energy&fuels

pubs.acs.org/EF

Article

# Machine Learning Algorithm to Predict Methane Adsorption Capacity of Coal

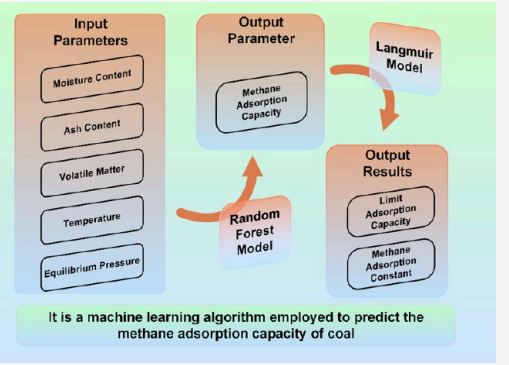
Wenshuo Li, Wei Li,* Andreas Busch, Liang Wang, Ferian Anggara, and Shilong Yang

Cite This: *Energy Fuels* 2024, 38, 23422−23432

Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** Accurately predicting methane adsorption capacity in coal is crucial for assessing coalbed methane resources and ensuring safe extraction. Conventional methane isotherm adsorption experiments often suffer from human error and experimental artifacts, leading to inaccurate and poorly reproducible outcomes. Furthermore, they are time-consuming to conduct, requiring specific and well calibrated experimental equipment. In this paper, a Random Forest (RF) algorithm is introduced to improve the accuracy and reliability of methane adsorption capacity prediction. Approximately 200 sets of experimental data, including parameters such as experimental temperature, equilibrium pressure, moisture, ash content, and volatile matter of coal samples, were collected and analyzed to establish a prediction model based on the RF algorithm. The robustness and reliability of the model were validated using K-Fold cross-validation and hyperparameter optimization. The results indicate that the Random Forest algorithm performs exceptionally well in predicting methane adsorption capacity, with optimal values for mean squared error (MSE) and the coefficient of determination ($R^2$), demonstrating a high correlation between predicted and actual values. Machine learning algorithms are innovatively combined with traditional experimental methods in this study. By training the model using large data sets, issues of error and reproducibility in traditional experiments are addressed, improving experimental efficiency and providing a more reliable method for evaluating coalbed methane resources. To some extent, the method can replace traditional methane isotherm adsorption experiments in coal, improving prediction accuracy and efficiency and demonstrating promising prospects for wide application.



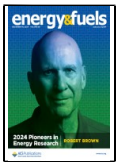



## 1. INTRODUCTION

The methane adsorption capacity in coal can indirectly assess the methane reserves in coal seams, which helps improve coalbed methane extraction techniques and increases gas recovery efficiency.[1] The Langmuir volume ($V_L$) and Langmuir pressure ($P_L$) are used to evaluate the methane adsorption capacity of coal,[2,3] thereby estimating the methane reserves in coal seams, which is crucial for preventing gas outbursts and explosions, playing a significant role in the assessment and safe extraction of coalbed methane resources.[4,5] Methane isothermal adsorption experiments are important methods for determining the limit adsorption capacity and adsorption constants of methane in coal. However, the results are often influenced by random errors and human factors, leading to poor reproducibility of the data.[6,7] Currently, the common methane isothermal adsorption experimental methods include the manometric and the gravimetric methods. The manometric method measures the adsorption capacity of coal samples at different pressures using the equation of state. It is simple to operate but can be affected by excess sorption volume. On the other hand, the gravimetric method measures the mass change of coal samples at different pressures using a high-precision balance, which avoids the influence of residual volume but requires higher standards for experimental equipment. Additionally, methane isothermal adsorption experiments are time-consuming, typically taking 2 to 3 days to complete a single set of experiments. To address these challenges, many effective methods have been proposed. Lei[8] developed a correction model to minimize errors by measuring the excess sorption volume of the adsorption tank using the vacuum He-charging method. Qiu[9] designed an automatic measurement device that determines the mass difference before and after adsorption equilibrium using the gravimetric method, thereby reducing human error.

With the development of big data and deep learning, machine learning algorithms such as KNN, DT, RF, and SVM have made significant advances in fields like image recognition, natural language processing, medical diagnostics, or autono-

**Received:** October 9, 2024
**Revised:** November 21, 2024
**Accepted:** November 25, 2024
**Published:** December 1, 2024






https://doi.org/10.1021/acs.energyfuels.4c04906
*Energy Fuels* 2024, 38, 23422−23432

mous driving.[10−12] By processing large volumes of data, these algorithms have greatly enhanced prediction accuracy and operational efficiency, while reducing the need for human intervention. By applying machine learning algorithms to coal methane adsorption experiments, models can be trained using large data sets to predict methane adsorption in coal.[13] Among various algorithms, Random Forest (RF) performs exceptionally well in predicting large data sets.[14,15] RF reduces the overfitting issue of individual decision trees by performing decision classification and regression based on the results of multiple decision trees. Additionally, it provides high robustness and excellent parallel processing capabilities. Many machine learning algorithms, including RF, are widely used in the coal and natural gas resource extraction field.[16] For instance, Zihao[17] employed the RF algorithm to identify coal structures, while Zheng[18] studied XGBoost parameter optimization based on grid search and metaheuristic algorithm to obtain the optimal parameter combination for outstanding prediction. Mohsen Tavakolian[19] and colleagues simulated methane and carbon dioxide adsorption capacities in tight reservoirs using various machine learning algorithms, including ANN, DT, and RF, and found that RF demonstrated advantages in adsorption simulations.

When using machine learning algorithms to predict methane adsorption in coal, identifying and understanding the key parameters that influence adsorption is crucial. Research has demonstrated that methane adsorption capacity in coal is significantly influenced by factors such as moisture content, ash content, volatile matter, and temperature. As moisture content increases, methane adsorption capacity and saturation values decrease, though the rate of decline gradually slows. Additionally, the influence of inherent moisture on methane adsorption varies according to coal rank.[20−22] Ash content, largely composed of mineral impurities, impacts adsorption since higher ash content reduces the effective specific surface area and pore volume of coal due to a greater proportion of inorganic components, thereby decreasing methane adsorption capacity.[23] The effect of volatile matter on methane adsorption exhibits a "U-shaped" trend. In lower ranks of coal, both volatile matter and adsorption capacity decrease as the degree of coalification increases. However, in highly metamorphosed coals, adsorption capacity gradually increases as volatile matter decreases.[23,24] Temperature also plays a role in methane adsorption, with rising temperatures leading to a reduction in adsorption capacity—a trend that tends to level off at higher temperatures.[22,25−28] Based on the above research findings, machine learning algorithm will be employed in this study to predict methane adsorption capacity using industrial analysis results of coal samples (moisture content, ash content, and volatile matter), temperature, and pressure as input parameters. The aim is to achieve a more accurate estimation of the methane adsorption capacity of coal samples, thereby providing a scientific basis for the development and safe extraction of coalbed methane resources.

Combining machine learning algorithms with methane isotherm adsorption experiments in coal may increase resource consumption, but it enhances prediction efficiency while ensuring accuracy. Furthermore, the application of machine learning algorithms can partially replace traditional methane isotherm adsorption experiments. By determining the industrial analysis parameters of coal samples and using pretrained models, isotherm adsorption curves can be generated to accurately assess the methane adsorption capacity of the coal samples. This approach not only improves experimental efficiency but also provides a more comprehensive understanding of the adsorption characteristics of coal samples, offering new methods and possibilities for research and applications in related fields.

## 2. RELEVANT THEORY

**2.1. Methane Isotherm Adsorption Experiments.** Block coal samples were collected on-site, sealed, and transported to the laboratory, where they were processed to the required particle size for experimentation. Basic petrophysical properties of the three coal samples were determined, including moisture, ash and volatile matter content. The volatile matter content can indirectly indicate the degree of coal maturity.[29] Methane adsorption and desorption tests were conducted on coal samples A, B, and C using an adsorption−desorption test system. Prior to testing, the samples were dried in a constant-temperature oven at 50 °C for 48 h. Leak tightness was also checked, and the samples were degassed in a vacuum environment. After degassing, the temperature was reduced to 30 °C to conduct the adsorption experiments. Using the above experimental methods,[29] methane adsorption capacity and equilibrium pressures for the three coal samples were measured. Scatter plots and Langmuir fitting curves were then generated. Based on the Langmuir fitting curves, maximum methane adsorption capacity $V_L$ (Langmuir volume) and the adsorption constant $K_L$ (the reciprocal of the Langmuir pressure $P_L$, or $1/P_L$) were calculated.

**2.2. Langmuir Adsorption Model.** The Langmuir adsorption model assumes a monolayer adsorption structure and is proposed based on molecular motion theory. Equation 1 provides the relationship between the adsorption pressure of methane on coal and the adsorption amount.[30−32]

$$Q_c = \frac{V_L K_L p}{1 + K_L p}$$
$$K_L = 1/p_L \quad (1)$$

where $Q_c$ is the adsorption capacity of coal per unit mass, $m^3/t$, $V_L$ is the Langmuir volume, $m^3/t$, and $K_L$ is the Langmuir adsorption constant, equivalent to the reciprocal of the Langmuir pressure $1/P_L$, 1/MPa.

In this experiment, the Langmuir equation was used, and eq 1 was converted into a linear form as shown in eq 2. A linear fit was then performed on the methane adsorption capacity and equilibrium pressure data obtained from the experiments. By calculating the parameters of the linear fit, $V_L$ and $K_L$ have been determined.

$$\frac{p}{Q_c} = \frac{p}{V_L} + \frac{1}{V_L K_L} \quad (2)$$

## 3. METHODS

**3.1. Data set Configuration.** The data set for predicting gas adsorption in coal comprises approximately 114 experiments with a total of around 1000 data points.[33−42] The data are organized and analyzed based on coal sample types,[43] and methane adsorption measurement methods, as summarized in Table 1. Analysis indicates a uniform distribution of coal sample types and formation periods, with the volumetric method primarily used for gas adsorption experiments. Data set parameters include experimental temperature, equilibrium pressure, moisture content, ash content, volatile matter, and their corresponding gas adsorption values. Experimental temperature,


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

**Table 1. Input Dataset Used for the Machine Learning Algorithm**

| types of coals | | | experimental methods | |
|---|---|---|---|---|
| types | volatile $V_{daf}$ (%) | number | methods | number |
| anthracite WY | ≤10 | 37 | gravimetric method | 4 |
| bituminous coal YM | 10−20 | 29 | | |
| | 20−28 | 21 | manometric method | 110 |
| | 28−37 | 19 | | |
| lignite HM | ≥37 | 8 | | |

equilibrium pressure, moisture content, ash content, and volatile matter are used as the model's input parameters, with gas adsorption values as the predicted output.

After determining the input and output variables of the model, data preprocessing is required for the data set to ensure data quality and improve model accuracy. The data preprocessing steps in this study mainly include the following:

*3.1.1. Data Cleaning.* For missing values in the data set, mean imputation is applied. Assuming a feature $X$ has missing values $X_s$, the imputed value can be calculated as follows:

$$\hat{X}_s = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3}$$

where $n$ represents the number of nonmissing values in feature $X$. For extreme outliers, the interquartile range ($IQR$) method is used for removal. By calculating the quartiles $Q_1$ and $Q_3$, the outlier range is defined as

$$\text{range} = [Q_1 - 1.5 \times \text{IQR},\ Q_3 + 1.5 \times \text{IQR}] \tag{4}$$

where IQR = $Q_3 - Q_1$ is the interquartile range. Values falling outside this range are considered outliers and are removed.

*3.1.2. Feature Scaling.* To eliminate the impact of differences in feature scales, all numerical features are standardized. Standardization is done using the Z-score method, which transforms feature $X$ into the standardized value $Z$:

$$Z = \frac{X - \mu}{\sigma} \tag{5}$$

Where $\mu$ and $\sigma$ are the mean and standard deviation of feature $X$, respectively. This process sets the mean of the feature values to 0 and the standard deviation to 1, thereby improving the convergence speed of the model during training.

**3.2. Random Forest (RF) Algorithm.** Random forests,[44] proposed by Leo Breiman in 2001, are an ensemble learning algorithm. The core of the random forest algorithm involves utilizing Bootstrap sampling with replacement to randomly select multiple subsets. Each subset trains a decision tree, and the prediction results of these decision trees are combined through computations to produce the final prediction of the Random Forest. In classification algorithms, decision trees are constructed using split criteria such as Gini Index or Information Gain. The random forest classification algorithm derives its prediction by taking the mode of the results from individual decision trees. In regression algorithms, the Mean Square Error (MSE) or Mean Absolute Error (MAE) of the model is usually calculated as the splitting criterion for constructing decision trees, and the average result of each decision tree is calculated as the random forest regression prediction result as shown in Figure 1.

Bootstrap sampling[45] is a resampling method primarily used to estimate the distribution of sample statistics, evaluate model performance, or assess the reliability of statistical estimates. In the random forest algorithm, for each decision tree, $n$ samples are randomly drawn with replacement from the original data set $D$ to form a Bootstrap sample set $D_b$. Decision trees are then trained using these generated Bootstrap sample sets $D_b$.

The Gini index is used in classification algorithms to measure the impurity of a data set. A lower Gini index indicates a purer node.

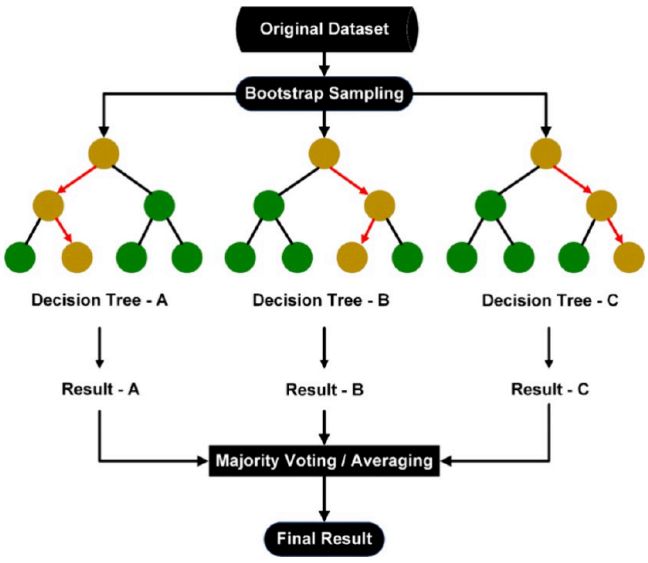



**Figure 1.** Schematic diagram of the Random Forest regression algorithm.

$$Gini(D) = 1 - \sum_{i=1}^{C} p_i^2 \tag{6}$$

where $p_i$ represents the proportion of class $i$ in data set $D$ and $C$ denotes the total number of classes.

Information gain is used to measure the change in entropy before and after splitting. In classification algorithms, a higher information gain indicates that the data set becomes purer after the split.

$$IG(D, A) = \text{entropy}(D) - \sum_{v \in \text{values}(A)} \frac{|D_v|}{|D|}\text{entropy}(D_v) \tag{7}$$

where $D$ is the original data set, $A$ is the splitting feature, $v$ is a value of feature $A$, $D_v$ is the subset of $D$ where feature $A$ has value $v$, $|D|$ is the size of data set $D$, and $Entropy(D)$ is the entropy of data set $D$.

Mean square error (MSE) is used in regression algorithms to minimize the variance of target values within nodes. Meanwhile, Mean Absolute Error (MAE) is used to minimize the absolute error of target values within nodes.[45]

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{8}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \bar{y}| \tag{9}$$

where $y_i$ represents the target value of sample $i$, $\bar{y}$ is the mean value of the target values, and $N$ is the number of samples.

**3.3. K-Fold Cross Validation.** K-fold cross-validation is a common model validation method, the core of which is to maximize the utilization of training data while minimizing the impact of validation data, in order to obtain more reliable and robust model evaluation results, reduce training bias caused by data set partitioning, and improve the generalization ability of the model. The value of $k$ affects the model evaluation accuracy, computation results bias, and the distribution of data. Therefore, selecting an appropriate $k$ value determines whether the model training can achieve the expected effectiveness.

Figure 2 illustrates its principle, with specific steps as follows: Given the original data set $D$ containing $n$ samples, i.e., $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, the data set is divided into $k$ similarly sized subsets. Subset $D_i$ is used as the testing set, and the remaining $k-1$ subsets are combined as the training set, i.e.:


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

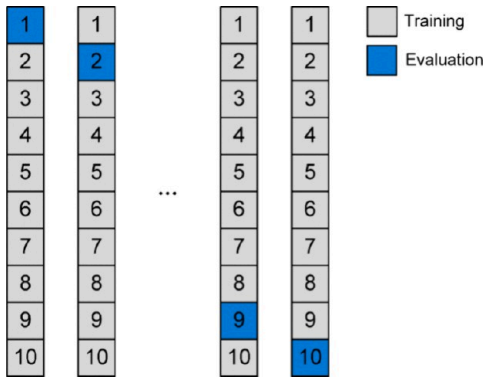



**Figure 2.** Diagram of k-fold cross-validation.

$$\begin{cases} T_k = D_1 \cup D_2 \cup \cdots \cup D_{k-1} \\ V_k = D_k \end{cases} \tag{10}$$

After training, we calculate the Mean Squared Error (MSE) and $R^2$ score for models with different $k$ values. We then compare these evaluations to determine the optimal $k$ value for model training.

**3.4. Evaluation Methods.** To assess the training effectiveness of the regression model predicting gas adsorption, predictions need to be made on the validation and test sets after training the model, followed by computing the loss values. The evaluation loss function used in the random forest regression algorithm is Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).[46] The principle of MSE is to calculate the mean of the sum of squares of the distance difference between the predicted value and the true value, in order to characterize the magnitude of the error loss value. At the same time, MSE can amplify the error values generated after model training, thus highlighting the impact of outliers. RMSE, the square root of MSE, reflects the standard deviation between predicted and actual values, aiming to standardize the units of original data and errors for intuitive understanding of prediction errors.

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{11}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{12}$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{13}$$

where $y_i$ represents the target value of sample $i$, $\overline{y}$ is the mean value of the target values, and $m$ is the number of samples.

Using evaluation loss functions can visually represent the actual performance of the model after training, but they do not characterize the goodness of fit between the actual and the predicted data points. To address this, the coefficient of determination $R^2$ (R-squared) is introduced. A higher $R^2$ value indicates a stronger explanatory power of the model.

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

## 4. RESULTS AND DISCUSSION

**4.1. Data Correlation and Distribution.** Using approximately 1000 data points from the data set, a heatmap of Pearson correlation coefficients was plotted among six parameters: experimental temperature, equilibrium pressure, moisture content, ash content, volatile matter, and gas adsorption. As shown in Figure 3, it can be observed from the heatmap that the correlation coefficients between the model's output gas adsorption and the input parameters equilibrium pressure and volatile matter are 0.48 and 0.38, respectively. During the coal maturation process, the relationship between volatile matter and adsorption capacity exhibits a

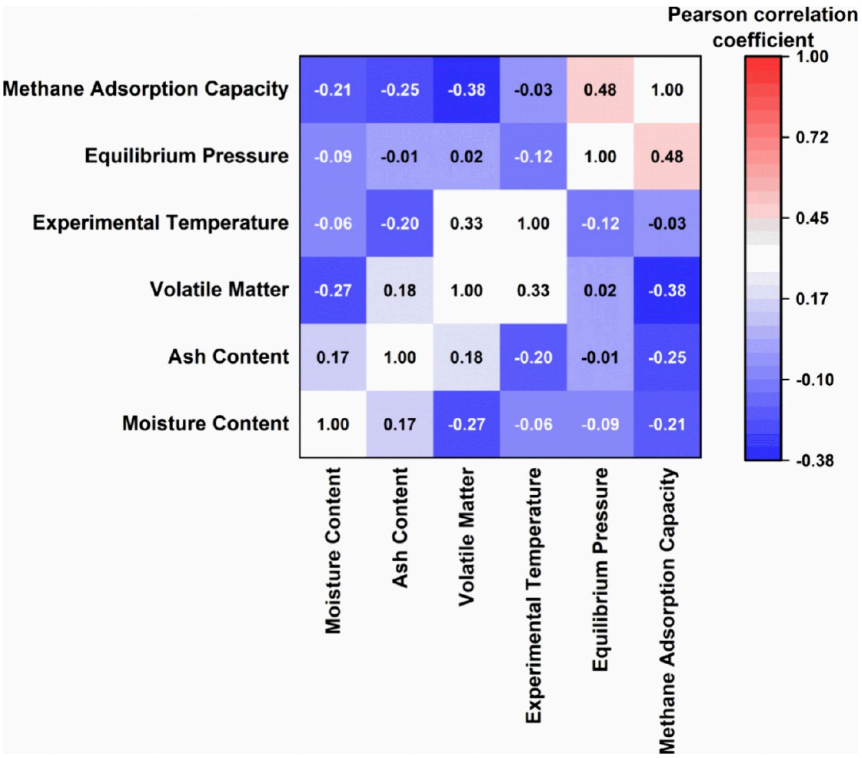



**Figure 3.** Pearson correlation coefficient heatmap of model parameters.



https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

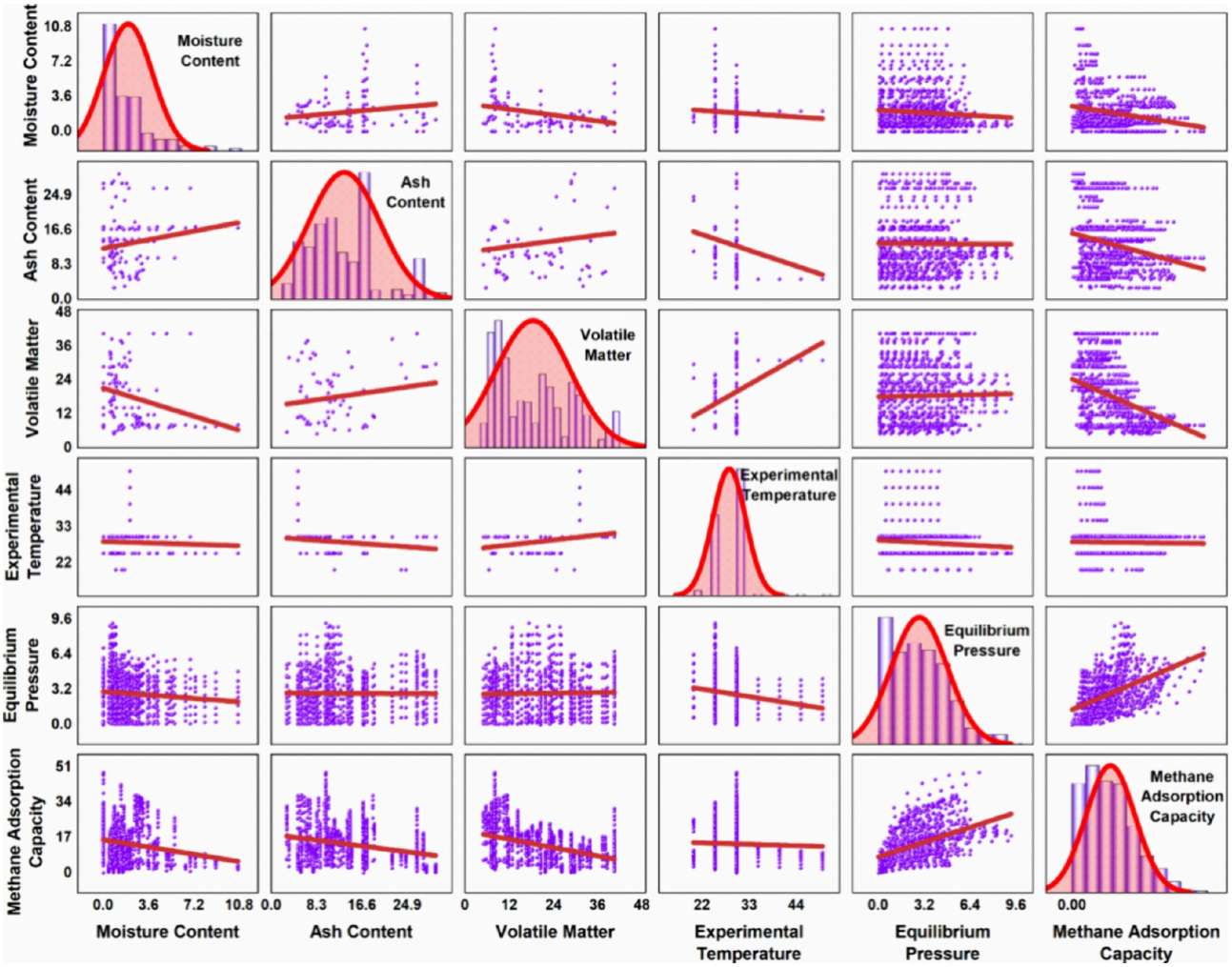



**Figure 4.** Input−output parameter distribution plots.

"U-shaped" trend.[24] In the low-rank metamorphic stage, volatile matter is relatively high, while adsorption capacity is low. As maturity increases, both volatile matter and adsorption capacity decrease, reaching their lowest point. However, in the high-rank metamorphic stage, although volatile matter is significantly reduced, the adsorption capacity gradually increases. From a mechanistic perspective, volatile matter indirectly influences the coal's ability to adsorb gas. Generally, the correlation for the five input parameters are low, which suggests that these parameters have minimal mutual influence, indicating a relatively independent impact on predicting gas adsorption values.

To eliminate the impact of data distribution on the model training results and thereby improve the generalization ability of the predictive model, it is essential that the input data for the model is uniformly distributed within the data range. To achieve this, it is necessary to establish distribution plots between the six input-output parameters, as shown in Figure 4. Here, it can be observed that the data points are uniformly distributed across the data ranges, and the mutual dependence between the input data points is low. The diagonal KDE (Kernel Density Estimation) plots indicate that the distribution of the six parameters approximates a normal distribution, exhibiting symmetrical characteristics. The distribution of methane adsorption $Q$ and the input parameters—moisture content $M_{ad}$, ash content $A_{ad}$, volatile matter $V_{daf}$, experimental temperature $T$, and equilibrium pressure $P$—indicates the following: $M_{ad}$ data points are within the range [0, 10], $A_{ad}$ within [0, 25], $V_{daf}$ within [5, 40], $T$ within [20, 50] at 5 °C intervals, and $P$ within [0, 10]. Additionally, the equilibrium pressure $P$ and gas adsorption $Q$ exhibit a nonlinear relationship similar to the Langmuir curve.

From the distribution plots, it can be observed that the gas adsorption $Q$ and the equilibrium pressure $P$ exhibit a Langmuir distribution trend. This conclusion is consistent with the experimental results discussed in the previous section. This finding lays the foundation for further prediction of gas adsorption and the calculation of the adsorption constants.

**4.2. Data Set Partitioning and Hyperparameter Optimization.** Before establishing and training the model, it is essential to partition the data set. Typically, data is split into 80% for training and 20% for testing. To validate the model before evaluation, the training set is further divided into 80% training data and 20% validation data.[47,48] This split aims to assess whether the model training meets the expected effectiveness. To fully utilize the limited data and explore the intrinsic relationships within the data, the study employs k-fold cross-validation to partition the data set. Given the large size of the training data set, we explore a range of commonly used $k$, typically from 2 to 10, to compare training outcomes. For each $k$, we compute the Mean Squared Error (MSE) and Coefficient of Determination ($R^2$). As shown in Figure 5, with a increase


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

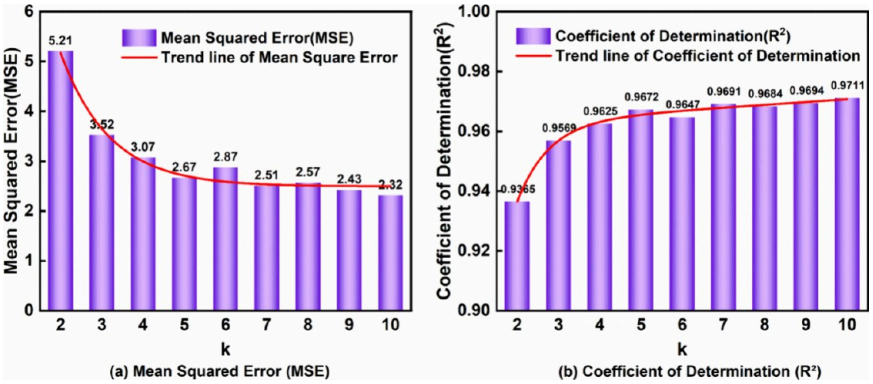



**Figure 5.** (a) Mean squared error (MSE) and (b) coefficient of determination ($R^2$) for models with different $k$ values.

in $k$, the MSE generated by the model after training gradually decreases, while the $R^2$ increases. When $k$ exceeds 7, the MSE value approaches 2.50, and the $R^2$ score approaches 0.97. Considering the balance between computational resource consumption and model accuracy, selecting $k$ = 10 is deemed more appropriate.

In machine learning, the selection of model parameters plays a crucial and direct role in determining the effectiveness of the trained model. Therefore, selecting model parameters wisely is key to achieving the expected training results. The hyperparameters of the RF algorithm are detailed in Table 2.

**Table 2. Values of Hyperparameters to Be Optimized and Their Corresponding Optimal Parameters**

| hyperparameter name | parameter range | explanation of parameters | optimal parameters |
|---|---|---|---|
| n_estimators | [50, 100, 200] | the number of decision trees | 100 |
| max_depth | [None, 10, 20, 30] | maximum depth | 20 |
| min_samples_split | [2, 5, 10] | minimum number of separated samples | 2 |
| min_samples_leaf | [1, 2, 4] | minimum number of leaf node samples | 2 |
| max_features | ['auto', 'sqrt', 'log2'] | maximum number of separated features | 'sqrt' |

Increasing the number of decision trees generally improves the model's accuracy, but it also results in longer training times and higher resource consumption. Although a larger number of trees enhances the model's ability to fit the training data, an excessively high number may lead to overfitting. The maximum depth of a tree reflects its complexity: greater depth enables a model to better capture the nuances of the training data but also increases the risk of overfitting. The minimum number of samples required to split an internal node determines how many samples are necessary for further splitting. Smaller values encourage deeper trees, improving the fit to the training data but also raising the risk of overfitting, while larger values limit tree depth, thereby enhancing the model's generalizability. The minimum number of samples required at a leaf node helps regulate tree complexity by preventing the formation of overly small leaf nodes, with higher values contributing to the model's stability. The maximum number of features considered for each split dictates the number of features evaluated at each node when seeking the optimal split in the RF algorithm. Lower values can reduce model variance and improve generalization ability.

We utilize the *GridSearchCV* method from the *sklearn* library for hyperparameter optimization and select optimal hyperparameter values within the permissible computational limits to achieve the best model performance. After hyperparameter optimization, the optimal results within the optimization range are obtained. Finally, the model is trained by calling the *RandomForestRegressor* method from the *sklearn* library with the optimal parameters.

**4.3. Model Training Results and Evaluation.** After data feature processing, data set partitioning, hyperparameter optimization, and model training, scatter plots of actual and predicted gas adsorption values for the training and test sets are generated, as shown in Figure 6. It visually demonstrates

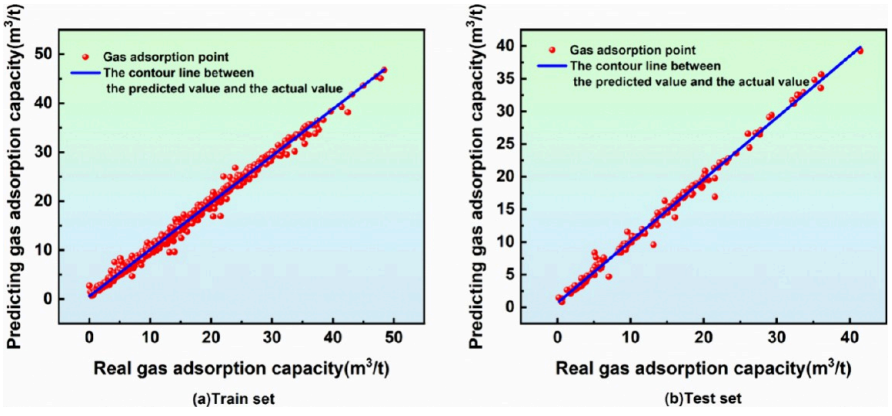



**Figure 6.** (a, b) Scatter plot of data from model training and testing sets.



https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

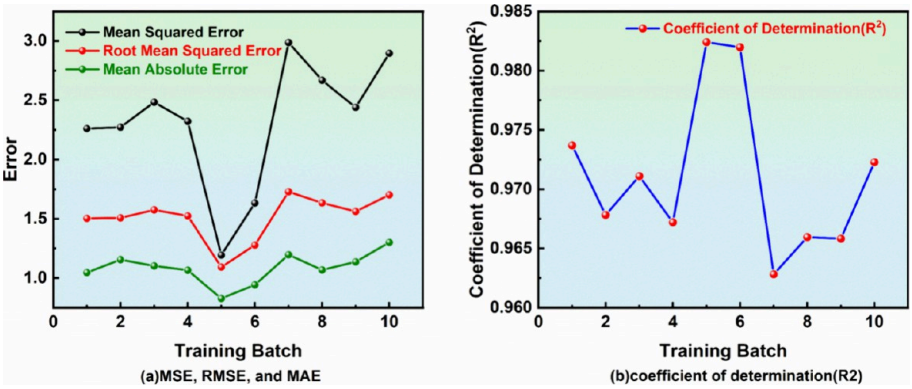



**Figure 7.** (a) Error value and (b) determination coefficient of the model.

**Table 3. Experimental and Model Prediction Data for Coal Samples A, B, and C**

| coal sample | temperature (°C) | moisture content (%) | ash content (%) | volatile matter (%) | equilibrium pressure (MPa) | $Q_t$ (m³/t) | $Q_p$ (m³/t) |
|---|---|---|---|---|---|---|---|
| A[38] | 30 | 0.57 | 7.11 | 31.73 | 0.235 | 3.760 | 3.725 |
| | | | | | 0.637 | 6.819 | 6.623 |
| | | | | | 1.156 | 8.925 | 8.835 |
| | | | | | 1.661 | 10.229 | 9.845 |
| | | | | | 2.166 | 11.176 | 11.241 |
| | | | | | 2.555 | 11.796 | 11.765 |
| | | | | | 3.404 | 12.650 | 12.623 |
| | | | | | 4.110 | 13.268 | 13.429 |
| | | | | | 5.004 | 13.943 | 13.567 |
| B[41] | 25 | 0.82 | 10.3 | 24.01 | 0.328 | 4.105 | 4.705 |
| | | | | | 0.666 | 6.823 | 7.191 |
| | | | | | 1.424 | 10.200 | 10.502 |
| | | | | | 1.574 | 11.070 | 10.821 |
| | | | | | 2.342 | 12.843 | 12.984 |
| | | | | | 3.258 | 14.335 | 14.594 |
| | | | | | 4.013 | 15.340 | 15.448 |
| | | | | | 4.903 | 16.134 | 16.308 |
| | | | | | 5.694 | 16.651 | 16.921 |
| | | | | | 6.497 | 17.202 | 17.275 |
| | | | | | 7.313 | 17.544 | 17.674 |
| | | | | | 8.140 | 17.781 | 17.942 |
| C[40] | 30 | 3.17 | 7.08 | 8.41 | 0.095 | 4.504 | 5.122 |
| | | | | | 1.233 | 17.912 | 18.712 |
| | | | | | 2.047 | 22.843 | 23.232 |
| | | | | | 2.923 | 26.209 | 26.863 |
| | | | | | 3.798 | 28.792 | 28.981 |
| | | | | | 4.691 | 31.019 | 31.420 |
| | | | | | 5.516 | 33.672 | 32.885 |

the relationship between the actual and predicted values, showing a high degree of linear correlation between them, which indicates that the model's predictions align well with the actual gas adsorption values and validate the effectiveness and accuracy of the model.

To further quantify the relationship between the actual and predicted values, four evaluation metrics are introduced: MSE, RMSE, MAE, and $R^2$. By analyzing these metrics at each step of the model training process, including their averages, one can accurately determine the optimal completion point of the trained model. As shown in Figure 7, the MSE for each training batch is concentrated in the range of 1.20 to 3.00, with a mean of approximately 2.32. By calculating MSE, the error value can be amplified and the error impact caused by outliers can be highlighted. The RMSE ranges from 1.1 to 1.7, with a mean of about 1.51, serving to standardize errors in relation to the original data units and visually representing error magnitude. $R^2$ ranges from 0.96 to 0.99, with a mean of 0.975. When the training batch $k$ is 5, the model achieves its lowest MSE, RMSE, and MAE, specifically 1.19, 1.09, and 0.82, respectively. At this point, with an $R^2$ score exceeding 0.982, the model performs optimally in this training batch, indicating that the training results align closely with expectations.

**4.4. Application of the Model.** To verify model effectiveness and generalization ability, the experimental data from three groups were summarized as follows. The three coal samples exhibit substantial variation in input characteristics. Samples A and C were tested at 30 °C, while sample B was tested at 25 °C, allowing the model's adaptability to different temperature conditions to be assessed. Likewise, moisture


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

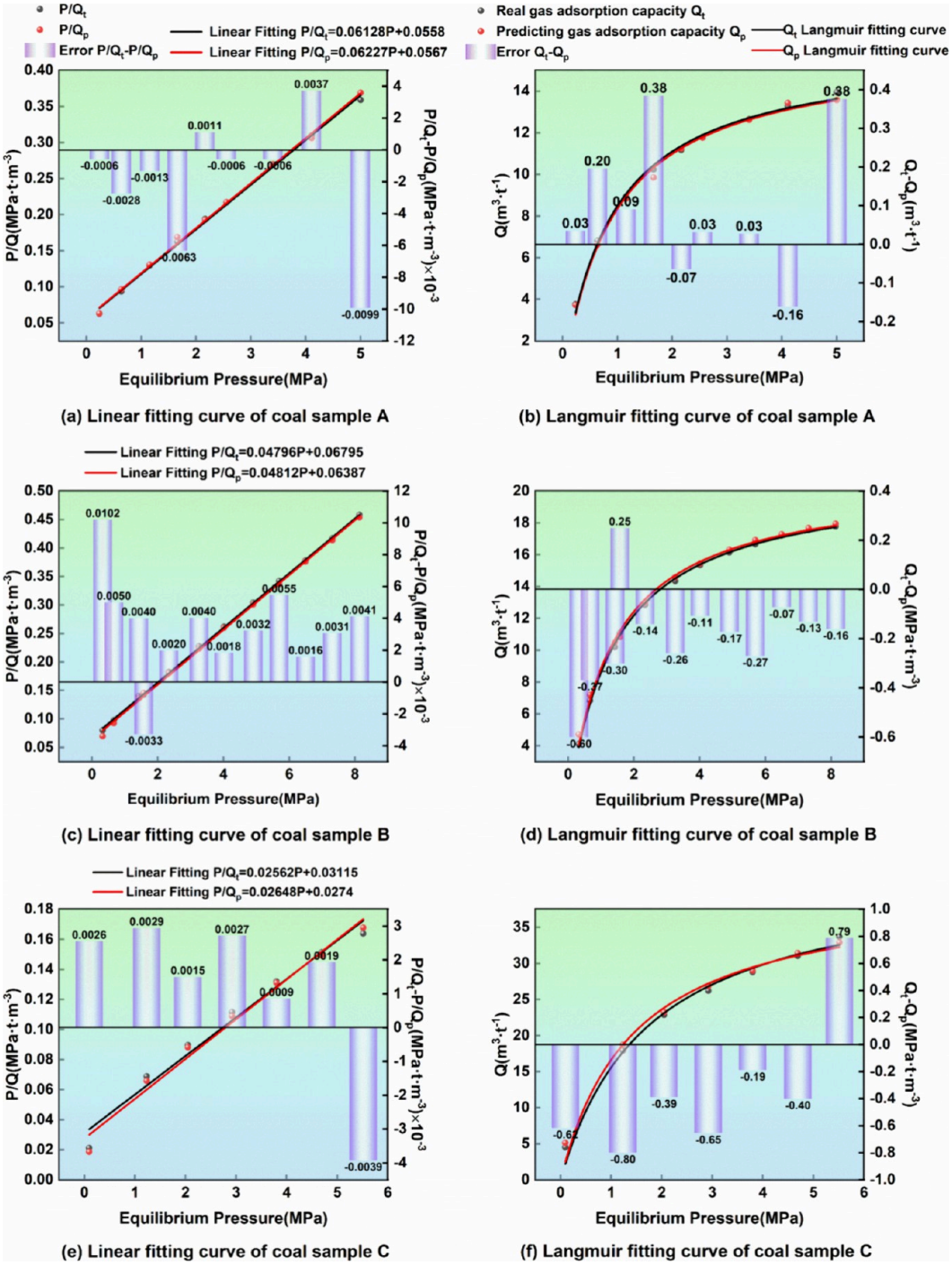



**Figure 8.** (a–f) Linear fitting and Langmuir fitting curves of coal samples and data errors.

**Table 4. Experimental and Predicted Values of Langmuir Volume $V_L$ and Adsorption Constants $K_L$ and Their Errors**

| | maximum adsorption capacity $V_L$ | | | adsorption constant $K_L(1/P_L)$ | | |
|---|---|---|---|---|---|---|
| coal sample | experiment $V_{Lt}$ | prediction $V_{Lp}$ | error $V_{Lt}-V_{Lp}$ | experiment $K_{Lt}$ | prediction $K_{Lp}$ | error $K_{Lt}-K_{Lp}$ |
| A | 16.045 | 16.058 | −0.013 | 1.122 | 1.082 | 0.040 |
| B | 20.836 | 20.779 | 0.057 | 0.703 | 0.743 | −0.039 |
| C | 42.934 | 40.694 | 2.239 | 0.564 | 0.690 | −0.127 |

content varies from 0.57% to 3.17%, ash content from 7.08% to 10.3%, and volatile matter from 8.41% to 31.73%, enabling evaluation of the model's sensitivity to changes in these parameters. Overall, the selection of these three samples provides data diversity, allowing for model validation across different experimental conditions and supporting an assessment of the model's accuracy and robustness in predicting methane adsorption capacity in coal. Using the linear equation


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

of the Langmuir given in Chapter 2.1, the experimental data from three sets of coal samples were utilized to plot linear regression curves and Langmuir curves depicting the real and predicted values of gas adsorption against their corresponding equilibrium pressure. In Table 3, it is important to specify that Moisture Content refers to equilibrium moisture content. Additionally, $Q_t$ represents the measured values of methane adsorption capacity in coal, while $Q_p$ represents the predicted values.

The data from the three sets of coal samples in Table 3 were fitted using eqs 1 and 2, and scatter plots along with fitting curves are shown in Figure 8. The parameters of the fitted Langmuir curves are shown in Table 4. The relationships between the true methane adsorption capacity $Q_t$, the predicted methane adsorption capacity $Q_p$, and the equilibrium pressure $P$ fitted by eq 1 are shown in subfigures (b), (d), and (f) within Figure 8. The error between the measured and predicted methane adsorption amounts is around 0.50. The fitting degree $R^2$ of the Langmuir curve fitted to the RF model-predicted gas adsorption amount $Q$ reaches 0.99. Compared to the errors introduced by traditional experimental methods, the RF prediction error is smaller, and the predicted values meet the basic experimental data requirements.

The Random Forest algorithm accurately predicts the methane adsorption capacity in coal, and estimates the limit adsorption capacity ($V_L$) and adsorption constant ($K_L$) of coal samples, providing a basis for coal mine safety and gas extraction. In future research, efforts will be directed toward optimizing model parameters, incorporating additional variables to enhance accuracy, and conducting a comparative analysis of alternative machine learning algorithms.

**4.5. Sensitivity Analysis.** To investigate the impact of each input feature on the methane adsorption capacity in coal, a comprehensive sensitivity analysis was conducted using the SHAP (SHapley Additive exPlanations) method and information gain calculation. SHAP values, derived from Shapley values in game theory, quantify the marginal contribution of each feature to the model's prediction, providing an intuitive visualization of feature importance. Information gain, on the other hand, calculates the importance score of each feature by accumulating its contribution at each split node in the RF model, offering a complementary assessment of global feature importance.

As shown in Figure 9, the SHAP analysis results indicate that the SHAP value distributions for equilibrium pressure and volatile matter are particularly significant, suggesting that these two features have high importance in the model and exhibit clear trends in influencing the prediction results. Specifically, as equilibrium pressure increases, SHAP values display a positive upward trend, consistent with the phenomenon of enhanced methane adsorption capacity at higher pressures. Additionally, the impact of volatile matter shows a "U-shaped" trend: in the low volatile matter region, SHAP values are lower, indicating that adsorption capacity decreases with increasing volatile matter; however, in the high volatile matter region, SHAP values gradually rise, suggesting a recovery in adsorption capacity. This trend aligns with the adsorption performance changes associated with coal rank, where low-rank coals generally exhibit lower volatile matter and adsorption capacity, while high-rank coals show increased adsorption as volatile matter decreases.

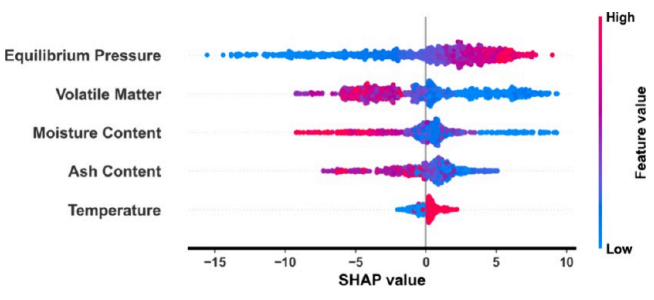



**Figure 9.** SHAP analysis results of each input feature.

The information gain calculation further validates these findings. Figure 10 presents the feature importance rankings,

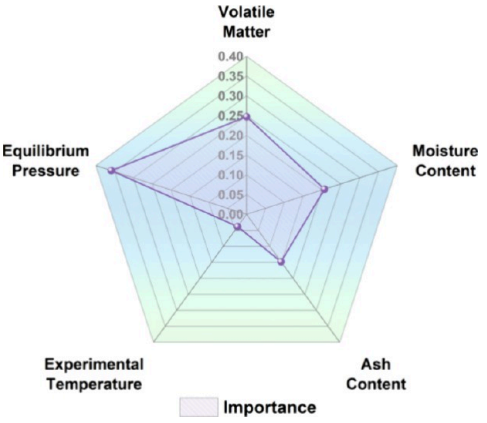



**Figure 10.** Importance of model input parameters in predicting gas adsorption capacity.

with cumulative information gain values for equilibrium pressure and volatile matter being relatively high, indicating their strong sensitivity in predicting methane adsorption. Equilibrium pressure and volatile matter not only reflect the physical adsorption characteristics of methane under different environmental conditions but also reveal the complex interactions between coal structure and methane adsorption. In contrast, the SHAP values for moisture and ash content are relatively flat, and their information gain scores are low, suggesting a minor influence on prediction results in the current data set, likely due to the narrow range of these features.

Furthermore, the feature importance ranking indicates that the influence of moisture, ash content, and temperature on the results is relatively smaller compared to equilibrium pressure and volatile matter. Analysis reveals that the limited impact of temperature on methane adsorption capacity is likely due to the temperature distribution in the data set, which is primarily concentrated between 25 and 35 °C and exhibits considerable dispersion. This observation is further supported by the SHAP values of temperature shown in Figure 9. Additionally, the distribution of moisture and ash content is more uniform compared to temperature, displaying a negative correlation within the examined range but having minimal impact on methane adsorption capacity. This can be attributed to the limited influence of moisture and ash content on the primary adsorption sites in coal, as they do not directly participate in the methane adsorption process. Specifically, moisture may occupy coal pores and surfaces, thereby reducing the available adsorption sites for methane. And, ash content, predominantly composed of mineral matter, has a relatively small specific


https://doi.org/10.1021/acs.energyfuels.4c04906
Energy Fuels 2024, 38, 23422−23432

surface area and weak adsorption capacity, leading to its negligible effect on methane adsorption.

By combining the trend analysis of SHAP values with the feature importance rankings from information gain, a potential functional relationship between methane adsorption capacity $Q$, volatile matter $V_{daf}$, and equilibrium pressure $P$ can be inferred. This study not only identifies the key features in the model but also reveals the impact patterns of feature value changes on methane adsorption prediction. This integrated sensitivity analysis enhances the interpretability of the model, providing a solid basis for optimizing model input features and improving prediction accuracy.

## 5. CONCLUSIONS

By integrating the Random Forest algorithm with traditional methane isotherm adsorption experiments, optimizing hyperparameters, and employing k-fold cross-validation, a high-accuracy prediction model for methane adsorption in coal was developed. Based on the model, the limit adsorption capacity $V_L$ and adsorption constant $K_L$ of the coal samples were estimated. The main conclusions are as follows:

1. By optimizing five parameters—number of decision trees, maximum depth, minimum samples for splitting, minimum samples per leaf, and maximum features for splitting—a model with a coefficient of determination $R^2$ of 0.98 was trained. The low prediction error indicates that the model has a strong ability to fit the data.
2. By applying k-fold cross-validation to partition the data set and comparing the MSE and $R^2$ of models with various $k$, $k$ = 10 was chosen as the optimal iteration number.
3. By estimating the limit adsorption capacity $V_L$ and adsorption constant $K_L$ of the coal samples from the methane adsorption capacity predicted by the model at various equilibrium pressures, the time-consuming experimental process can be avoided. The results indicate that the computed $V_L$ and $K_L$ are close to the experimental values, with errors in the decimal or even percentage range, validating the effectiveness of the model and method. Additionally, sensitivity analysis of the model's parameters confirmed that there is a functional relationship between the volatile matter $V_{daf}$, equilibrium pressure $P$, and methane adsorption capacity $Q$, further verifying the influence of volatile matter and equilibrium pressure on methane adsorption capacity in coal.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Data available on request from the authors.

## ■ AUTHOR INFORMATION


### Corresponding Author

**Wei Li** − *State Key Laboratory of Coal Mine Disaster Prevention and Control and National Engineering Research Center for Coal Gas Control, China University of Mining and Technology, Xuzhou 221116, China*; Email: lweisafety@cumt.edu.cn

### Authors

**Wenshuo Li** − *State Key Laboratory of Coal Mine Disaster Prevention and Control and National Engineering Research Center for Coal Gas Control, China University of Mining and Technology, Xuzhou 221116, China*; orcid.org/0009-0005-6526-8395

**Andreas Busch** − *Lyell Centre, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*; orcid.org/0000-0002-3279-5202

**Liang Wang** − *State Key Laboratory of Coal Mine Disaster Prevention and Control and National Engineering Research Center for Coal Gas Control, China University of Mining and Technology, Xuzhou 221116, China*

**Ferian Anggara** − *Department of Geology, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*

**Shilong Yang** − *State Key Laboratory of Coal Mine Disaster Prevention and Control and National Engineering Research Center for Coal Gas Control, China University of Mining and Technology, Xuzhou 221116, China*; orcid.org/0000-0001-9316-5813

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.energyfuels.4c04906


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported by the Fundamental Research Funds for the National Natural Science Foundation of China (grant nos. 52374243 and 51874295) and International Cooperation and Exchange Program (grant no. 51911530199, NSFC-RS)

## ■ REFERENCES


(1) Wu, G.; Ye, Z.; Zhang, L.; Tang, J. Bituminous Coal Sorption Characteristics and Its Modeling of the Main Coal Seam Gas Component in the Huaibei Coalfield, China. *Sustainability* **2023**, *15*, 9822.

(2) Guo, P.; Tang, X.; Wen, L.; Wu, B.; Luo, F.; Liu, Y. Geological characteristics and coalbed methane adsorbability of shallow coal rock in Qinshui Basin, China. *Journal of Petroleum Exploration and Production Technology* **2024**, *14*, 2901.

(3) Li, Z.; Ren, T.; Black, D.; Qiao, M.; Abedin, I.; Juric, J.; Wang, M. In-situ gas contents of a multi-section coal seam in Sydney basin for coal and gas outburst management. *International Journal of Coal Science & Technology* **2023**, *10* (1), 62.

(4) Li, W.; Guo, J.; Jiao, Y.; Deng, D.; Zhong, Y.; Yang, S. Investigation on tectonically deformed coal particle crushing by DEM under triaxial loading: Implication for coal and gas outburst. *Advanced Powder Technology* **2023**, *34* (10), No. 104155.

(5) Li, W.; Ren, T.; Busch, A.; den Hartog, S. A. M.; Cheng, Y.; Qiao, W.; Li, B. Architecture, stress state and permeability of a fault zone in Jiulishan coal mine, China: Implication for coal and gas outbursts. *International Journal of Coal Geology* **2018**, *198*, 1−13.

(6) Zeng, Q.; Wang, Z.; Sui, T.; Huang, T. Adsorption Mechanisms of High-Pressure Methane and Carbon Dioxide on Coals. *Energy Fuels* **2021**, *35* (16), 13011−13021.

(7) Zhang, J.; Feng, Q.; Zhang, X.; Hu, Q.; Yang, J.; Wang, N. A Novel Data-Driven Method to Estimate Methane Adsorption Isotherm on Coals Using the Gradient Boosting Decision Tree: A Case Study in the Qinshui Basin China. *Energies* **2020**, *13*, 5369.

(8) Lei, H. Experimental study on determination of residual space volume of adsorption tank by vacuum filled He method. *Coal Sci. Technol.* **2024**, *52* (S1), 86−93.

(9) Qiu, F. Automatic Device for the Determination of Coal Gas Adsorption Constant. *Autom. Instrum.* **2021**, No. 8, 207−210.

(10) Xiouras, C.; Cameli, F.; Quilló, G. L.; Kavousanakis, M. E.; Vlachos, D. G.; Stefanidis, G. D. Applications of Artificial Intelligence and Machine Learning Algorithms to Crystallization. *Chem. Rev.* **2022**, *122* (15), 13006−13042.
(11) Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science* **2021**, *2* (3), 160.
(12) Mohammed, M.; Khan, M. B.; Bashier, E. B. M. *Machine Learning: Algorithms and Applications*; 2016. CRC Press DOI: .
(13) Meng, M.; Zhong, R.; Wei, Z. Prediction of methane adsorption in shale: Classical models and machine learning based models. *Fuel* **2020**, *278*, No. 118358.
(14) Raparthi, M.; Soni, M.; Tiwari, V.; Dhumane, A.; Sharma, R. Scalable Implementation of Random Forests for Big Data Classification on Cloud Infrastructure. In *Deep Learning and Visual Artificial Intelligence*, 2024; Goar, V.; Sharma, A.; Shin, J.; Mridha, M. F., Eds.; Springer Nature: Singapore, pp 493−512.
(15) Yates, D.; Islam, M. Z. FastForest: Increasing random forest processing speed while maintaining accuracy. *Information Sciences* **2021**, *557*, 130−152.
(16) Xue, S.; Zheng, X.; Yuan, L.; Lai, W.; Zhang, Y. A review on coal and gas outburst prediction based on machine learning. *J. China Coal Soc.* **2024**, *49* (2), 664−694.
(17) Wang, Z.; Cai, Y.; Liu, D.; Qiu, F.; Sun, F.; Zhou, Y. Intelligent classification of coal structure using multinomial logistic regression, random forest and fully connected neural network with multisource geophysical logging data. *International Journal of Coal Geology* **2023**, *268*, No. 104208.
(18) Xiaoliang, Z.; Wenhao, L.; Lei, Z.; Sheng, X. Quantitative evaluation of the indexes contribution to coal and gas outburst prediction based on machine learning. *Fuel* **2023**, *338*, No. 127389.
(19) Tavakolian, M.; Najafi-Silab, R.; Chen, N.; Kantzas, A. Modeling of methane and carbon dioxide sorption capacity in tight reservoirs using Machine learning techniques. *Fuel* **2024**, *360*, No. 130578.
(20) Hao, D.; Zhang, L.; Li, M.; Tu, S.; Zhang, C.; Bai, Q.; Wang, C. Experimental study of the moisture content influence on CH4 adsorption and deformation characteristics of cylindrical bituminous coal core. *Adsorption Science & Technology* **2018**, *36* (7−8), 1512−1537.
(21) Chen, M.-y.; Cheng, Y.-p.; Li, H.-r.; Wang, L.; Jin, K.; Dong, J. Impact of inherent moisture on the methane adsorption characteristics of coals with various degrees of metamorphism. *Journal of Natural Gas Science and Engineering* **2018**, *55*, 312−320.
(22) Crosdale, P. J.; Moore, T. A.; Mares, T. E. Influence of moisture content and temperature on methane adsorption isotherm analysis for coals from a low-rank, biogenically-sourced gas reservoir. *International Journal of Coal Geology* **2008**, *76* (1), 166−174.
(23) Cheng, Y.; Jiang, H.; Zhang, X.; Cui, J.; Song, C.; Li, X. Effects of coal rank on physicochemical properties of coal and on methane adsorption. *International Journal of Coal Science & Technology* **2017**, *4* (2), 129−146.
(24) Busch, A.; Han, F.; Magill, C. R. Paleofloral dependence of coal methane sorption capacity. *International Journal of Coal Geology* **2019**, *211*, No. 103232.
(25) Guo, F.; Liu, G.; Zhang, Z.; Lv, R.; Xian, B.; Lin, J.; Barakos, G.; Chang, P. A Fractal Adsorption Model on Methane in Coal with Temperature Effect Dependence. *Fractal and Fractional* **2024**, *8* (7), 370.
(26) Wang, Z.; Li, Y.; Wang, Z.; Zhou, L. Factors Influencing the Methane Adsorption Capacity of Coal and Adsorption Heat Variations. *Energy Fuels* **2023**, *37* (17), 13080−13092.
(27) Wang, Z.; Wang, X.; Zuo, W.; Ma, X.; Li, N. The influence of temperature on methane adsorption in coal: A review and statistical analysis. *Adsorption Science & Technology* **2019**, *37* (9−10), 745−763.
(28) Guan, C.; Liu, S.; Li, C.; Wang, Y.; Zhao, Y. The temperature effect on the methane and CO2 adsorption capacities of Illinois coal. *Fuel* **2018**, *211*, 241−250.
(29) Cheng, Y.; Liu, Q.; Ren, T. *Coal Mechanics*; Springer: Singapore, 2021. DOI: .
(30) Swenson, H.; Stadie, N. P. Langmuir's Theory of Adsorption: A Centennial Review. *Langmuir* **2019**, *35* (16), 5409−5426.
(31) Kondo, S.; Ishikawa, T.; Abe, I. *Adsorption Science*; Chemical Industry Press: Beijing 2006.
(32) Langmuir, I. THE ADSORPTION OF GASES ON PLANE SURFACES OF GLASS, MICA AND PLATINUM. *J. Am. Chem. Soc.* **1918**, *40*, 1361−1403.
(33) Liu, Z.; Wang, Z. Influence of metamorphic degree on gas diffusion law in coal. *J. Heilongjiang Univ. Sci. Technol.* **2024**, *34* (2), 201−206.
(34) Liu, Y.; Hao, C.; Wang, Z.; Xie, J.; Zhao, W.; Meng, F.; Han, Y. Micropore distribution and methane adsorption process and mechanism in bituminous coals: A molecular dynamics simulation study. *Journal of Environmental Chemical Engineering* **2024**, *12* (2), No. 112139.
(35) Xie, Z.; Zhang, B.; Gan, R.; Fu, Z. Determination and Regularity of Gas Adsorption Constants of Coal Samples from the 5# Coal Seam in Datong Coalfield. *Energy Energy Conserv.* **2023**, *05*, 65−67+113.
(36) Li, S. Experimental Study on Effect of Temperature Change on Gas Adsorption Characteristics of Coal. *Shaanxi Coal* **2023**, *42* (4), 39−43+54.
(37) Zhang, Q.; Liu, X.; Nie, B.; Wu, W.; Wang, R. Methane sorption behavior on tectonic coal under the influence of moisture. *Fuel* **2022**, *327*, No. 125150.
(38) Hu, B. *Methane Adsorption Behavior Characteristics of Multi-scale Pore Structure in Coal and Its Microscopic Influencing Mechanism*. China University of Mining and Technology, 2022, https://link.cnki.net/doi/10.27623/d.cnki.gzkyu.2022.000167.
(39) Yang, H.; Bi, W.; Zhang, Y.; Yu, J.; Yan, J.; Lei, D.; Ma, Z. Effect of tectonic coal structure on methane adsorption. *Journal of Environmental Chemical Engineering* **2021**, *9* (6), No. 106294.
(40) Wang, X.; Cheng, Y.; Zhang, D.; Yang, H.; Zhou, X.; Jiang, Z. Experimental study on methane adsorption and time-dependent dynamic diffusion coefficient of intact and tectonic coals: Implications for CO2-enhanced coalbed methane projects. *Process Safety and Environmental Protection* **2021**, *156*, 568−580.
(41) Guo, X. *Deformation and Metamorphism Characteristics of Deformed Coal and Its Effect on Methane Adsorption*. doctor, China University of Mining and Technology (Beijing), 2020. https://link.cnki.net/doi/10.27624/d.cnki.gzkbu.2020.000045.
(42) Qin, X. Gas Adsorption Property of Bituminous Coal under Water Influence. *Coal Min. Technol.* **2018**, *23* (1), 104−107.
(43) Voncken, J. H. L. The Origin and Classification of Coal. In *Geology of Coal Deposits of South Limburg, The Netherlands: Including Adjacent German and Belgian Areas*, Voncken, J. H. L., Ed.; Springer International Publishing, 2020; pp 25−40.
(44) Biau, G. Analysis of a random forests model. *J. Mach. Learn. Res.* **2012**, *13* (1), 1063−1095.
(45) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Resampling Methods. In *An Introduction to Statistical Learning: with Applications in R*, James, G.; Witten, D.; Hastie, T.; Tibshirani, R., Eds.; Springer: US, 2021; pp 197−223.
(46) Rainio, O.; Teuho, J.; Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **2024**, *14* (1), 6086.
(47) Hastie, T.; Tibshirani, R.; Friedman, J. Random Forests. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Hastie, T.; Tibshirani, R.; Friedman, J., Eds.; Springer: New York, 2009; pp 587−604.
(48) Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* - Vol. 2, Morgan Kaufman Publishing: Montreal, Quebec, Canada, 1995.