processes

MDPI

Article

Prediction Model for Harmful Gas Risk Levels in Non-Coal Strata Tunnels Based on SSA-CatBoost

Zengchan Mao [1], Wenpin Luo [2], Jianhua Wu [1,*], Peidong Su [3], Xiaojin Wang [1] and Peng Yang [2]

[1] Sichuan Water Development Investigation, Design & Research Co., Ltd., Chengdu 610072, China
[2] Sichuan Tingzikou Irrigation Area Construction and Development Co., Ltd., Nanchong 637000, China
[3] School of Geoscience and Technology, Southwest Petroleum University, Chengdu 610500, China
* Correspondence: jianhuawu123@163.com

Abstract

Harmful gas is a major hazard in underground engineering construction, and accurate prediction of its risk level is essential for tunnel safety. Existing prediction methods for harmful-gas risk in non-coal strata tunnels are limited by empirical scoring, subjective indicator assignment, and insufficient quantitative characterization of reservoir performance. To address these limitations, this study proposes an SSA-CatBoost prediction model for harmful-gas risk levels in non-coal strata tunnels. Eight influencing indicators were selected as input variables. Among them, reservoir performance was quantitatively characterized by measured porosity and permeability, while the other six indicators were quantified using engineering-based scoring criteria. A database containing 138 real harmful-gas tunnel cases was constructed, and CatBoost was used as the base classifier. The Sparrow Search Algorithm was introduced to optimize the hyperparameters of CatBoost. The proposed SSA-CatBoost model achieved an average accuracy of 93.63% in five-fold cross-validation and an accuracy of 92.86% on the independent test set. Compared with CatBoost, SSA-SVM, and SSA-XGBoost, the proposed model showed the highest cross-validation accuracy. Engineering validation further showed that all selected validation samples were correctly classified. In addition, replacing empirical reservoir-performance scoring with measured porosity and permeability improved the recognition performance of adjacent risk levels, with the F1-scores of Level III and Level IV increasing from 0.667 and 0.727 to 0.909, respectively. The novelty of this study lies in integrating measured reservoir-performance parameters into a machine-learning-based harmful-gas risk prediction framework, thereby reducing the subjectivity of conventional scoring systems and improving the quantitative characterization of non-coal strata tunnel gas hazards.

Keywords: harmful gas; tunnel; machine learning; prediction; classification

Academic Editor: Dicho Stratiev

Received: 3 May 2026
Revised: 15 June 2026
Accepted: 22 June 2026
Published: 6 July 2026

1. Introduction

With the increasing global demand for infrastructure, the rise in underground engineering construction has led to frequent underground geological disasters, especially gas outbursts and combustion accidents [1–4]. Due to the toxicity and flammability of gases, they pose extreme dangers in underground environments, often causing serious injuries and fatalities. Faced with numerous new projects, achieving accurate and rapid prediction of harmful-gas risk levels is crucial for ensuring construction safety [5–7]. Machine learning, as an emerging technology, can effectively achieve this goal, particularly in multi-factor and highly coupled prediction tasks, demonstrating powerful predictive capabilities to

handle the complex nonlinear relationships involved in gas outburst processes [8]. Unlike coal-seam-gas hazards, harmful-gas risks in non-coal strata tunnels are usually related to complex geological conditions, including adjacent oil and gas reservoirs, gas migration pathways, caprock integrity, groundwater conditions, tunnel burial depth, and the storage and transport capacity of the surrounding strata. In such tunnels, harmful gases may accumulate or be released during excavation, leading to toxic, flammable, or explosive hazards. Therefore, the key problem addressed in this study is the prediction of harmful-gas risk levels in non-coal strata tunnels under multi-factor coupling conditions.

Currently, extensive research has been conducted on coal-seam-gas prediction and gas-related hazard assessment in underground engineering, including tunnel construction and deep coal mining. Zhang and Lowndes [9] proposed a prediction method combining fault tree analysis and artificial neural networks to evaluate the potential risk of coal and gas outbursts. Wang et al. [10] established a coal and gas outburst prediction model based on extension theory, constructing an evaluation system with multiple indicators and risk levels to achieve quantitative outburst-risk prediction. Bi et al. [1] developed a coal mine gas emission prediction model based on a hybrid machine-learning framework, which improved the prediction accuracy of gas emission. Lin et al. [11] proposed a coal mine gas emission prediction method based on multi-factor time series analysis, combining Recursive Feature Elimination with Cross-Validation (RFECV) feature selection with a Bidirectional Long Short-Term Memory (Bi-LSTM) model to improve the accuracy and stability of gas emission prediction. Zheng et al. [12] used a parameter-optimized Extreme Gradient Boosting (XGBoost) model to predict coal and gas outbursts and quantitatively analyzed the contribution of each evaluation indicator. In addition to these methods, artificial neural network (ANN) models have also been applied to nonlinear prediction and sensitivity-analysis problems in geotechnical engineering. For example, Jolfaei and Lakirouhani [13] used an ANN model to analyze the relationship between input parameters and borehole failure geometry and to evaluate the sensitivity of effective parameters in borehole failure.

However, due to the strong uncertainty and spatial heterogeneity of harmful-gas distribution in non-coal strata tunnels, the direct application of coal-seam-gas prediction models to non-coal tunnel engineering remains limited. Liu et al. [14] proposed a risk prediction model for oil-type gas tunnels based on factors such as the burial depth of adjacent oil and gas reservoirs and their distance from the tunnel. Zhu et al. [15] further introduced factors such as groundwater conditions and the burial depth of adjacent oil and gas production areas to enrich the indicator inputs, thereby enhancing the model's explanatory power and evaluation effectiveness. However, most input indicators in the above two studies were quantified using scoring methods. In particular, reservoir performance was mainly characterized by empirical grading assignment. Although this method is reasonable for indicators that are difficult to measure directly, such as the distance to adjacent oil and gas areas, empirical scoring may not fully reflect the physical characteristics of reservoir performance, which can be directly measured through porosity and permeability tests.

Based on this research gap, this study aims to improve the objectivity and predictive reliability of harmful-gas risk level assessment in non-coal strata tunnels by reducing the dependence on empirical scoring. While retaining scoring indicators that are mainly obtained from geological surveys, this study replaces the empirical scoring of reservoir performance with measured porosity and permeability parameters, which can better characterize the gas storage and migration capacity of the strata. A database containing 138 real harmful-gas tunnel cases was constructed, and an SSA-CatBoost prediction model was established for harmful-gas risk levels. Compared with ANN-based methods, the proposed SSA-CatBoost model is a tree-based ensemble learning framework that is suitable for structured tabular

engineering data and provides more direct feature-importance information. In addition, SSA was introduced to optimize the hyperparameters of CatBoost, thereby improving the stability and predictive performance of the model. The model performance was evaluated using five-fold cross-validation, an independent test set, engineering validation cases, and comparisons with other machine-learning models. The average accuracy of five-fold cross-validation reached 93.63%, and the accuracy on the independent test set reached 92.86%. In the engineering validation, all selected validation samples were correctly classified. The novelty of this study lies in integrating measured reservoir-performance parameters into a machine-learning-based prediction framework, thereby reducing the subjectivity of conventional scoring systems and improving the quantitative characterization of harmful-gas risk in non-coal strata tunnels.

2. Materials and Methods

2.1. Sampling and Physical Property Experiments

A core drill was used as the operating platform on the construction site, implementing rotary drilling during geological exploration. Rock core samples for testing were all taken from relatively complete rock formations. A total of 138 rock gas samples were collected from tunnel boreholes, and all samples were tested for porosity and permeability.

Porosity was determined using the gas expansion method, with nitrogen or helium as the measurement medium, based on Boyle's law. During the test, a known volume of standard gas at a set initial pressure undergoes isothermal expansion into the core holder, which is initially at atmospheric pressure, allowing the gas to diffuse into the core pores. Based on the pressure change before and after expansion and the known volumes, combined with the gas state equation, the effective pore volume and grain volume of the tested rock sample can be calculated, and then the sample porosity is obtained [10].

Permeability was tested using a core permeability tester, based on Darcy's law. Under a certain pressure condition, air flows through the core after passing through a flow meter. Due to differences in permeability among different rock samples, different pressure drops are generated when the airflow passes through the two ends of the core. Pressure sensors transmit this information to a computer, which calculates the pressure difference across the core ends and the corresponding flow rate, substitutes them into relevant formulas, and finally obtains the sample permeability [10].

2.2. Sparrow Search Algorithm (SSA)

The Sparrow Search Algorithm (SSA), proposed in 2020, is a heuristic algorithm that simulates the behavior of sparrow flocks during foraging and predator evasion [16]. During foraging, sparrows exhibit three role behaviors: producers, responsible for finding food, delineating foraging areas, and guiding the direction; scroungers, which guide the flock to forage together after the producer finds food; and guards, responsible for identifying danger and transmitting warning signals.

Assuming the matrix representation of the sparrow population is shown in Equation (1):

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} \dots & x_{1,d} \\ x_{2,1} & x_{2,2} \dots & x_{2,d} \\ \cdot\cdot & \cdot & \cdot \\ \cdot\cdot & \cdot & \cdot \\ \cdot\cdot & \cdot & \cdot \\ x_{n,1} & x_{n,2} \dots & x_{n,d} \end{bmatrix} \quad (1)$$

where X represents the position matrix of the sparrow population; n is the number of sparrows; d is the dimension of the decision variables; and $x_{i,j}$ denotes the value of the

jth decision variable of the ith sparrow. In this study, each sparrow represents a candidate combination of CatBoost hyperparameters, and each element in the matrix corresponds to the value of one hyperparameter to be optimized.

Then the fitness value of each sparrow in the population can be expressed as shown in Equation (2):

$$X = \begin{bmatrix} f([x_{1,1}x_{1,2}\ldots x_{1,d}]) \\ f([x_{2,1}x_{2,2}\ldots x_{2,d}]) \\ \cdot \\ \cdot \\ \cdot \\ f([x_{n,1}x_{n,2}\ldots x_{n,d}]) \end{bmatrix} \tag{2}$$

where n is the number of sparrows; d is the dimension of the variables.

According to the different roles sparrows play during foraging, the producer's position is updated each time the population is updated, as shown in Equation (3):

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} exp\left(\frac{-i}{aN}\right), R_2 < ST \\ X_{i,j}^{t} + QL, R_2 \geq ST \end{cases} \tag{3}$$

where $X_{i,j}^{t+1}$ represents the position information of the i-th sparrow in the j-th dimension; t is the current iteration number; a is a random number in the range [0, 1]; N is the maximum number of iterations; Q is a random number following a standard normal distribution; L is a $1 \times d$ matrix with all elements equal to 1; R_2 is the alarm value, $R_2 \in [0, 1]$; ST is a safety value, $ST \in [0.5, 1]$ [13].

If R2 < ST, i.e., the alarm value is lower than the safety value, it indicates that the population is in a safe state, and the search range can be expanded. Once R2 $\geq$ ST, it indicates that the alert system has detected danger and communicated this information to the population. At this point, all sparrows must quickly evacuate to other safe areas to forage.

The scrounger position is updated as shown in Equation (4):

$$X_{i,\ j}^{t+1} = \begin{cases} Qexp\left(\frac{X_{worst} - X_{i,j}^{t}}{i^2}\right), i > \frac{n}{2} \\ X_{p}^{t+1} + \left|X_{i,j} - X_{p}^{t+1}\right| A^{+}L, otherwise \end{cases} \tag{4}$$

where Xp is the optimal position currently occupied by the producer; Xworst is the current global worst position; A is a $1 \times d$ matrix with elements randomly assigned 1 or -1, and $A^{+} = A^{T}\left(AA^{T}\right)^{-1}$. When i > n/2, it indicates that the i-th participant with a lower fitness value may be hungry and needs to fly elsewhere to forage for more energy. In other cases, the scrounger will search for food near the current producer's best position.

The guard position is updated according to Equation (5):

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta\left|X_{i,j}^{t} - X_{best}^{t}\right|, f_i > f_g \\ X_{i,j}^{t} + K\left|\frac{X_{i,j}^{t} - X_{worst}^{t}}{(f_i - f_w) + \varepsilon}\right|, f_i = f_g \end{cases} \tag{5}$$

where Xbest is the current global optimal position; β is the step size control parameter; K is a random number, $K \in [-1,1]$; $\mathbf{f_i}$ is the fitness value of the current sparrow individual; $\mathbf{f_g}$ and $\mathbf{f_w}$ are the global best and worst fitness values; ε is the smallest constant. When $\mathbf{f_i} = \mathbf{f_g}$, it indicates that those vigilant individuals can perceive the danger of their own position and will actively approach other sparrows to reduce the risk of being preyed upon; when

$\mathbf{f_i} > \mathbf{f_g}$, it indicates that the number of sparrows is approaching its limit, and they face a high risk of predator attack.

2.3. CatBoost Model

Categorical Boosting (CatBoost) is a high-performance machine-learning algorithm based on gradient boosted trees [17,18]. Its core design goal is to efficiently handle categorical features while avoiding gradient bias and prediction shift through an ordered boosting strategy, thereby improving model generalization while maintaining high accuracy. This study uses CatBoost to construct a harmful-gas risk level prediction model for non-coal strata tunnels, mainly based on its unique advantages in the following aspects:

(1) Symmetric Trees and Categorical Feature Processing

CatBoost uses symmetric trees as base learners, where each level splits based on the same feature, making model inference faster and less prone to overfitting. Its core innovation is native support for categorical features. Traditional Gradient Boosting Decision Tree (GBDT) models require preprocessing of categorical features (e.g., one-hot encoding or label encoding), while CatBoost directly converts them into numerical features using target statistics. Specifically, for a categorical feature k, its encoded value is calculated as follows.

$$x_{i,k} = \frac{\sum_{j=1}^{n} \left[x_{j,k} = x_{i,k}\right] \cdot Y_j + \alpha \cdot P}{\sum_{j=1}^{n} \left[x_{j,k} = x_{i,k}\right] + \alpha} \tag{6}$$

where P is the prior term (usually the target mean of the dataset), and α is the smoothing coefficient used to reduce noise from low-frequency categories.

(2) Ordered Boosting

Traditional GBDT models use the same dataset for multiple iterations when calculating gradients, which can lead to gradient bias accumulation. CatBoost proposes an ordered boosting mechanism, which achieves unbiased gradient estimation by randomly permuting the sample order and sequentially training multiple intermediate models. The specific process is as follows:

(1) Randomly permute the training set to generate a sequence $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$;
(2) For the i-th sample, use only the first i−1 samples to train an auxiliary model M_i to calculate the gradient for that sample;
(3) The final gradient estimate is based on the outputs of all auxiliary models, ensuring an unbiased gradient direction for each iteration.
(4) Loss Function for Multi-classification Tasks

For the multi-classification problem of predicting harmful-gas risk levels in non-coal strata tunnels, CatBoost uses the multi-class cross-entropy loss function:

$$L = -\sum_{i=1}^{N} \sum_{k=1}^{K} 1_{[y_i k]} log \left(\frac{e^{F_k(x_i)}}{\sum_{j=1}^{K} e^{F_j(x_i)}} \right) \tag{7}$$

where K is the number of classes (equivalent to the number of harmful-gas risk levels in non-coal strata tunnels in this study), and $F_k(x_i)$ is the raw output of the model for the k-th class. The output is transformed into a probability distribution via the Softmax function, and the loss function is optimized iteratively within the gradient boosting framework.

2.4. Cross-Validation

To improve the robustness and generalization ability of the prediction model, K-fold cross-validation was used for model training and evaluation. This method mitigates evaluation bias caused by a single data split through data resampling. Specifically, the training set was divided into K mutually exclusive subsets of equal size. In each iteration, one subset is selected as the validation set, and the remaining K-1 subsets are combined as the training set. The model is then trained, and its performance metrics are calculated. This process is repeated K times, ensuring each sample is used for validation once. Finally, model performance is estimated by the average of the K results. This ensemble evaluation strategy provides a more comprehensive and stable estimate of the model's performance on unseen data [9].

2.5. Classification Evaluation Metrics

To comprehensively evaluate the performance of the multi-classification model, this study uses Accuracy and F1-score as core evaluation metrics. Accuracy measures the proportion of correctly predicted samples overall and is intuitive. The F1-score combines Precision and Recall, providing a more robust evaluation when class distribution is imbalanced [12]. The complementary use of these two metrics helps systematically analyze the model's predictive effectiveness across different classes. Their calculation formulas are as follows:

$$precision = \frac{TP_i}{TP_i + FP_i} \tag{8}$$

$$recall = \frac{TP_i}{TP_i + FN_i} \tag{9}$$

$$F1\ score = \frac{2 \times P_i \times R_i}{P_i + R_i} \tag{10}$$

where TP represents true positives, FP represents false positives, FN represents false negatives, N is the total number of classes, and i represents the index of each class in the dataset.

$$accuracy = \frac{1}{n}\sum_{i=1}^{n} k, ii \tag{11}$$

where n represents the total number of samples in the dataset, and k,ii represents the number of correctly predicted samples in class i.

3. Development of the SSA-CatBoost Harmful Gas Prediction Model

3.1. Prediction Indicators and Scoring Criteria

For the classification scoring criteria of tunnels affected by oil- and gas-bearing strata, this study referred to the existing scoring frameworks proposed by Liu et al. [14] and Zhu et al. [15]. For the reservoir-performance characterization of the surrounding strata, empirical scoring was no longer used. Instead, measured porosity and permeability were adopted as petrophysical parameters to characterize gas storage and migration capacity. The scoring criteria for the remaining six indicators are provided in Supplementary Table S1. These scores were used only to quantify the input indicators K1–K6 for machine-learning modeling. The final harmful-gas risk levels were not determined by summing the scores in Supplementary Table S1 but were assigned according to field monitoring records, engineering investigation data, and relevant safety criteria. In addition, the definitions of K3 and K4 consider the influence of geological structures, fractures, slip surfaces, and caprock integrity on gas migration, sealing, and accumulation conditions [19,20] (see Figure 1 and Table 1).

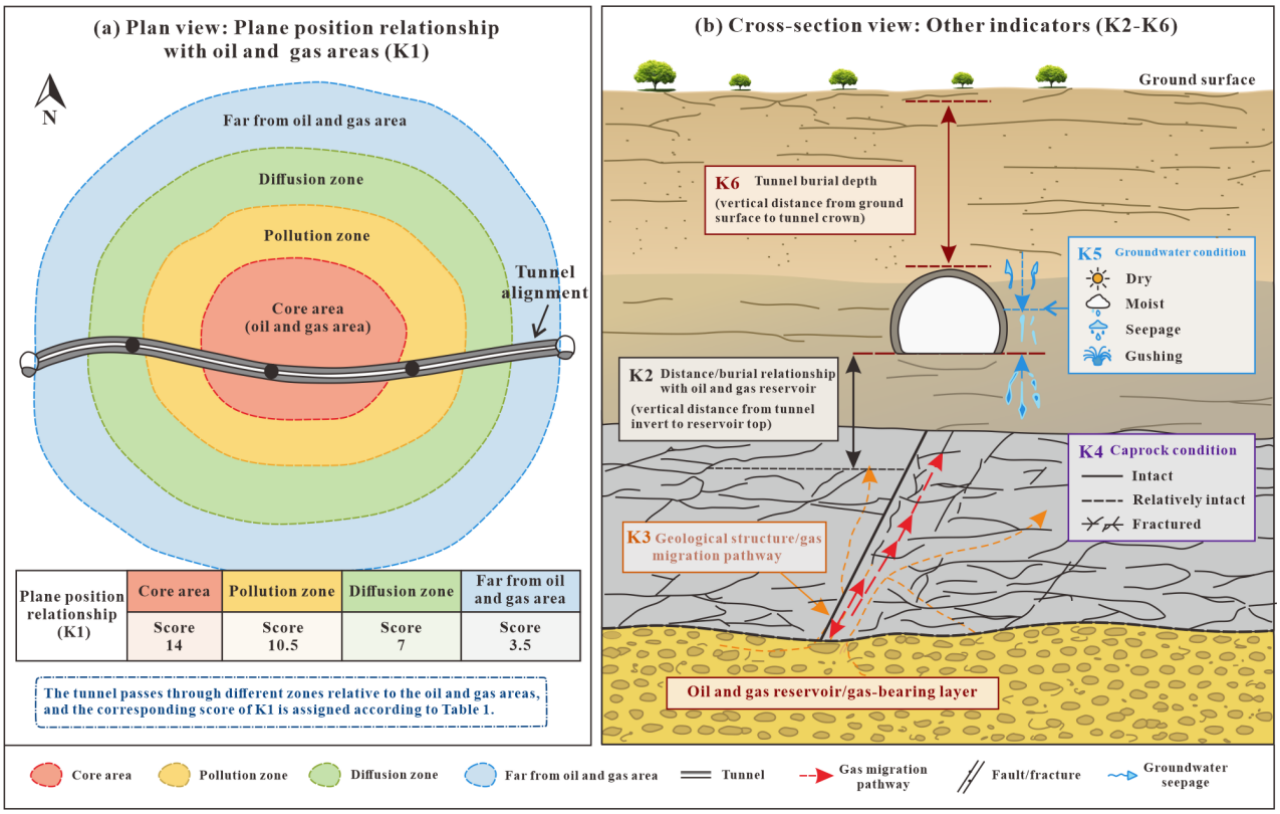


Figure 1. Schematic illustration of the semi-quantitative scoring indicators for harmful-gas risk assessment in tunnel sections. The six scoring-based indicators (K1–K6), together with the measured petrophysical parameters, namely porosity and permeability, constitute the eight input variables used in the model.

Table 1. Quantitative scoring table for gas sections in shallow natural gas tunnels.

Influencing Factors	Gas Section Characteristics	Risk Criterion and Score 1	Risk Criterion and Score 2	Risk Criterion and Score 3	Risk Criterion and Score 4
Oil and Gas Factors	Plane position relationship with oil and gas areas (K1)	Crossing the core area: 14	Crossing the pollution zone: 10.5	Crossing the diffusion zone: 7	Far from the oil and gas area: 3.5
Oil and Gas Factors	Relative burial-depth relationship with oil/gas reservoirs (K2)	0~100: 20	100~500: 15	500~1000: 10	>1000: 5
Geological Factors	Gas migration and accumulation condition of geological structures (K3)	Favorable for natural gas storage: 20	Favorable for natural gas escape: 15	Non-oil and gas migration direction area: 10	Not connected to oil and gas areas: 5
Geological Factors	Caprock characteristics (K4)	Intact stratum: 10	Relatively intact stratum: 7.5	Relatively fractured stratum: 5	Fractured stratum: 2.5

Table 1. *Cont.*

Influencing Factors	Gas Section Characteristics	Risk Criterion and Score 1	Risk Criterion and Score 2	Risk Criterion and Score 3	Risk Criterion and Score 4
Geological Factors	Groundwater conditions in the tunnel (K5)	Dry: 6	Moist: 4.5	Seepage: 3	Gushing/flowing water: 1.5
Engineering Factors	Tunnel burial depth (m) (K6)	>200: 20	100~200: 15	60~100: 10	<60: 5

3.2. Database Construction

This study extensively surveyed existing tunnel engineering data from the main canal and irrigation-area tunnels of the Tingzikou Irrigation Area Phase I Project and the Pidu River Water Supply Phase II Project. Geological information was collected, and scoring evaluation and physical-property tests were conducted for 138 harmful-gas tunnel cases. The actual harmful-gas risk levels of these cases were determined based on field monitoring and engineering records. The scoring process followed the evaluation criteria proposed in Section 3.1. Among the collected samples, there were 45 Level I samples (no risk, 32.6%), 38 Level II samples (slight risk, 27.5%), 30 Level III samples (moderate risk, 21.8%), and 25 Level IV samples (high risk, 18.1%). Some representative sample data are provided in Table S1.

3.3. SSA-CatBoost Tunnel Harmful Gas Prediction Model

In this paper, 138 collected harmful-gas tunnel samples were included in the SSA-CatBoost model for analysis. To reduce the influence of class imbalance on model evaluation, a stratified 80%/20% split was used to divide the dataset into a training set and an independent test set so that the class distribution was approximately preserved in both subsets. The training set contained 110 samples, and the independent test set contained 28 samples. The independent test set was retained only for final model evaluation and was not used during SSA hyperparameter optimization. The mapminmax function was used to normalize the input data before model training. Because the input variables have different units and numerical ranges, including scoring indicators, porosity, and permeability, normalization was applied to maintain preprocessing consistency and facilitate model comparison. It should be noted that CatBoost is a tree-based ensemble model, and monotonic normalization is not strictly required because tree-based splitting mainly depends on feature ordering and threshold selection. Therefore, normalization was mainly used for preprocessing consistency rather than as a necessary condition for improving model performance. Considering that the prediction performance of CatBoost is affected by several key hyperparameters, SSA was used to optimize the main CatBoost parameters in this study. The optimized parameters included the number of iterations, tree depth, learning rate, and L2 leaf regularization coefficient. The search ranges were set as follows: iterations = 50~300, depth = 3~8, learning rate = 0.01~0.30, and l2 leaf reg = 1~10. The SSA parameters were set as follows: population size = 30, maximum number of iterations = 50, safety threshold (ST) = 0.8, producer ratio = 0.2, and guard ratio = 0.1. During the optimization process, the cross-validation accuracy was used as the fitness function, and the parameter combination with the highest fitness value was selected as the optimal CatBoost parameter set. Based on this, the SSA-CatBoost harmful-gas risk prediction model was constructed (Figure 2).

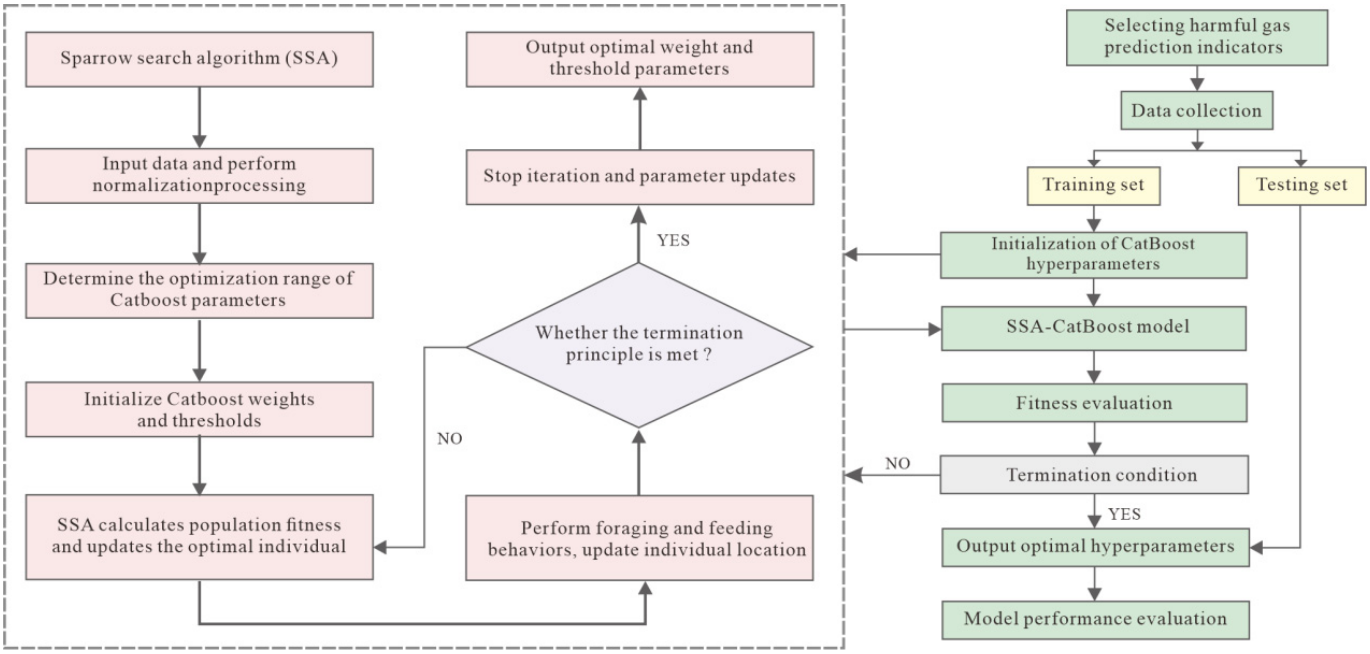


Figure 2. Flowchart of the SSA-CatBoost prediction model.

4. Results and Validation

4.1. Test Results

The 138 samples were divided into a training set and an independent test set using a stratified 80%/20% split. Five-fold cross-validation was performed on the training set for model training and evaluation, achieving an average accuracy of 93.63%. The independent test set contained 28 samples, among which 26 samples were correctly classified. The overall accuracy on the independent test set was therefore 92.86%. Considering the limited size of the independent test set, the Wilson 95% confidence interval of the test-set accuracy was calculated as 77.35–98.02%. This confidence interval reflects the uncertainty associated with the relatively small number of test samples. Table 2 presents the confusion-matrix-based evaluation metrics for each risk level on the independent test set.

Table 2. Confusion-matrix-based evaluation metrics for each risk level on the independent test set.

Harmful Gas Risk Level	Confusion Matrix Measurement Indicators							
	TP	FN	FP	TN	Class-Wise Accuracy (%)	Precision (%)	Recall (%)	F1-Value
Level I (no risk)	8	1	0	19	96.43	100.00	88.89	0.941
Level II (slight risk)	8	0	1	19	96.43	88.89	100.00	0.941
Level III (moderate risk)	5	1	0	22	96.43	100.00	83.33	0.909
Level IV (high risk)	5	0	1	22	96.43	83.33	100.00	0.909

As shown in Table 2, the one-vs-rest class-wise accuracy values were 96.43% for all four risk levels. The F1-scores for Levels I–IV were 0.941, 0.941, 0.909, and 0.909, respectively, indicating satisfactory multi-class prediction performance on the independent test set. The recall values for Level II and Level IV both reached 100%, suggesting that all samples in these two risk levels were correctly identified. The precision values for Level I and Level III both reached 100%, indicating that the model showed a low tendency to misclassify other risk levels as Level I or Level III.

For the remaining metrics, the recall for Level I was 88.89%, indicating that one Level I sample was misclassified into another risk category. The recall for Level III was 83.33%, suggesting that this risk level had relatively higher classification difficulty, possibly because of feature overlap with adjacent risk levels. Overall, the results of both five-fold cross-validation and independent test-set validation indicate that the SSA-CatBoost model achieved good classification performance for harmful-gas risk levels in non-coal strata tunnels. However, because of the limited size of the independent test set, the test-set results should be interpreted with appropriate caution.

4.2. Engineering Validation

To verify the predictive performance of the SSA-CatBoost harmful-gas prediction model in actual engineering, two other hydraulic tunnels were used for engineering validation. The Shizishan Tunnel is located in Qingshan Village, Peng'an County, Nanchong City. The tunnel mileage stake numbers are from Zong 131 + 956.64 to Zong 135 + 089.00, with a total length of 3235.59 m. It traverses the Jurassic Upper Suining Formation (J_3S), with lithology mainly interbedded sandstone and siltstone. The structure is primarily controlled by the east–west trending Dacheng, Guanyinchang, Nanchong, and Xianduhe anticlines. The tunnel is deeply buried (100–200 m) and belongs to the Yingshan oil and gas influence area. Its main oil-producing layers are the Lianggaoshan Formation, Sha-1 Member, and Daanzhai Member, with lithology mainly gray-black shale interbedded with fine siltstone and brown–gray shell limestone interbedded with black shale.

The Jinjiliang Tunnel project is located in Yingshan County, Nanchong City. It starts from the middle slope foot of the Xiaoshiti mountain, passes through Fengdou Village, Huashiban, Jinjiliang, Hejiagou, Mawangzhai, and ends at the middle slope of the Tu Diya mountain. The entrance stake number is Zong 116 + 949.19, and the exit stake number is Zong 121 + 231.24, with a total length of 4282.05 m and a maximum burial depth of 163.4 m. The strata traversed by the tunnel body are J3s silty mudstone interbedded with thin to medium-thick sandstone and argillaceous siltstone. The geological structure is simple with gentle rock layers. Joint fissures are developed, the rock mass is soft, and rock permeability ranges from slight to moderate. The rock mass softens easily when exposed to water. The formation itself lacks hydrocarbon generation capacity; the gas source is the Jurassic Lianggaoshan Formation and Ziliujing Formation source rocks. The scores, porosity and permeability test results, and prediction results for the two tunnels are shown in Table 3. The actual risk level was determined based on the maximum gas concentration obtained from on-site construction monitoring, compared with the US "Handbook for Methane Control in Mining" and the International Tunneling Association's "Guidelines for Good Occupational Health and Safety Practice in Tunnel Construction". The results show that all selected validation samples were correctly classified, suggesting that the model has potential applicability to additional engineering cases (Figures 3 and 4, Table 3).

Table 3. Engineering Validation Results.

Project	K1	K2	K3	K4	K5	K6	Porosity (%)	Permeability (10^{-3} μm^2)	Maximum Concentration from On-Site Monitoring (LEL%)	Actual Level	Predicted Level
Jinjiliang Tunnel 1	14	5	20	10	6	13	10.74	0.231	20.514	IV	IV
Jinjiliang Tunnel 2	14	5	20	10	6	14	3.44	0.1815	20.012	IV	IV

Table 3. *Cont.*

Project	K1	K2	K3	K4	K5	K6	Porosity (%)	Permeability (10^{-3} μm^2)	Maximum Concentration from On-Site Monitoring (LEL%)	Actual Level	Predicted Level
Jinjiliang Tunnel 3	14	5	20	10	6	13.5	9.03	0.252	25.020	IV	IV
Jinjiliang Tunnel 4	14	5	20	10	6	14	8.58	0.148	13.750	III	III
Jinjiliang Tunnel 5	14	5	20	7.5	4.5	5	2.85	0.0626	5.028	II	II
Shizishan Tunnel 1	6	5	5	2.5	3	2.5	8.76	0.0152	0.000	I	I
Shizishan Tunnel 2	5.5	3	16	7	3.52	13	7.65	0.105	7.370	II	II
Shizishan Tunnel 3	5.8	3	15.5	7.5	4	17	8.46	0.132	15.790	III	III
Shizishan Tunnel 4	6	3	9	8	3.5	13	4.39	0.0539	0.000	I	I
Shizishan Tunnel 5	6	3	17	7.5	4	17	8.33	0.118	17.110	III	III
Shizishan Tunnel 6	6.5	3	16	5	4	7.5	6.34	0.0289	2.840	II	II
Shizishan Tunnel 7	6	5	5	2.5	2.5	2.5	5.92	0.0156	0.000	I	I

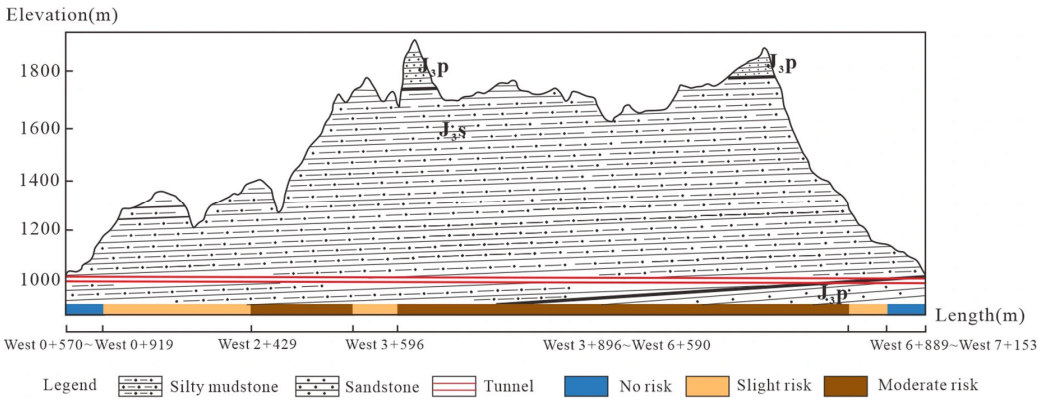


Figure 3. Profile and gas zoning map of the Shizishan Tunnel.

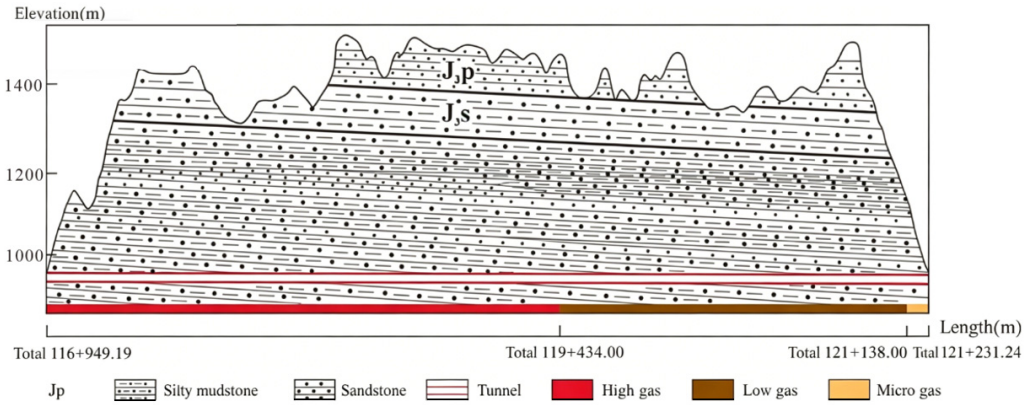


Figure 4. Profile and gas zoning map of the Jinjiliang Tunnel.

5. Discussion

5.1. Comparison with Other Machine-Learning Algorithms

To further evaluate the statistical performance of the proposed SSA-CatBoost model, five-fold cross-validation was performed on SSA-CatBoost, CatBoost, SSA-SVM, and SSA-XGBoost. The results are shown in Table 4. Among all compared models, SSA-CatBoost achieved the highest average accuracy of 93.63%, followed by ANN (90.91%), PSO-Catboost (88.18%), SSA-XGBoost (86.36%), CatBoost (85.45%), and SSA-SVM (81.82%). Furthermore, the 95% confidence interval for SSA-CatBoost was 90.54–96.72%, indicating that the model exhibited relatively stable predictive performance on the current dataset (Table 4).

Table 4. Statistical results of five-fold cross-validation for different models.

Model	Fold 1 (%)	Fold 2 (%)	Fold 3 (%)	Fold 4 (%)	Fold 5 (%)	Mean Accuracy (%)	Std (%)	95% CI (%)
SSA-Catboost	95.45	90.90	90.90	95.45	95.45	93.63	2.49	90.54–96.72
Catboost	95.45	86.36	77.27	86.36	81.82	85.45	6.74	77.08–93.82
SSA-SVM	81.82	81.82	77.27	86.36	81.82	81.82	3.21	77.83–85.81
SSA-XGBoost	81.82	86.36	90.90	86.36	86.36	86.36	3.21	82.37–90.35
ANN	90.90	86.36	90.90	90.90	95.45	90.91	3.21	86.91–94.89
PSO-Catboost	81.82	86.36	90.90	90.90	90.90	88.18	4.06	83.13–93.22

To further test the statistical advantage of SSA-CatBoost compared to the other models in cross-validation, paired *t*-tests were conducted between SSA-CatBoost and the other models based on the accuracy of each fold. The results are shown in Table 5. The results indicate that the differences between SSA-CatBoost and CatBoost (p = 0.037), SSA-SVM (p = 0.00045), and SSA-XGBoost (p = 0.035) are all statistically significant at the 0.05 significance level.

Table 5. Paired *t*-test results between SSA-Catboost and comparison models.

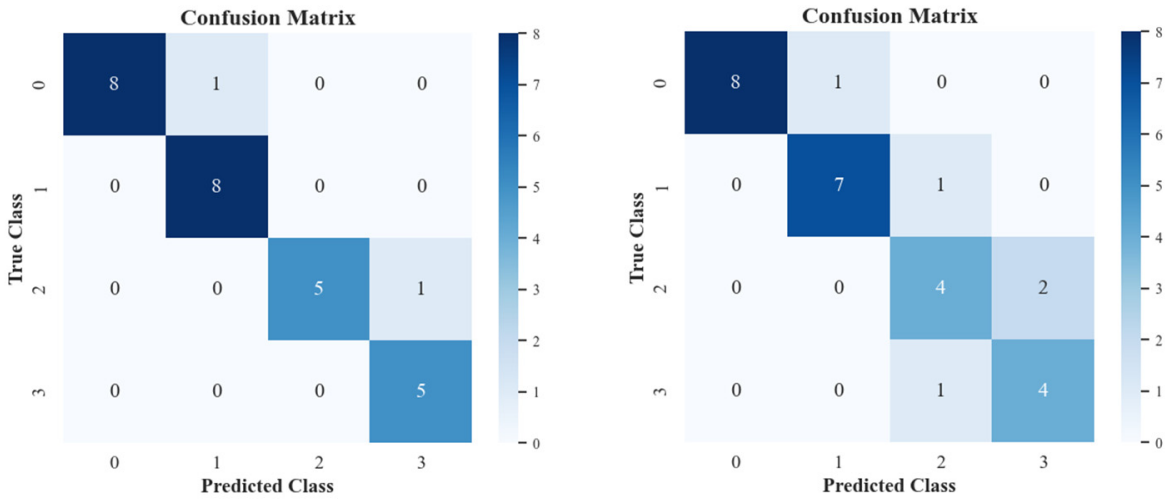
Comparison	Mean Difference (%)	t Value	*p* Value
SSA-Catboost vs. Catboost	8.18	3.09	0.037
SSA-Catboost vs. SSA-SVM	11.81	10.61	0.00045
SSA-Catboost vs. SSA-XGBoost	7.27	3.14	0.035

5.2. Indicator Importance Evaluation

To verify the improvement in model performance brought by replacing empirical reservoir-performance scoring with measured porosity and permeability, an ablation comparison was conducted between the scoring-based reservoir-performance variant and the measured-parameter variant. In the scoring-based variant, reservoir performance was represented by empirical scores, whereas in the measured-parameter variant, this empirical score was replaced by measured porosity and permeability.

The independent test-set results are shown in Figure 5 and Table 6. Compared with the model using empirical reservoir-performance scoring, the measured-parameter variant showed improved classification performance on the independent test set. The recognition result for the no-risk category remained largely stable, whereas the precision, recall, and F1-score for the slight-risk, moderate-risk, and high-risk categories improved to varying degrees. Among them, the improvement for the moderate-risk and high-risk categories was the most evident, with their F1-scores increasing from 0.667 and 0.727 to 0.909 and 0.909, respectively. This result indicates that converting reservoir performance from empirical scores to measurable petrophysical parameters, namely porosity and permeability, helps

reduce information loss caused by subjective assignment, enhances the model's ability to characterize class boundaries, and particularly improves the discrimination of adjacent risk levels.



(**a**) (**b**)

Figure 5. Confusion matrices of the SSA-Catboost independent test set before and after replacing indicators: (**a**) after indicator replacement; (**b**) before indicator replacement.

Table 6. Confusion matrix evaluation results for each category in the independent test set before indicator replacement.

Harmful Gas Risk Level	Evaluation Metrics of the Confusion Matrix							
	TP	FN	FP	TN	Class-Wise Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Level I (No risk)	8	1	0	19	96.43	100.00	88.89	0.941
Level II (Slight risk)	7	1	1	19	92.86	87.50	87.50	0.875
Level III (Moderate risk)	4	2	2	20	85.71	66.67	66.67	0.667
Level IV (High risk)	4	1	2	21	89.29	66.67	80.00	0.727

To further examine whether this improvement was only caused by a single independent test split, five-fold cross-validation was also conducted for the two reservoir-performance variants. As shown in Table 7, the scoring-based reservoir-performance variant achieved a mean accuracy of 91.81%, whereas the measured-parameter variant achieved a higher mean accuracy of 93.63%. This result suggests that the performance improvement after introducing porosity and permeability was also reflected in cross-validation.

Table 7. Five-fold cross-validation comparison between the scoring-based and measured-parameter reservoir-performance variants.

Variant	Fold 1 (%)	Fold 2 (%)	Fold 3 (%)	Fold 4 (%)	Fold 5 (%)	Mean Accuracy (%)	Std (%)	95% CI (%)
Measured-parameter variant	95.45	90.90	90.90	95.45	95.45	93.63	2.49	90.54–96.72
Scoring-based reservoir-performance variant	90.90	90.90	90.90	90.90	95.45	91.81	2.03	89.28–94.34

It should be noted that high harmful-gas risk does not necessarily correspond to high matrix permeability. Some high-risk samples showed very low permeability, such as 0.0085×10^{-3} μm^2, indicating that permeability alone cannot determine the final risk level. Physically, low matrix permeability may limit gas dissipation and favor gas retention under sealed conditions. During tunnel excavation, however, faults, fractures, bedding-parallel slip surfaces, or excavation-induced damage may provide local release pathways for the retained gas. Therefore, a high-risk level may still occur in low-permeability strata when strong gas-source proximity, favorable structural migration or accumulation conditions, effective caprock sealing, and sufficient tunnel burial depth are present. This result further indicates that harmful-gas accumulation and release in non-coal strata tunnels are controlled by multi-factor coupling rather than by a single petrophysical parameter.

To further improve the interpretability of the SSA-CatBoost model, SHAP analysis was conducted to evaluate the contribution of each input variable to the prediction of different harmful-gas risk levels. The class-wise SHAP summary plots are shown in Figure 6. In the SHAP plots, a positive SHAP value indicates that the corresponding feature increases the model output for a given risk level, whereas a negative SHAP value indicates a decreasing effect.

For Level I, which represents no gas risk, K3 and K1 were the most influential variables. Low values of K1 and K3 generally produced positive SHAP values for Level I, indicating that tunnel sections far from oil/gas areas and without favorable gas migration or accumulation structures were more likely to be classified as no-risk sections. In contrast, high values of K1 and K3 tended to reduce the probability of Level I classification.

For Level IV, which represents high gas risk, K1, K3, K4, and K6 showed the strongest influence. High values of these indicators generally contributed positive SHAP values, suggesting that proximity to oil/gas areas, favorable gas migration and accumulation conditions, better caprock sealing conditions, and greater tunnel burial depth increased the probability of high-risk classification. This result is consistent with the geological understanding that harmful-gas risk is jointly controlled by gas-source proximity, structural migration pathways, sealing conditions, and engineering burial depth.

For Level II and Level III, the SHAP distributions were more dispersed, and several indicators showed overlapping effects. This indicates that the intermediate risk levels were controlled by multiple coupled factors and had more pronounced feature overlap with adjacent classes. This result also explains why the classification difficulty of the intermediate risk levels was greater than that of the no-risk and high-risk levels.

Porosity and permeability did not dominate the prediction results alone, but they contributed to the discrimination of different risk levels, especially for the moderate- and high-risk classes. Their inclusion provides measured petrophysical information on gas storage and migration capacity, supplementing the semi-quantitative geological and engineering indicators. Therefore, the SHAP results further support the replacement of empirical reservoir-performance scoring with measured porosity and permeability in the proposed model.

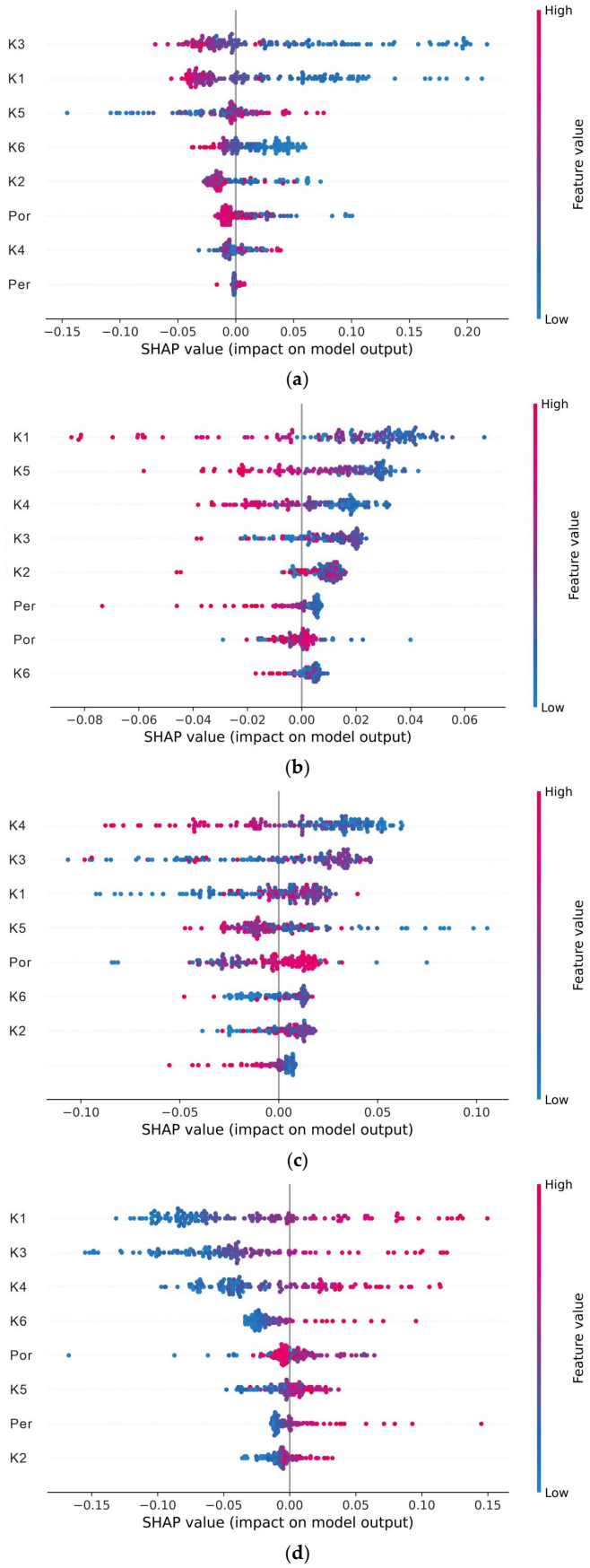


Figure 6. Class-wise SHAP summary plots of the SSA-CatBoost model for different harmful-gas risk levels: (**a**) Level I, no risk; (**b**) Level II, slight risk; (**c**) Level III, moderate risk; and (**d**) Level IV, high risk.

5.3. Limitations

This study still has certain limitations. First, constrained by data sources, the research sample size is relatively limited for a four-class classification problem. Although 138 harmful-gas tunnel samples were collected, the smallest class, Level IV, contained only 25 samples. In addition, the independent test set contained only 28 samples, meaning that a single misclassification may noticeably affect the reported accuracy and class-wise metrics. Therefore, although the model achieved good performance on the current dataset, the independent test-set results should be interpreted with caution. Future work should incorporate more engineering cases, especially high-risk samples, and supplement test data for key parameters such as permeability and porosity to enhance the robustness and applicability of the model.

Second, the current model evaluation strategy still has limitations. The independent test set was retained only for final model evaluation and was not used during SSA hyperparameter optimization. However, the five-fold cross-validation procedure does not fully separate hyperparameter tuning from model evaluation in the same way as nested cross-validation. Therefore, the cross-validation results may still contain some optimistic bias. Future studies should adopt repeated or nested cross-validation when larger datasets become available.

Third, the quantification of the evaluation indicator system remains incomplete. Although this study quantitatively improved the reservoir-performance indicator by introducing porosity and permeability, other indicators were still incorporated into the modeling process in the form of scores or grades. Parameters such as RQD, in situ stress, joint density, and other rock-mass integrity indicators may also influence gas migration, storage, and excavation-induced release. However, complete and consistent data for these parameters were not available for all samples in the current database. Subsequent research should further promote the development of the evaluation system from empirical scoring toward measured-parameter and mechanism-constrained characterization, thereby enhancing the scientificity and interpretability of the model.

6. Conclusions

(1) A harmful-gas risk level prediction model based on CatBoost was developed, and the Sparrow Search Algorithm (SSA) was introduced to optimize the key hyperparameters of the model, thereby improving its classification performance and generalization ability under multi-factor coupling conditions. Finally, the SSA-CatBoost harmful-gas risk level prediction model was constructed.

(2) Eight key input parameters were selected to construct a database containing 138 real harmful-gas tunnel cases, in which reservoir performance was quantitatively characterized by permeability and porosity, and the remaining indicators were characterized by scoring methods. Based on this, the SSA-CatBoost prediction model was established. The results show that the model achieved an average accuracy of 93.63% in five-fold cross-validation and a prediction accuracy of 92.86% on the independent test set, demonstrating high prediction accuracy and good generalization ability.

(3) The SSA-CatBoost model was systematically compared with other machine-learning models. The results show that the proposed model has higher prediction accuracy and superior comprehensive classification performance, reflecting good engineering applicability. Furthermore, compared with the model constructed using the scoring method to characterize reservoir performance, the improved model incorporating permeability and porosity showed enhanced recognition performance across all risk levels, indicating that the quantitative characterization of reservoir performance parameters contributes to improving the objectivity and accuracy of harmful-gas risk prediction.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/pr14132204/s1, Table S1: Representative sample data used in the harmful gas risk-level prediction database.

Author Contributions: Z.M.: Writing—original draft, Validation, Software, Methodology. W.L.: Writing—review & editing, Supervision, Resources. J.W.: Writing—review & editing, Supervision. P.S.: Writing—review & editing. X.W.: Data curation. P.Y.: Data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Science and Technology Special Project of Sichuan Province: Key Technology Research on Digital Preservation and Resource Utilization of Major Railway Physical Data (2024ZDZX0009).

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: Authors Zengchan Mao, Jianhua Wu and Xiaojin Wang are employed by Sichuan Water Development Investigation, Design & Research Co., Ltd., authors Wenpin Luo and Peng Yang are employed by Sichuan Tingzikou Irrigation Area Construction and Development Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Bi, S.; Shao, L.; Qi, Z.; Wang, Y.; Lai, W. Prediction of Coal Mine Gas Emission Based on Hybrid Machine Learning Model. *Earth Sci. Inform.* **2023**, *16*, 501–513. [CrossRef]
2. Copur, H.; Cinar, M.; Okten, G.; Bilgin, N. A Case Study on the Methane Explosion in the Excavation Chamber of an EPB-TBM and Lessons Learnt Including Some Recent Accidents. *Tunn. Undergr. Space Technol.* **2012**, *27*, 159–167. [CrossRef]
3. ITA. *Guidelines for Good Occupational Health and Safety Practice in Tunnel Construction*; ITA Report No. 001; ITA: Lausanne, Switzerland, 2008.
4. Jaffe, W.; Lockyer, R.; Howcroft, A. The Abbeystead Explosion Disaster. *Ann. Burn. Fire Disasters* **1997**, *10*, 1–4.
5. Kang, X.B.; Xu, M.; Luo, S.; Xia, Q. Study on Formation Mechanism of Gas Tunnel in Non-Coal Strata. *Nat. Hazards* **2013**, *66*, 291–301. [CrossRef]
6. Proctor, R.J. The San Fernando Tunnel Explosion, California. *Eng. Geol.* **2002**, *67*, 1–3. [CrossRef]
7. Tang, S.-H.; Zhang, X.-P.; Liu, Q.-S.; Xie, W.-Q.; Wu, X.-L.; Chen, P.; Qian, Y.-H. Control and Prevention of Gas Explosion in Soft Ground Tunneling Using Slurry Shield TBM. *Tunn. Undergr. Space Technol.* **2021**, *113*, 103963. [CrossRef]
8. Kissell, F.N. *Handbook for Methane Control in Mining*; Information Circular 9486; National Institute for Occupational Safety and Health, Pittsburgh Research Laboratory: Pittsburgh, PA, USA, 2006.
9. Ruilin, Z.; Lowndes, I.S. The Application of a Coupled Artificial Neural Network and Fault Tree Analysis Model to Predict Coal and Gas Outbursts. *Int. J. Coal Geol.* **2010**, *84*, 141–152. [CrossRef]
10. Wang, W.; Wang, H.; Zhang, B.; Wang, S.; Xing, W. Coal and Gas Outburst Prediction Model Based on Extension Theory and Its Application. *Process Saf. Environ. Prot.* **2021**, *154*, 329–337. [CrossRef]
11. Lin, H.; Li, W.; Li, S.; Wang, L.; Ge, J.; Tian, Y.; Zhou, J. Coal Mine Gas Emission Prediction Based on Multifactor Time Series Method. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110443. [CrossRef]
12. 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*, 127389. [CrossRef]
13. Jolfaei, S.; Lakirouhani, A. Sensitivity Analysis of Effective Parameters in Borehole Failure, Using Neural Network. *Adv. Civ. Eng.* **2022**, *2022*, 4958004. [CrossRef]
14. Liu, Y.; Su, P.; Qiu, P.; Luo, T.; Yang, C.; Lu, X. Research on Gas Tunnel Prediction in Central Sichuan Using Energy Valley Optimizer and Support Vector Machine. *Bull. Eng. Geol. Environ.* **2025**, *84*, 50. [CrossRef]
15. Zhu, X.; Su, P.; Yu, J.; Pei, J.; Teng, Z.; Li, Y.; Liu, Y. A Prediction Model for Hazard Levels of Shallow Natural Gas in Tunnel Based on K-Means Clustering and Tabular Prior-Data Fitted Network. *Results Eng.* **2025**, *27*, 106873. [CrossRef]
16. Xue, J.; Shen, B. A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm. *Syst. Sci. Control. Eng.* **2020**, *8*, 22–34. [CrossRef]
17. Xue, Y.; Bai, C.; Qiu, D.; Kong, F.; Li, Z. Predicting Rockburst with Database Using Particle Swarm Optimization and Extreme Learning Machine. *Tunn. Undergr. Space Technol.* **2020**, *98*, 103287. [CrossRef]
18. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *arXiv* **2019**, arXiv:1706.09516.

19. Zhu, H.; Qi, B.; Li, J.; Li, C.; Raza, A.; Guo, C. Unlocking the Shiny Surface Features of Shale Shear Fractures at Micro-Nanoscale. *Adv. Geo-Energy Res.* **2025**, *18*, 202–206. [CrossRef]
20. Zhu, H.; Ju, Y.; Lu, Y.; Yang, M.; Feng, H.; Qiao, P.; Qi, Y. Natural Evidence of Organic Nanostructure Transformation of Shale during Bedding-Parallel Slip. *Bull. Geol. Soc. Am.* **2025**, *137*, 2719–2746. [CrossRef]