RESEARCH ARTICLE

Construction of a geological identification model for small faults in coal fields based on Bayes-XGBoost

Jieqi Liu*

Human Resource Department, Shaanxi Energy Institute, Xianyang, Shaanxi, China.

Received: February 28, 2024; **accepted:** April 16, 2024.

Geologic identification of small faults in coalfields is an important issue in coal mining. The existence and nature of small faults have an important impact on mine safety and coal mining efficiency. To solve the problem of inaccurate geological identification of small faults in coalfields, this study proposed a method based on Bayes algorithm and limit gradient lifting to construct an efficient and accurate geological identification model of small faults in coalfields by combining the advantages of Bayesian theory and limit gradient boosting algorithm to achieve accurate identification of small defects by probabilistic modeling of data using Bayesian theory. Meanwhile, the study utilized the efficient feature selection and model construction ability of the limit gradient boosting algorithm, which made the proposed model capable to not only handle high-dimensional data, but also overcome the noise and incompleteness problems in the data, thus effectively performing the geological identification of small faults. The results showed that the model's average accuracies on the two experimental datasets were 93.6% and 92.4%, respectively, while the average F1 values of the model were 90.6% and 93.0%, respectively, which were all better than that of control model. The results suggested that the proposed model was advanced and addressed the issues of low efficiency and time consumption in traditional coalfield exploration methods, which offered a new detection technology for coalfield exploration and mining, as well as promoted the development of related industries.

Keywords: Bayes algorithm; XGBoost framework; recognition model; coalfield minor fault; parameter tuning.

***Corresponding author:** Jieqi Liu, Human Resource Department, Shaanxi Energy Institute, Xianyang 712000, Shaanxi, China. Emails: 15109113736@163.com.

Introduction

Coal, one of the world's most significant energy resources, is crucial to both economic growth and the availability of electricity. However, the existence of fault geology in the exploration and mining process of coal fields poses great challenges to mine safety and coal resource evaluation [1]. The existence of coalfield minor fault (CMF) brings multiple effects on mine mining and coal resource evaluation, such as deformation and rupture of coal seams, changes in distribution and permeability of coalbed methane, and increased safety risk of coal mining [2]. In addition, the existence of small faults may lead to the instability of the coal mine working face and the expansion of coal seam fissures, thus affecting the productivity and safety of the mine [3]. Conventional techniques for identifying fault geology mostly depend on engineering surveys and geological exploration, which includes seismic surveys, geological surveys, and geological exploration boreholes. However, these methods have problems such as

inconsistent data quality, difficulty in feature selection, and complex model construction [4]. Accurately identifying and evaluating the location and nature of small faults can aid in formulating plans for coal resource exploration and mining and improve resource utilization efficiency. However, small faults in coalfields often have characteristics such as large dip angles, short lengths, small displacements, and complex shapes, which make comprehensive detection and identification difficult using ordinary methods [5].

The eXtreme gradient boosting (XGBoost) algorithm is a popular technology that has been applied in many fields and has been optimized and improved by many researchers to address its shortcomings. Asselman *et al*. found that Bayes algorithm (BA) demonstrated high prediction accuracy and tried to apply this algorithm to analyze students' behavior during the learning process. They proposed an XGBoost based PFA method which gave the PFA model a better scalability through the XGBoost framework to improve the prediction accuracy of students' performance. The results showed that the proposed scalable XGBoost-PFA method was experimentally validated to outperform other assessment models and significantly improved the prediction accuracy of student achievement [6]. Budholiya *et al*. proposed a diagnostic system that utilized an optimized XGBoost framework as the core to construct a classifier to predict heart disease. The results indicated that the proposed model had higher accuracy than most of the same type methods [7]. Osman *et al*. proposed an accurate groundwater level prediction model to study the groundwater level in Malaysia using Xgboost algorithm and artificial neural network to predict the local groundwater level for 11 months. The research compared the accuracy of different model outputs using local rainfall, temperature, and evapotranspiration as inputs and a support vector machine model as a control. The results showed that the accuracy of the proposed model reached 92%, which was 11% higher than that of control model [8]. A new deep learning model for the classification problem using convolutional limit gradient was described by Thongsuwan *et al*. based on the XGBoost algorithm and convolutional neural networks. The model demonstrated good generalizability. In addition to processing image data, the proposed model could support classification and prediction after feature learning. The results showed that the performance of the proposed model was superior to that of the convolutional and XGBoost models alone [9]. Zhu *et al*. proposed a robust probabilistic machine learning model to predict the distribution of rockheads using spatial geographic information. The model combined natural gradient boosting with XGBoost to construct the basic learner and predict the spatial information of rockheads, addressing the problem of uncertainty in their variation. The results indicated improved accuracy in predicting rockhead spatial information compared to previous neural networks [10].

The Bayesian optimization algorithm is a method of probability inference based on the Bayesian theorem, which can be applied to various machine learning problems including probability inference, classification, regression, and clustering. The algorithm is one of the most popular algorithms at present and hence several researchers have optimized it. Wickramasinghe and Kalutarage attempted to classify medical products based on BA. A simple but efficient model known as plain BA classification model was developed. Since the assumptions of BA could not fully fit the real data, the plain BA utilized dynamic changes to satisfy the general data. The results suggested that the plain BA classification model showed strong robustness in numerous classification tasks [11]. Liu *et al*. found that the experience of random effects in the multilevel model always deviated from the overall mean. Therefore, they proposed an empirical prediction model based on BA. The model utilized BA to update and adjust the random effects in the multilevel model to reflect its real situation more accurately. By introducing priori information and real-time data, the model

could gradually correct the empirical predictions of the random effects. The results proved that the proposed model showed high accuracy in applications [12]. Jospin *et al*. constructed an efficient assessment tool based on Bayes' quantitative features and deterministic fusion of deep learning algorithms. The uncertainty associated with deep neural network prediction was understood and quantified through BA, which greatly reduced the computational complexity of the model. Tests showed that, compared to conventional deep learning algorithms, the model performed better in terms of accuracy and computational speed [13]. Sun and Zhou found that convex optimization algorithms were useful in solving the problems faced by BA in applications and used biased random key genetic algorithm to improve BA. The performance of the proposed algorithm was enhanced by an improvement procedure that employed the local optimization model as its decoder, while an XSS attack was used to assess the rebuilt model's cybersecurity. The enhanced algorithm performed better in cybersecurity [14]. Laumont *et al*. proposed a Monte Carlo sampling scheme to optimize BA. The traditional BA often encountered the problems of large computation and slow convergence when dealing with large-scale datasets, which made the algorithm very limited in practical applications. The optimized BA solved the problem of no known proof of convergence, and at the same time, it had a large improvement in terms of computational speed and accuracy [15].

This study utilized the previous research data to construct an efficient model for identifying small faults in coal fields using Bayes-XGBoost algorithm to address the limitations of conventional methods for detecting small faults in coal fields, which are inefficient and time-consuming. To achieve fast and efficient identification of small faults in coal fields, a Bayesian optimization method was explored to automatically adjust the hyperparameters of the XGBoost model, which improved the performance and generalization ability of the model. By automatically adjusting the hyperparameters, the method could be better adapted to different datasets and task requirements, thus improving the accuracy and stability of the model. The results of this study would solve the fault geology identification problem and could enhance the efficiency of the coal mining surveying industry, promoting its development. Additionally, it would offer guidance for geological identification issues in other fields.

Materials and methods

Geological recognition model construction based on XGBoost framework

XGBoost is a powerful machine learning framework that is based on the gradient boosting tree algorithm [16]. It performs well in many data science competitions and real-world applications and is widely used in tasks such as regression, classification, sorting, and recommender systems [17]. XGBoost framework has better recognition classification results for small scale targets, so the algorithm could be used for CMF recognition. The structure of XGBoost framework was shown in Figure 1. The core of the XGBoost framework was the gradient boosting tree, which was an integrated learning method that progressively improved the performance of the model by iteratively training multiple decision trees [18]. Each decision tree was trained on the residuals of the previous tree, thus gradually reducing the prediction error of the model. The gradient boosting tree computation process was shown in Figure 2. However, the XGBoost framework has defects such as difficulty in parameter tuning and overfitting, so it is necessary to improve the XGBoost framework. The original objective function of XGBoost is shown in equation (1).

$$F(x)=\sum_{i=1}^{n} l\left(y_i,\overline{y}_i\right)+\alpha T+\frac{1}{2}\beta\sum_{j=1}^{T} w_j^2 \qquad (1)$$

where $l\left(y_i,\overline{y}_i\right)$ was the loss function of the XGBoost framework. x was the input sample.

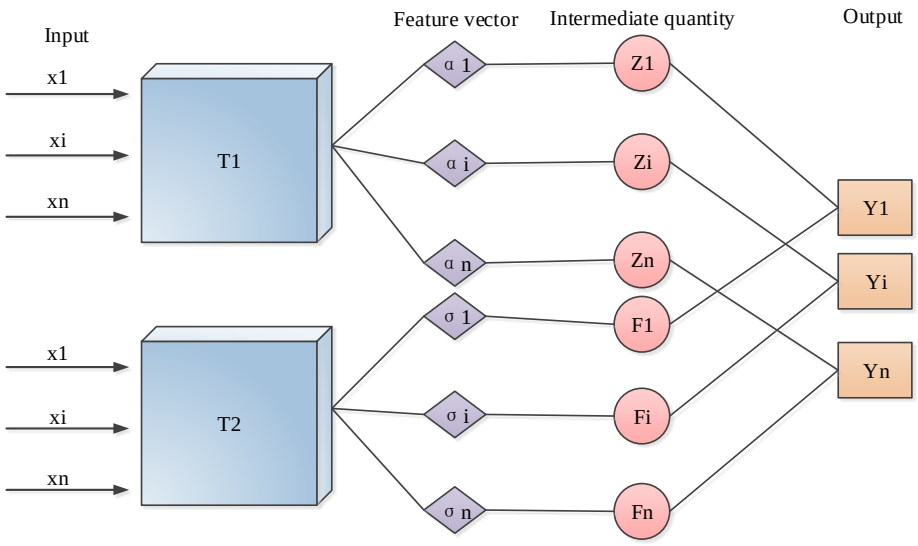

Figure 1. XGBoost framework structure.

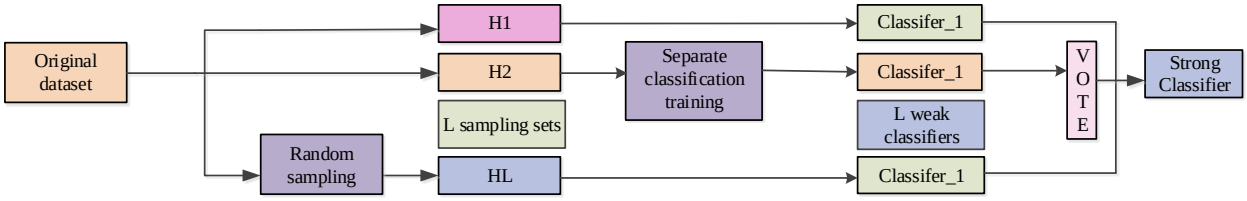

Figure 2. Gradient boosting tree calculation flowchart.

y_i was the true value of the sample. $\overline{y}_i$ was the predicted value of the input sample. T was the number of nodes. n was the input samples. α was the paradigm coefficient of the node shrinkage coefficient. β was the node shrinkage coefficient. w was the weight matrix. The goal of binary classification, a popular machine learning job, was to divide data samples into two groups or labels. In a binary classification, each sample could only belong to one of these two categories. After binary classification classified the input samples, it reduced the computational cost of XGBoost in recognition and played a positive role in the recognition efficiency of the model [18]. After introducing binary classification, the objective function would increase the weights of certain samples. The optimized objective function would no longer adjust the parameters using the previous loss function, but instead use the cross-entropy function. The computational expression of the cross-entropy function was shown in equation (2).

$$s = -\frac{1}{n}\sum_{i=1}^{n} y_i \log^{(p_i)} + (1-y_i)\log^{(1-p_i)} \qquad (2)$$

where p_i was the distribution probability of the sample predicted values. The distribution of the model's anticipated probabilities for each category was the distribution probability of the sample predicted values. The cross-entropy function was used to measure the difference between the actual samples and the model predictions, which included the model's predicted probabilities for each category [19]. The predicted probability distribution was gradually optimized by minimizing the cross-

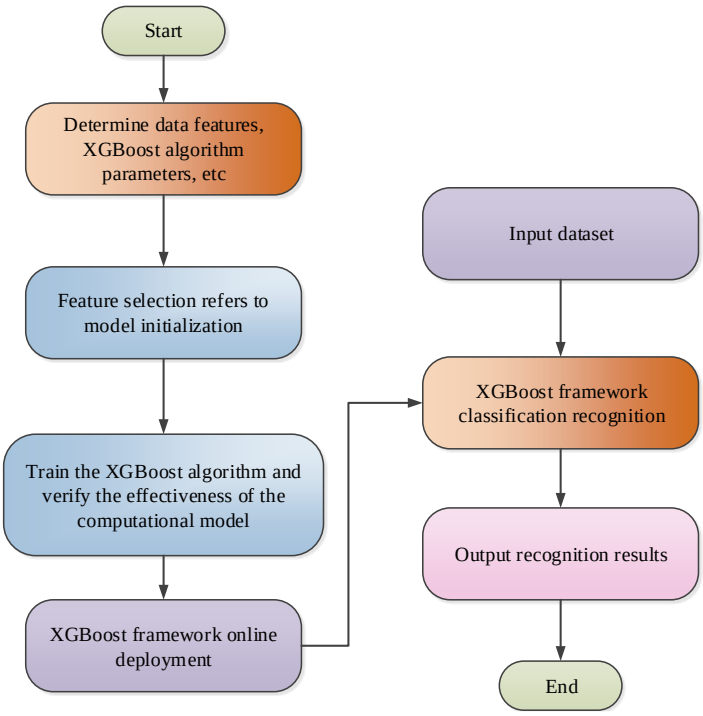

Figure 3. XGBoost framework runtime flowchart.

entropy loss during model computation to make it closer to the distribution of the actual labels. The formula for calculating the distribution probability of the sample predicted values was shown in equation (3).

$$p_i = \frac{1}{1+e^{-\bar{y}_i}} \tag{3}$$

The cross-entropy function was relatively simple for the gradient calculation of model parameters, which could further reduce the computational complexity of the model. After the introduction of the cross-entropy function, XGBoost could recognize CMFs more quickly and did the real-time feedback needs. The optimized objective function expression was shown in equation (4).

$$F'(x) = -(1+Ky_i)\left[\sum_{i=1}^{n} y_i \log^{(p_i)} + (-y_i+1)\log^{(1-p_i)}\right] \tag{4}$$

where K was the weight of the true value of the sample. Since the improved objective function could realize the adjustment of the weights, the model could change the weights according to the characteristics of the input data to achieve a better recognition effect. The expression of weights adjustment was shown in equation (5).

$$w' = -\frac{G_i}{H_i + \lambda} \tag{5}$$

where G_i was the first order derivative sum of the regularization term $\phi(x_i)$. H_i was the sum of second order derivatives of $\phi(x_i)$. The expressions for G_i and H_i were shown in equation (6).

$$\begin{cases} G_i = \sum_{i=1}^{n} g_i \\ H_i = \sum_{i=1}^{n} h_i \end{cases} \tag{6}$$

where g_i and h_i were the first order derivatives as well as the second order derivatives of $\phi(x_i)$, respectively. λ was the node weight coefficients. Where the expression of the regularization term $\phi(x_i)$ was shown in equation (7).

$$\phi(x_i) = \gamma T + \lambda \sum_{j=1}^{T} w_j^T \quad (7)$$

where γ was $L2$ paradigm and λ was node weight coefficient. After optimization, the XGBoost framework had advantages in terms of performance, scalability, accuracy, and robustness. The final output function of XGBoost was shown in equation (8).

$$f_{out}(x) = \sum_{i=1}^{n} l(y_i, \overline{y}_i) + \sum_{i=1}^{T} \phi(x_i) \quad (8)$$

At this point, the construction of the geo-RM of CMF based on XGBoost framework was completed, and the operation flowchart of this model was shown in Figure 3. It was found in the study that XGBoost had certain shortcomings when recognizing CMFs. For example, the model might be interfered by data noise when dealing with complex geologic structures, leading to a decrease in prediction accuracy. In addition, XGBoost's ability to recognize different types of small fault lands varied, and the model might misjudge or miss some special types of small fault lands.

Construction of geological recognition model for XGBoost framework based on Bayes optimization approach

Bayes optimization is an iterative method for optimizing an objective function [20]. It is suitable for those cases where the objective function is difficult to optimize directly, such as highly nonlinear, noisy disturbances, or black-box functions (without explicit expressions) [21]. The core idea of Bayes optimization is to use Bayes' theorem to model the objective function and use a posteriori probability to guide the next optimization decision. The schematic diagram of BA is shown in Figure 4.

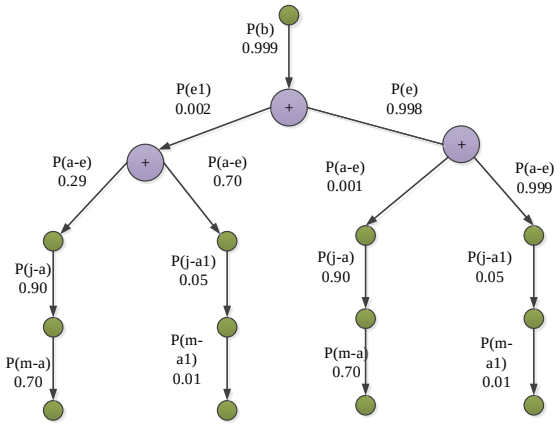

Figure 4. Schematic diagram of Bayes algorithm.

In each iteration, Bayes optimization synthesized previous observations and prior knowledge to construct a Gaussian process model (or Gaussian process regression model), which estimated the unknown region of the objective function [22]. In the Bayes optimization method, the acquisition function was the most important part. The acquisition function was a criterion used to select the next sample point to be evaluated, which combined the predicted value of the objective function and uncertainty to find possible improvements in unexplored regions [23]. The common acquisition functions were PI function, UCB function, *etc*., and the expression of PI function was shown in equation (9).

$$a_t(x, d_t) = P(\phi(x) \le c - \theta) = \varphi\left(\frac{c - \theta - u_t(x)}{\sigma_t(x)}\right) \quad (9)$$

where c was the optimal solution of the function. θ was the balancing parameter in the current collection, which was usually used to balance the difference between the local searches made. $\varphi(.)$ was the density function of

the global sample normal distribution. d_t was the relative entropy of the PI function. $u_t(x)$ was the average estimate of the sample before time step t. $\sigma_t(x)$ was the number of times the sample x being selected before time step t. The PI function was a parameter-free method based on the predicted results of the model [24]. It did not depend on the specific form of the model or parameter settings but evaluated the importance of the features by randomly rearranging them. The UCB function, on the other hand, was mostly used in multiple classification tasks, and the mathematical expression of the UCB function was shown in equation (10).

$$\alpha_t\left(x, d_t'\right) = u_t(x) + \tau\sigma_t(x) \tag{10}$$

where τ was the tuning parameter of the classification term, which optimized the classification process by changing the value of confidence in the multinomial classification and thus optimizing the classification process. The XGBoost framework's inability to do global searches led to the use of the PI function as the model's acquisition function in this investigation [25]. However, the PI function also suffered from sampling bias and was not robust to highly noisy objective functions. So, further improvements to the function were needed. This study mainly improved the adaptive ability of the balance parameters of the PI function and avoided the problem of acquisition bias by introducing an adaptive algorithm to adjust and optimize the balance parameters. Equation (11) displayed the revised PI function's expression.

$$PI(x)_p = \varphi\left(\frac{u_t(x) - y_{\max} - \varsigma}{\sigma_t(x)}\right) \tag{11}$$

where ς was the adaptive function. $y_{\max}$ was the maximum value of the objective function of the current input sample. The computational expression of the adaptive function was shown in equation (12).

$$\varsigma = 1 - \frac{1}{e^{y' - y_{\max}}} \tag{12}$$

where y' was the theoretical maximum of the objective function. Equation (13) displayed the function's expression at the point where the adaptive function approached the theoretical maximum.

$$\lim_{(y' - y_{\max}) \to 0} 1 - \frac{1}{e^{y' - y_{\max}}} = 0 \tag{13}$$

where $\lim$ was the seeking limit. $(y' - y_{\max}) \to 0$ was the condition of the limit, i.e., the limit when y' converged to $y_{\max}$. The acquisition function in this state was mainly used for data acquisition and categorization, while the acquisition function was mainly used for exploring the connection between the data when y' was much larger than $y_{\max}$, and the expression for this case was shown in equation (14).

$$\lim_{(y' - y_{\max}) \to \infty} 1 - \frac{1}{e^{y' - y_{\max}}} = 1 \tag{14}$$

After the improvement of the PI function, the XGBoost framework based on the Bayes optimization method of geologic RM optimization was completed. The model was referred to as the Bayes-XGBoost model. The operation flowchart of Bayes-XGBoost model was shown in Figure 5.

Testing of model performance

To analyze the performance of the Bayes-XGBoost model, a Dell G7 series laptop with a CPU of i9-9960X, 16GB of working memory, and an RTX 4090 D graphics card was used with Microsoft's Windows 11 operating system and

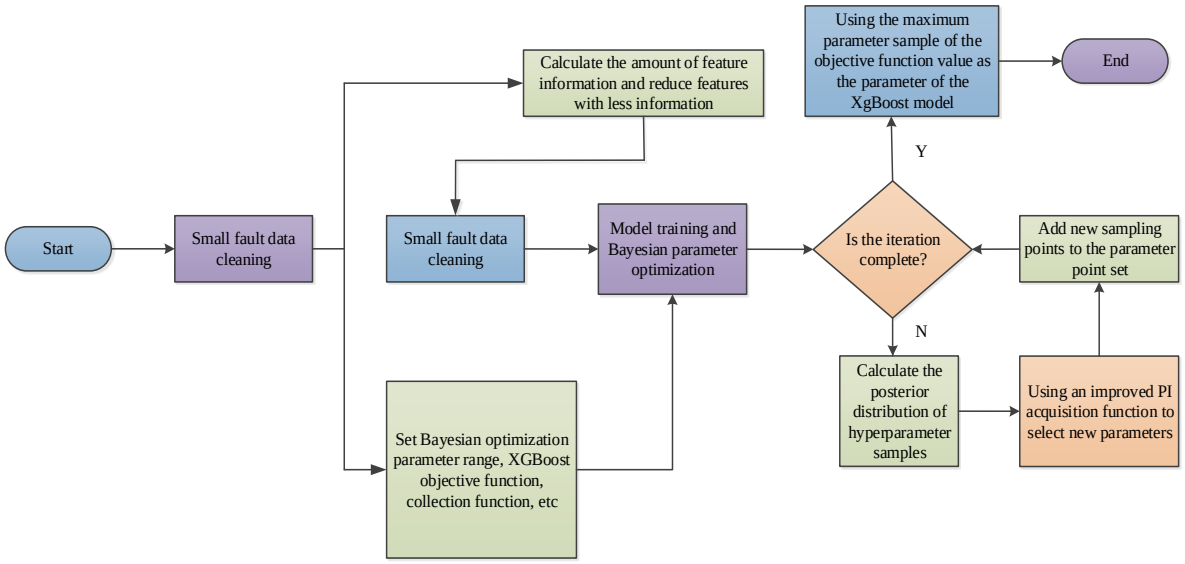

Figure 5. Bayes-XGBoost model running flowchart.

Microsoft Visual C++ (2019). The tests employed the deep learning coal mine image dataset (https://doi.org/10.57760/sciencedb.j00001.00793) and the CDSet dataset (https://pan.baidu.com/s/1KCOQc0Qq9HVSYLc7v2Rigw), a line segment detection dataset, which could be used to simulate the distribution of coal field interruption layers and included five types of underground structural coal in Chinese coal mines from 2020 to 2021 [26]. The reference models including deep coal seam mining area recognition model based on the backpropagation neural network (BPNN) algorithm [27] and coal gangue position and shape recognition model based on the regional convolutional neural network (RCNN) algorithm [28]. Those two popular models used as the controls in this study performed well in terms of recognition accuracy.

Results

Performance comparison of various models

Two datasets, coal mine image dataset and CDNet, were utilized as inputs in this study, respectively, to compare the prediction accuracies of BPNN, RCNN, and Bayes-XGBoost models. The results showed that the Bayes-XGBoost model demonstrated the highest accuracy of 92.3% and 89.6% on the two datasets, respectively, with the average accuracy higher than that of the control models (Figure 6). In addition, the Bayes-XGBoost model outperformed the RCNN and BPNN models by 4.7% and 5.6%, respectively, with an average accuracy of 92.4% on the coal mine image dataset (Figure 7a) and 93.6% on the CDNet dataset (Figure 7b). In contrast, RCNN and BPNN had accuracy rates of 90.2% and 86.1%, respectively.

ROC curves of various models

The receiver operating characteristic (ROC) curve, a popular tool for assessing the effectiveness of classification prediction models, was employed to explore the relationship between true positive rate (TPR) and false positive rate (FPR) under various classification levels. The area under the curve (AUC) value increased with increasing area enclosed by the curve and reference line in the ROC plot, indicating improved model performance. The results showed that the Bayes-XGBoost model had the largest AUC values on both datasets, indicating a better performance of Bayes-XGBoost model than that of BPNN and RCNN models (Figure 8).

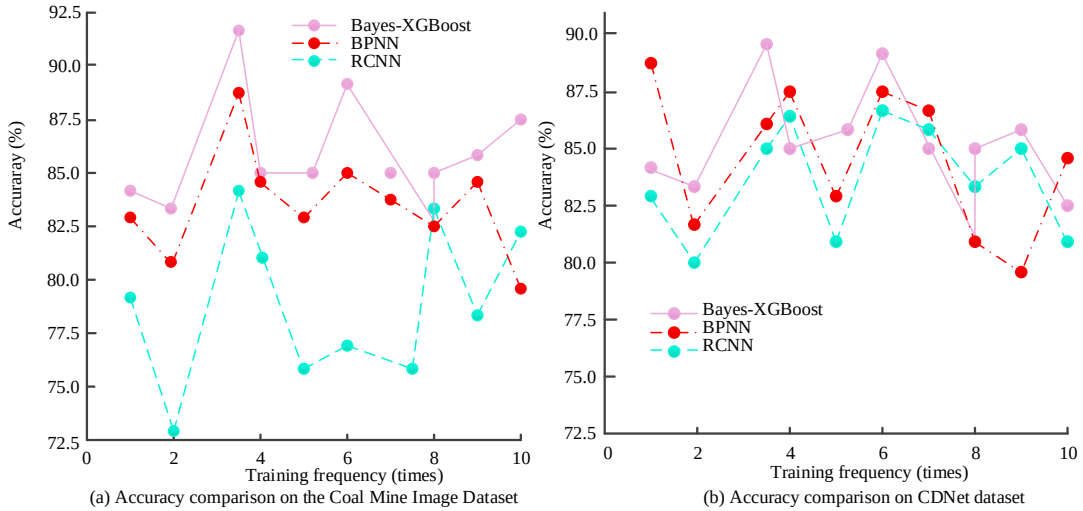

Figure 6. Comparison of accuracy of various models.

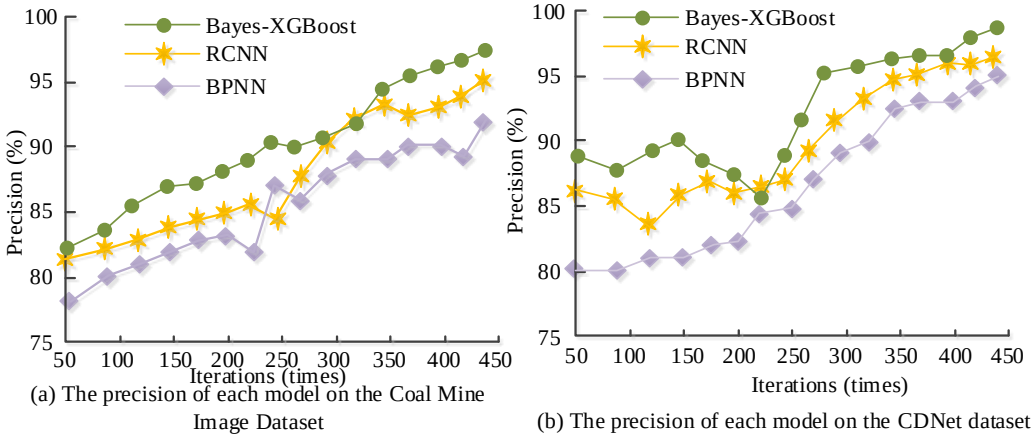

Figure 7. Comparison of precision of various models.

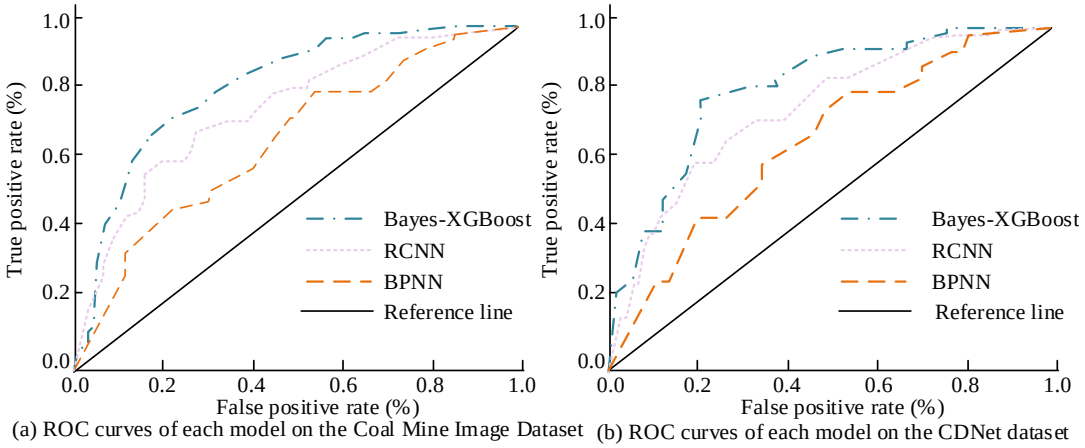

Figure 8. Comparison of ROC curves of various models.

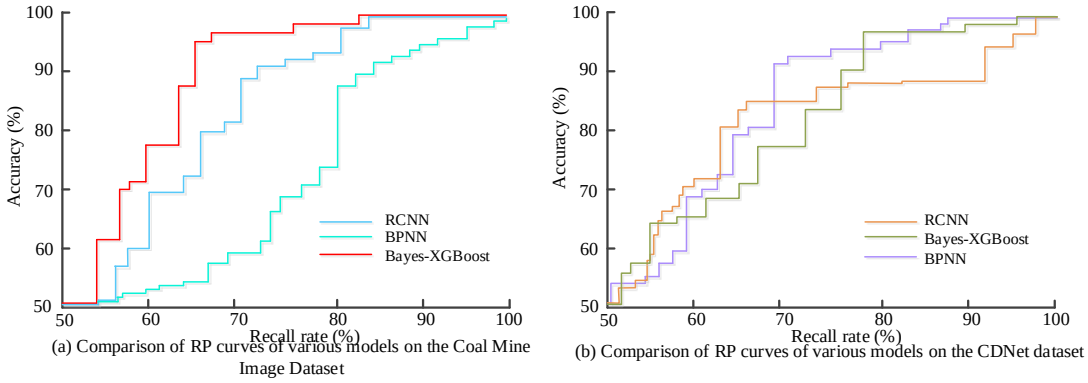

Figure 9. Comparison of RP curves of various models.

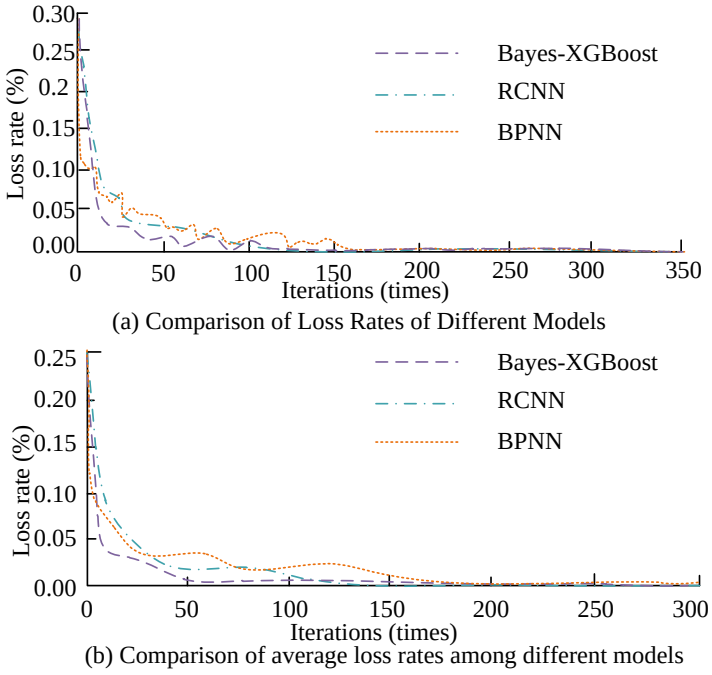

Figure 10. Comparison of loss rate and average loss rate.

Precision-recall curves of various models

The precision-recall curve (PRC) is a common tool for evaluating model performance. The curve is based on the relationship between precision and recall and is plotted by calculating precision and recall at different thresholds. The results demonstrated that the slope of the PRC of the Bayes-XGBoost model was closer to 1, thus the model had a more balanced growth of precision and recall, and the stability of the model was better than other models (Figure 9).

The loss rates of various models

The study analyzed the loss rates of the BPNN, RCNN, and Bayes-XGBoost models *vs.* the number of iterations, as well as the average loss rate after convergence to test the loss rates of the proposed model using the coal mine image dataset as input. The results showed that the Bayes-XGBoost model loss rate decreased the fastest, and the loss rate tended to be close to 0 after about 120 iterations (Figure 10a). After multiple tests, the average loss rate of each

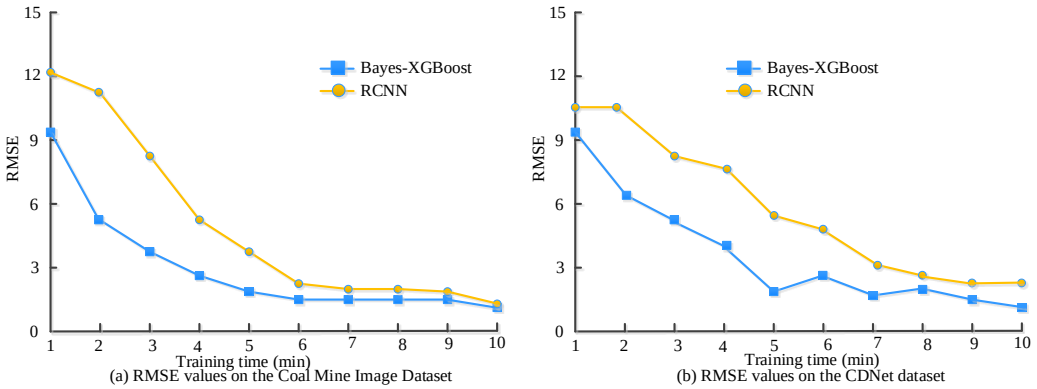

Figure 11. Comparison of RMSE values among different models.

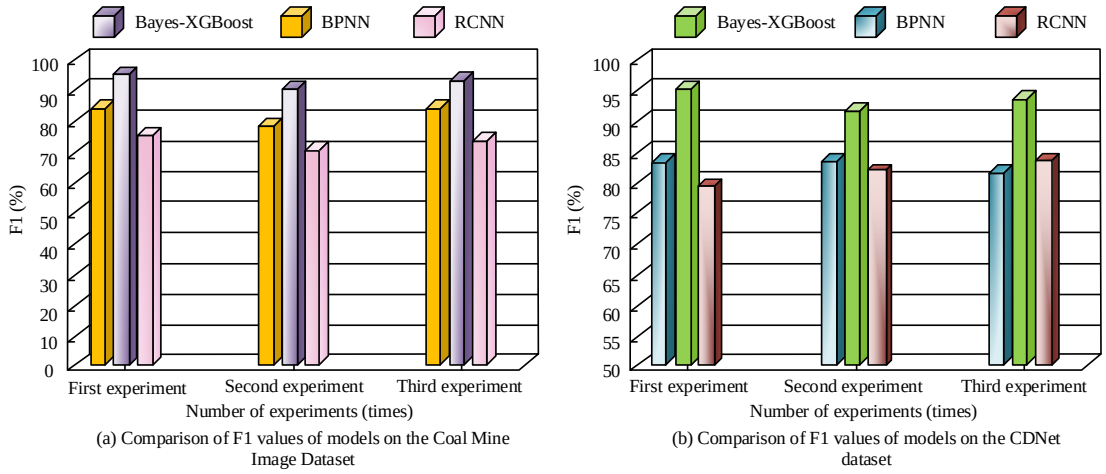

Figure 12. Comparison of F1 values among different models.

model and the trend of the average loss rate change was roughly the same as that demonstrated in Figure 10(a) (Figure 10b). The results indicated that the Bayes-XGBoost model's overall loss rate was lower than that of the control models.

Comparison of Root Mean Square Errors

Regression model performance is often evaluated using the Root Mean Square Error (RMSE) measure, which calculates the mean difference between the values that the model predicts and the actual values that are observed. The study compared the RMSE values of the models using the two datasets as inputs. The Bayes-XGBoost model showed the lowest RMSE value of 1.47, which was 0.02 lower than that of the RCNN model. The evolution of RMSE values of the models on coal mine image dataset regarding the training time was shown in Figure 11a. The trend of RMSE values of the models on CDNet dataset was shown in Figure 11b. The RMSE value of Bayes-XGBoost model on this dataset reached 1.45, which was smaller than that in the control models at any training moment.

The F1 value is a common metric for model evaluation. This study used coal mine image and CDNet datasets as inputs to compare the F1 values of BPNN, RCNN, and Bayes-XGBoost models. The tests were performed three times in the study to eliminate the impact of random variables. The F1 values of the BPNN, RCNN, and

Bayes-XGBoostsang models on coal mine image dataset were 79.2%, 71.7%, and 90.6%, respectively (Figure 12a). On the CDNet dataset, the average F1 value of the Bayes-XGBoostsang model was 93.0%, while the average F1 values of the BPNN and RCNN were 82.7% and 80.8%, respectively (Figure 12b).

Discussion

To test the stability of the proposed model, several control experiments were designed. The results of these experiments demonstrated that the Bayes XGBoost model accurately predicted and recognized faults, which was due to the Bayesian optimization algorithm's ability to classify samples quickly and accurately. Therefore, during dataset sample processing, the model only needed to recognize them by category. The Bayesian optimization algorithm enhanced the feature learning ability of the XGBoost algorithm in samples of the same category, thereby improving the model's overall predictive and recognition performance. Additionally, the XGBoost algorithm exhibited strong anti-interference ability and fault tolerance. When abnormal sample recognition occurred, the XGBoost algorithm would provide feedback and re-identify. This feedback mechanism resulted in a proposed model with a higher number of positive samples than negative samples. As a result, the proposed model outperformed the control models in terms of recall, P-R curve, F1 value, and other indicators. The study also compared the proposed model with PBNN and RCNN models and consulted other materials. Zhang *et al*. proposed a coal gangue loading rate model based on 3D image features and grey wolf optimized support vector machine [29]. Jabeur *et al*. also proposed an XGBoost algorithm and SHAP interaction value prediction model [7]. The model performance tests demonstrated that the proposed model had advantages in computational efficiency and training difficulty. The experimental results revealed that the proposed coal field small fault identification model effectively solved the difficulties encountered in traditional coal field surveys. The Bayes-XGBoost model could be used to obtain the distribution of coal mine faults quickly and accurately below the bottom layer, providing reliable geological data for coal mining. The proposed model had a wide range of applications, including identifying and predicting small faults in coal fields and other geological exploration research. Its proposal promoted the development of the coal mining industry and provided new ideas for geological exploration, hydrological, and water conservancy surveys. Overall, the proposed model had practical significance.

Conclusion

The identification of fault geology is a crucial task in the process of coal field exploration and mining. However, traditional fault identification methods face many challenges in practice, such as variable data quality, accuracy of feature selection, and complexity of model construction. These factors make the accurate identification of small faults a challenging task. Therefore, this study proposed a Bayes-XGBoost model that aimed to promote the development of geo-RM for CMF and provide better safety for the coal mining industry. The results indicated that the proposed model outperformed the control models in terms of recognition accuracy, average precision, F1 value, and other metrics. Therefore, the proposed model was progressive. Some shortcomings of the proposed model were also found in the study. Since the generalization ability of the Bayes-XGBoost model depended on the quality and diversity of the training data, the generalization ability of the model could be improved when the training data types were not duplicated. The proposed model in this study was expected to solve the shortcomings of insufficient generalization ability by other optimization methods, which would bring new ideas for the development of geo-RM for CMF.

References

1. Hou C, Jiang B, Yang Y, Li M, Song Y. 2022. Characteristics of fracture and microdeformation in coal seam distributed around the small-scale strike reverse fault. Energy Fuels. 36(13):6868-6880.
2. Hoque R, Das S, Hoque M, Haque E. 2024. Breast cancer classification using XGBoost. J Adv Res. 21(2):1985-1994.
3. Feng S, Chen X, Dong X, Wang L, Li G. 2023. Study on the effect of small faults on the gas content in No.3 coal seam of the changping mine field. ACS Omega. 8(19):16800-16808.
4. Wang D, Cheng Y, Yuan L, Wang C, Wang L. 2023. Implications of geological conditions on gas contents: A case study in the Pingdingshan coalfield. Energy Fuels. 37(9):6465-6478.
5. Jia T, Yan J, Liu X, Feng Z, Wei G, Cao L. 2022. Analysis method of the occurrence law of coalbed gas based on gas-geology units: a case study of the Guhanshan Mine Field, Jiaozuo Coalfield, China. ACS Omega. 7(14):12296-12306.
6. Asselman A, Khaldi M, Aammou S. 2023. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Envir. 2023. 31(6):3360-3379.
7. Jabeur SB, Mefteh-Wali S, Viviani JL. 2024. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Oper Res. 334(1):679-699.
8. [8] Osman A I A, Ahmed A N, Chow M F, Huang Y F, El-Shafie A. 2021. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Eng J. 12(2):1545-1556.
9. [9] Thongsuwan S, Jaiyen S, Padcharoen A, Agarwal P. 2021. ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost. Nucl Eng Technol. 53(2):522-531.
10. Zhu X, Chu J, Wang K, Wu S, Yan W, Chiam K. 2021. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. J Rock Mech Geotech. 13(6):1231-1245.
11. Wickramasinghe I, Kalutarage H. 2021. Naive Bayes: Applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Methodol. Appl. 25(3):2277-2293.
12. Liu S, Kuppens P, Bringmann L. 2021. On the use of empirical Bayes estimates as measures of individual traits. Assess. 28(3):845-857.
13. Jospin LV, Laga H, Boussaid F, Buntine W, Bennamoun M. 2022. Hands-on Bayesian neural networks—A tutorial for deep learning users. IEEE Comput Intell Mag. 17(2):29-48.
14. Sun B, Zhou Y. 2022. Bayesian network structure learning with improved genetic algorithm. Int J Intell Syst. 37(9):6023-6047.
15. Laumont R, Bortoli VD, Almansa A, Delon J, Durmus A, Pereyra M. 2022. Bayesian imaging using plug & play priors: When langevin meets tweedie. Siam J Imaging Sci. 15(2):701-737.
16. Jabeur SB, Stef N, Carmona P. 2023. Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. Comput Econ. 61(2):715-741.
17. Piraei R, Afzali SH, Niazkar M. 2023. Assessment of XGBoost to estimate total sediment loads in rivers. Water Resour Manag. 37(13):5289-5306.
18. Zhang J, Wang R, Lu Y, Huang J. 2024. Prediction of compressive strength of geopolymer concrete landscape design: application of the novel hybrid RF–GWO–XGBoost algorithm. Buildings-Basel. 14(3):591.
19. Ma B, Yan G, Chai B, Hou X. 2022. XGBLC: An improved survival prediction model based on XGBoost. Bioinformatics. 38(2):410-418.
20. Zhang J, Chen L, Hou X, Ren X, Li J, Chen Y. 2022. Hydrogeochemical processes of carboniferous limestone groundwater in the Yangzhuang coal mine, Huaibei coalfield, China. Mine Water Environ. 41(2):504-517.
21. Mokayed H, Quan TZ, Alkhaled L, Sivakumar V. 2023. Real-time human detection and counting system using deep learning computer vision techniques. Artif Intell Appl. 1(4):221-229.
22. Hebbi C, Mamatha H. 2023. Comprehensive dataset building and recognition of isolated handwritten Kannada characters using machine learning models. Artif Intell Appl. 1(3):179-190.
23. Wang X, Hou R, Xia Y, Zhou X. 2021. Structural damage detection based on variational Bayesian inference and delayed rejection adaptive Metropolis algorithm. SHM. 20(4):1518-1535.
24. Xie W, Nie W, Saffari P, Robledo LF, Descote PY, Jian W. 2021. Landslide hazard assessment based on Bayesian optimization–support vector machine in Nanping City, China. Nat Hazards. 109(1):931-948.
25. Frey D, Shin JH, Musco C, Modestino MA. 2022. Chemically-informed data-driven optimization (ChIDDO): Leveraging physical models and Bayesian learning to accelerate chemical research. React Chem Eng. 7(4):855-865.
26. Kang H, Gao F, Xu G, Ren H. 2023. Mechanical behaviors of coal measures and ground control technologies for China's deep coal mines–A review. J Rock Mech Geotech. 15(1):37-65.
27. Zhang J, Hou X, Liu S, Chen L, Wang Y. 2023. New data-driven method for *in situ* coalbed methane content evolution: A BP neural network prediction model optimized by grey relation theory and particle swarm. Energy Fuels. 37(14):10344-10354.
28. Yan PC, Zhang H, Kan XY, Chen FX, Wang CX, Liu ZC. 2024. Lightweight detection method of coal gangue based on multispectral and improved YOLOv5s. Int J Coal Prep Util. 44(4):399-414.
29. Zhang C, Dou D, Li X. 2024. Measurement of coal-carrying rate in gangue based on three-dimensional image features and gray wolf optimization-support vector machine. Int J Coal Prep Util. 44(3):275-290.