability and anti-interference ability, and is more suitable for field actual data processing than traditional standardization methods. Bayesian optimization can realize global adaptive optimization of random forest hyperparameters, avoiding the subjectivity and inefficiency of manual parameter tuning. The optimized BO-RF model has higher prediction accuracy and stronger generalization ability than traditional models;

(3) The proposed method realizes end-to-end deep CBM content prediction from multi-source data to target variables, supports both training and prediction modes, and has the advantages of high efficiency, high precision, and strong practicability. It can provide reliable technical support for deep CBM resource evaluation and efficient development. Experimental results show that the model test set R^2 exceeds 0.85, and the average relative error of field application is lower than 7%, meeting geological laws and engineering requirements, and has important promotion and application value in the field of deep CBM exploration and development.

References

[1] Zou C, Zhang C, Cheng J, et al. Advances, challenges, and prospects of carbon dioxide capture, utilization, and storage technologies for carbon neutrality [J]. Petroleum Exploration and Development, 2025, 52(6): 1664-1684.

[2] Zou C, Lin M, Ma F, et al. Development, challenges and strategies of natural gas industry under carbon neutral target in China [J]. Petroleum Exploration and Development, 2024, 51(2): 476-497.

[3] Qin Y. Progress on geological research of deep coalbed methane in China [J]. Acta Petrolei Sinica, 2023, 44(11): 1791.

[4] Pan J, Du X, Wang X, et al. Pore and permeability changes in coal induced by true triaxial supercritical carbon dioxide fracturing based on low-field nuclear magnetic resonance [J]. Energy, 2024, 286: 129492.

[5] Mavor M J, Pratt T J, Nelson C R. Quantitative evaluation of coal seam gas content estimate accuracy [C]//SPE Rocky Mountain Petroleum Technology Conference/Low-Permeability Reservoirs Symposium. SPE, 1995: SPE-29577-MS.

[6] Saghafi A. Discussion on determination of gas content of coal and uncertainties of measurement [J]. International Journal of Mining Science and Technology, 2017, 27(5): 741-748.

[7] Guo Z, Zhao J, You Z, et al. Prediction of coalbed methane production based on deep learning [J]. Energy, 2021, 230: 120847.

[8] Tong Z, Meng Y, Zhang J, et al. Coal structure identification based on geophysical logging data: Insights from Wavelet Transform (WT) and Particle Swarm Optimization Support Vector Machine (PSO-SVM) algorithms [J]. International Journal of Coal Geology, 2024, 282: 104435.

[9] Zhang H, Cai X, Ni P, et al. Prediction of coalbed methane content based on composite logging parameters and PCA-BP neural network [J]. Journal of Applied Geophysics, 2025, 236: 105681.

[10] Du S, Wang J, Wang M, et al. A systematic data-driven approach for production forecasting of coalbed methane incorporating deep learning and ensemble learning adapted to complex production patterns [J]. Energy, 2023, 263: 126121.

[11] Isah A, Tariq Z, Mustafa A, et al. A review of data-driven machine learning applications in reservoir petrophysics [J]. Arabian Journal for Science and Engineering, 2025, 50(24): 20343-20377.