RESEARCH ARTICLE

Lithology identification of uranium drilling based on SMOTE algorithm logic-CNN

Shengming Shi*

The Second Geological and Mineral Exploration Institute, Gansu Provincial Bureau of Geology and Mineral Exploration and Development, Lanzhou, Gansu, China.

Received: May 6, 2024; **accepted:** July 6, 2024.

With the development of society and the continuous increase in energy consumption, timely exploration of mineral resources is very necessary. Lithology identification is an important way to understand the structure of underground rock layers, but current lithology identification techniques face problems such as imbalanced logging data and difficulty in predicting small proportions of lithology. To improve the accuracy of lithology identification, this research proposed a method from two perspectives including data processing and model optimization. A synthetic minority oversampling technique (SMOTE) was designed to address the issue of imbalanced logging data to improve the recognition rate of minority lithology, while a hybrid algorithm combining particle swarm optimization algorithm and gradient boosting decision tree (GBDT) was designed and the hyper-parameters of the GBDT were optimized through particle swarm optimization algorithm. Further, a lithology recognition model based on logical three-dimensional convolutional neural network (logic-CNN) algorithm was constructed to investigate the lateral correlation between lithology and geophysical data. The results showed that, after using the SMOTE, the recognition rates of mudstone, argillaceous siltstone, siltstone, volcanic rock, and tuff increased by 1.30%, 3.03%, 1.90%, 1.03%, and 12.09%, respectively. The training and testing time of a hybrid algorithm combining particle swarm optimization algorithm and GBDT was 43 s and 279 ms, respectively. The lithology recognition model based on logic-CNN algorithm had a prediction accuracy of over 75% in drilling lithology. The designed algorithms and models demonstrated good performance and could provide technical support for lithology identification in uranium mining drilling.

Keywords: SMOTE; logic-CNN; lithology; identification; GBDT.

***Corresponding author:** Shengming Shi, The Second Geological and Mineral Exploration Institute, Gansu Provincial Bureau of Geology and Mineral Exploration and Development, Lanzhou 730020, Gansu, China. Email: shishengming2023@163.com.

Introduction

With the development of human civilization, the consumption of energy by humans is also continuously increasing. Uranium ore is mainly used in the nuclear industry and has significant strategic significance [1]. Well logging data is one of the attributes of information in mineral exploration with high resolution, which can reflect the changes in geophysical properties in the vertical direction. Lithological identification is one of the key methods in logging technology and an important way to understand the structure of underground rock layers [2]. At present, common methods for rock classification and recognition include intersection plot method, probability and statistical method, cluster analysis method, and machine learning method. The intersection

diagram method is the earliest proposed method, which has the advantage of simple use, but also has the problems of large errors, high time and labor costs. Probability and statistical methods are technically complex and require a significant amount of information [3, 4]. With the rapid development of computer technology and machine learning, the support vector machine (SVM) and neural networks have gradually been applied to lithology recognition.

As the significance of sandstone-type uranium resources continues to be recognized, many scientists are engaged in research on the development of lithology identification technology. Bajwa *et al*. selected color composite images under different band combinations to explore the drawing effect of land satellite thematic imaging instruments on rock lithology and used Erdas Image software to process the images. By changing the color tone of the image, different rock formations could be identified, and different bands were suitable for identifying rock lithology in different periods [5]. To improve the efficiency of uranium production blocks, a previous study conducted theoretical information analysis on the process flow of uranium production, determined the regression relationship between the main indicators of the production process cycle, and calculated the core geological and technical indicators, which identified problem areas and effectively reduced development costs and improved efficiency [6]. Chen *et al*. designed an algorithm that combined laser-induced breakdown spectroscopy and 2D deep convolutional neural network (CNN) to identify the lithology and main elements of rocks, using two different outputs to complete classification and regression tasks simultaneously and resulted a high accuracy rate of over 99% [7]. Yan *et al*. designed a recognition model grounded on CNN to lift survey efficiency and constructed it into a specialized recognition system. This model could not only identify lithology, but also identify the degree of weathering and showed good performance with recognition accuracy of over 95% and 91% for weathering degree and lithology, respectively [8]. Further, Hossain *et al*. designed a lithology prediction technique based on rough set theory to predict the complex lithology of the formation by constructing a logical rule for data from multiple logging data. Rock debris data was used to verify the performance of this prediction technique, and it was found that the misclassification rate of this method was low to about 18%, which could effectively predict lithology [9]. To analyze reservoirs in complex geology, Hemiram *et al*. designed wavelet transform to decompose the frequency spectrum and determined the characterization wavelet of reservoir lithology. A few advanced attribute analysis processes could use spectral decomposition volume. Gaussian wavelets had better performance, and the images generated by attribute analysis could reflect the range of reservoir lithology [10]. To explore the influence of deep geological conditions on the mechanical properties of shale, scientists designed real-time high temperature triaxial compression experiments by constructing a thermal coupling model and utilizing micro mechanical parameters to construct a numerical fracturing mechanism model to verify that the influence of confining pressure on rock mechanics was greater than that of temperature [11]. Sun *et al*. proposed two special conditions and established an inverted L-shaped model to understand the evolution of permeability during rock shear displacement, which involved matrix and cracks and could analyze the evolution of permeability and internal morphological changes in cracks. This model could analyze the evolution of permeability well, and both the initial crack size and the surface roughness of the crack would have an impact on the evolution of permeability [12].

There is currently a wealth of research on uranium ore lithology identification technology, and the methods involved are also diverse. However, the current lithology identification models suffer from issues such as imbalanced logging data, difficulty in predicting small-scale lithology, and low optimization efficiency [13]. To improve the accuracy of lithology identification, this study focused on two perspectives including

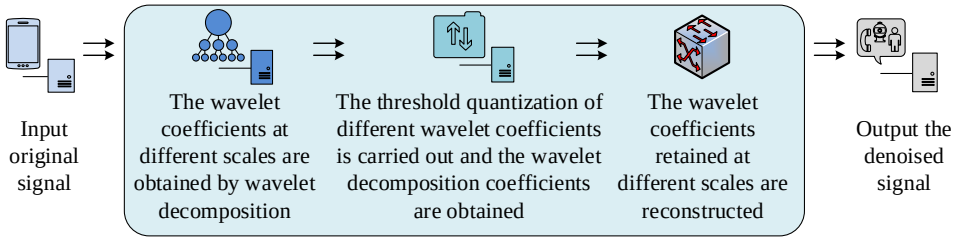


Figure 1. The denoising process of NWTTM.

data processing and model optimization. The objectives of this study were to improve the recognition rate of minority rock types, improve the optimization efficiency of recognition models, provide information technology support uranium mining, and improve the efficiency of uranium mining. This study designed the synthetic minority oversampling technique (SMOTE) algorithm and constructed a combined particle swarm optimization (PSO) - gradient boosting decision tree (GBDT) algorithm. Then, a lithology recognition model was built based on the logical three-dimensional convolutional neural network (logic-CNN) algorithm. The results of this study would enrich the research achievements in the field of lithology recognition and provide guidance for subsequent research.

Materials and methods

Design of SMOTE algorithm for imbalanced logging data

In response to the issue of imbalanced logging data, this study employed SMOTE to achieve data balance and preprocesses the logging data. The correlation function method was taken to perform depth correction on logging curves, and the nonlinear wavelet transform threshold method (NWTTM) was used to denoise logging data. The minimum and maximum normalization method (MMNM) was then applied to standardize the logging data. The correlation coefficient between the shift point b of the logging curve and the benchmark logging curve was calculated by equation (1).

$$A(b)=\frac{\sum_{g=c}^{c+d-1}\left(h_g-\bar{h}\right)\left(j_{g+b}-\bar{j}\right)}{\sqrt{\sum_{g=c}^{c+d-1}\left(h_g-\bar{h}\right)^2\sum_{g=c}^{c+d-1}\left(j_{g+b}-\bar{j}\right)^2}},b\in L\,\&\,b\in[-M,+M] \quad (1)$$

where h was the baseline logging curve. j was the comparison logging curve. c was the starting comparison position on h. d was the quantity of sampling points for the comparison length segment within h. h_g was the g-th sampling value of h. $\bar{h}$ was the average value of h in the comparison length segment. b was the number of points j moves relative to h. j_{g+1} was the $g+b$-th sampling value of j. $\bar{j}$ was the average value of j in the comparison length segment. L was the maximum amount of points for h. M was the maximum points that j could move in one direction. NWTTM not only could process Gaussian white noise in logging data, but also had significant advantages in processing time. The denoising process of this method was illustrated in Figure 1 [14], where the first step of NWTTM denoising was to input the original signal, and the second step was to perform wavelet decomposition and obtain wavelet coefficients at different scales. The third step was to perform threshold quantization on different wavelet coefficients and obtain wavelet decomposition coefficients, while the fourth step was to reconstruct the wavelet coefficients preserved at different scales, and the fifth step was to output the denoised signal. The normalization process of MMNM was shown in equation (2) [15].

$$N_{op}=\frac{n_{op}-n_{o\min}}{n_{o\max}-n_{o\min}},p=1,2,\ldots,\Pi \quad (2)$$

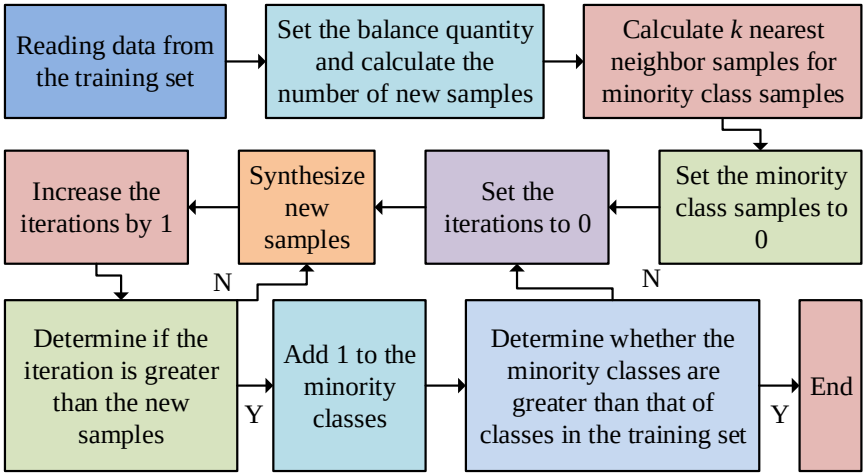


Figure 2. The main process of SMOTE algorithm.

where n_{op} was the value of the p-th sample under the o -th logging curve. N_{op} was the corresponding standardized data. $n_{o\min}$ was the minimum value. $n_{o\max}$ was the maximum value. Π was the quantity of samples in the dataset. The purpose of classification models was generally to reduce the loss function of the entire dataset, which could easily lead to the misclassification of minority rock samples into majority classes, affecting the accurate identification of minority rock samples. Therefore, this study adopted SMOTE algorithm to balance the preprocessed logging data. SMOTE had the advantages of diverse synthetic samples and high data utilization, effectively handling the matter of data imbalance in multi-classification problems [16]. The main process of SMOTE included the steps of (1) reading the data from the training set, (2) setting the balance number and calculating the new samples, (3) calculating the A nearest neighbor samples of minority class samples, (4) setting the minority class samples to 0, (5) setting the number of iterations to 0, (6) synthesizing a new sample, (7) increasing the iterations by 1, and (8) determining whether the iterations were greater than the new samples (Figure 2). If it was determined to be true, 1 was added to the minority classes, otherwise, the process went back to step 6. The final step 9 was to determine whether the minority classes were greater than that in the training. If yes, the process ended, otherwise, the process returned to step 5. The calculation of the new sample was shown in equation (3).

$$R_i \approx \frac{R_{balance}}{R^i_{minority}} - 1, i = 1, 2, ..., \Upsilon \tag{3}$$

where $R_{balance}$ was the number of balanced samples. $R^i_{minority}$ was the number of i -th minority class samples. Υ was the number of minority classes. The calculation of the new sample data was as below.

$$s_{new} = s_{i\partial} + \text{rand}(0,1) \times \left(s_\beta - s_{i\partial} \right) \tag{4}$$

where $s_{i\partial}$ was a minority class sample. s_β was a randomly selected sample from $s_{i\partial}$'s k nearest neighbor samples. $\text{rand}(0,1)$ was the generation of a random number within (0,1).

Validation of SMOTE algorithm

To validate the performance of SMOTE, this study analyzed its performance in minority lithology recognition and overall lithology recognition and preprocessed the data before the experiment. The data used in the study was obtained from well logging data of sandstone

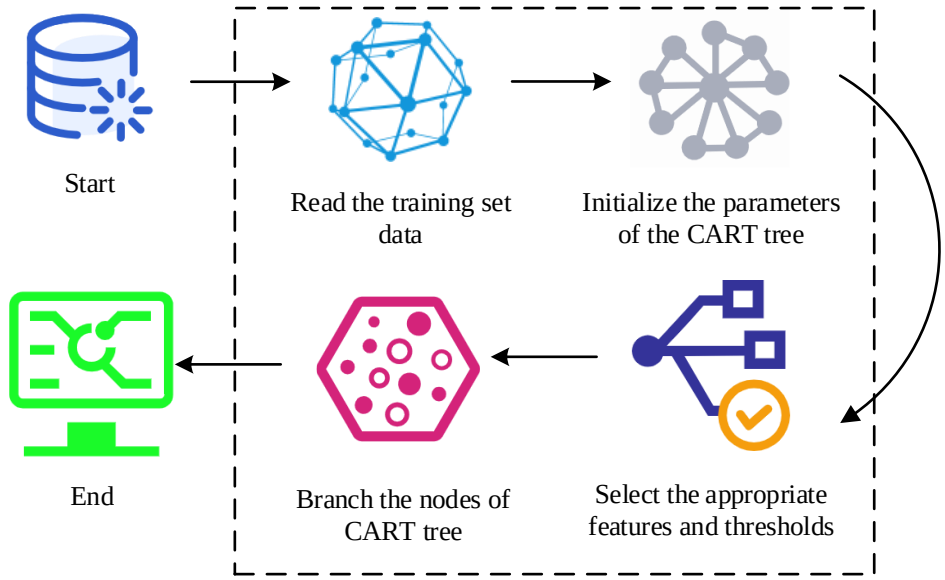


Figure 3. The training process of CART.

type uranium deposits in a uranium mining area at the southern end of Daqing Placanticline in the Songliao Basin, Heilongjiang province, China. The data was subjected to secondary survey screening and original drilling verification by the Tianjin Geological Survey Center of the China Geological Survey (Tianjin, China). An one-week drilling data from 6 logging curves and 6 logging sample points in 2018 were selected for this study. The preprocessed logging data was randomly divided into 20% for model training and the remaining 80% for model testing. The operating system used in the experiment was Windows 11 with an Intel Core i5 12600K processor, a main frequency of 3.7 GHz, and a maximum memory of 128 GB. For the identification and analysis of minority rock types, this study constructed GBDT classification models before and after using the SMOTE algorithm and analyzed the results using a confusion matrix.

Design of PSO-GBDT algorithm for improving optimization efficiency

To improve the accuracy of lithology identification, a PSO-GBDT algorithm combining PSO and GBDT was constructed. The GBDT algorithm that uses classification and regression tree (CART) as a weak learning machine has the advantages of short prediction time and high recognition accuracy, which can effectively solve multi-classification problems [17, 18]. The training process of CART included steps of reading the training set data, initializing the parameters of the CART tree, selecting appropriate features and thresholds, and branching the nodes of the CART (Figure 3). Due to the significant impact of hyper-parameter selection on the classification results of machine learning algorithms, it was necessary to make optimal selection that could be seen as multi-objective optimization. PSO can reduce the complexity of problems and is widely used for solving multi-objective optimization problems. Therefore, this study adopted the PSO-GBDT algorithm for hyper-parameter optimization and identifies lithology. The process of PSO-GBDT included the steps as (1) defining the position of particles in the PSO algorithm, (2) setting the particle swarm size, maximum number of iterations, and minimum particle fitness, (3) determining the fitness function, (4) setting the particle activity interval, (5) initializing the particle swarm position vector and velocity vector, (6) calculating the current fitness of the particles and initializing the local optimal solution

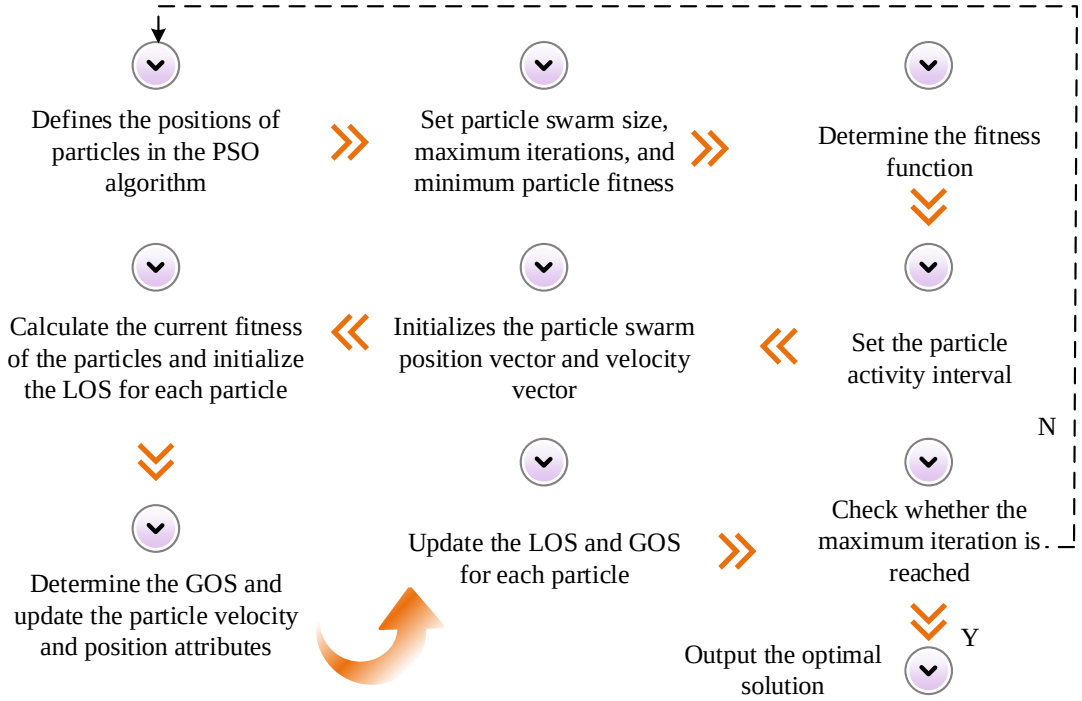


Figure 4. The process of PSO-GBDT algorithm.

(LOS) for each particle, (7) determining the global optimal solution (GOS) and updating the particle velocity and position attributes, (8) updating the LOS and GOS for each particle, (9) determining whether the maximum iterations had been completed. If yes, then output the optimal solution, otherwise, returning to step 1 (Figure 4). The representation of particle x was shown in equation (5).

$$x = \begin{bmatrix} \text{learning rate} \\ \text{number of decision trees} \\ \text{subsample ratio of columns} \\ \text{maximum depth} \end{bmatrix} \quad (5)$$

where "learning rate" was the learning rate of the algorithm. "number of decision trees" was the number of CART generated. "subsample ratio of columns" and "maximum depth" were the proportion of samples required to train a single CART and the growth depth of a single CART, respectively. The calculation of the fitness function was shown in equation (6) [19].

$$f(x) = \mathrm{GBDT}_{solve}(x) \quad (6)$$

where $\mathrm{GBDT}_{solve}(x)$ was the recognition error of the GBDT model on the test set when the hyper-parameter value was the particle value x. The calculation of the LOS for each particle was as follows.

$$\mathrm{pbest}_t = f(x_t) \quad (7)$$

where x_t was the particle swarm position vector. t was the particle number. The calculation of the GOS was as below.

$$\mathrm{gbest} = \min(\mathrm{pbest}) \quad (8)$$

where pbest was the LOS of all particles. The updates of particle velocity and position attributes were shown in equation (9).

$$\begin{cases} v_t' = v_t + \alpha_1 \times \mathrm{rand}(0,1) \times (\mathrm{pbest}_t - x_t) + \alpha_2 \times \mathrm{rand}(0,1) \times (\mathrm{gbest} - x_t) \\ x_t' = x_t + v_t \end{cases} \quad (9)$$

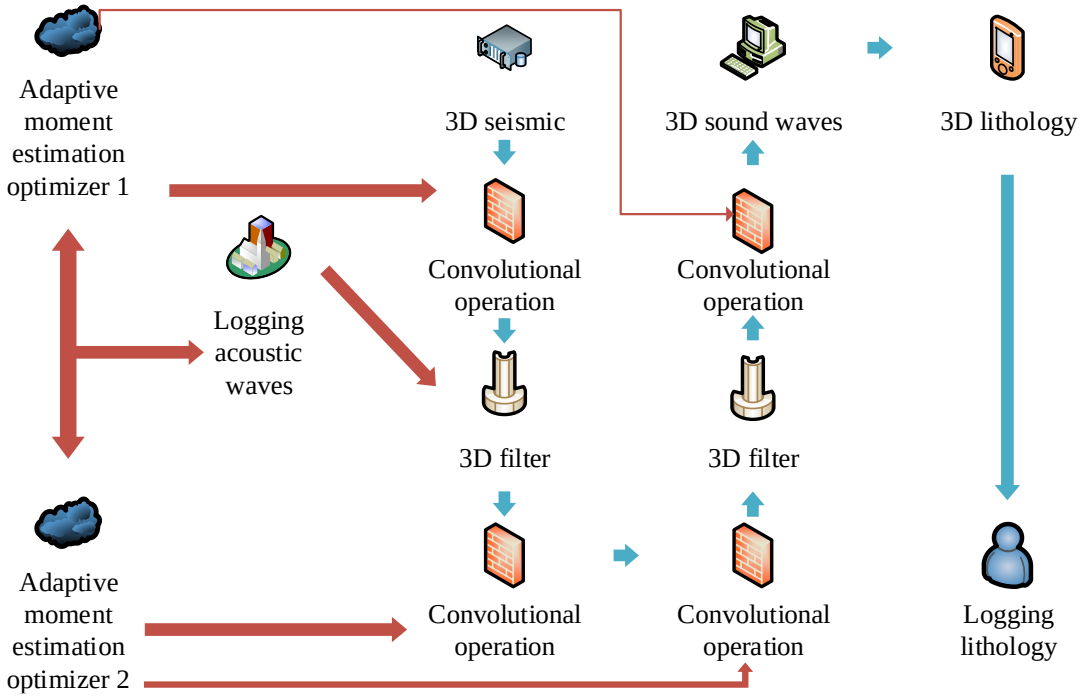


Figure 5. Structural diagram of Logics-CNN model.

where v_t was the velocity vector. α_1 and α_2 were both learning factors. The updates of local and GOSs for each particle were then shown in equation (10).

$$\begin{cases} \text{pbest}_t' = f(x_t), if \quad f(x_t) < \text{pbest}_t \\ \text{gbest}' = \min(\text{pbest}), if \quad \text{gbest} > \min(\text{pbest}) \end{cases} \quad (10)$$

where pbest_t' and gbest' were the updated LOS and GOS for each particle, respectively.

Validation of PSO-GBDT algorithm

To verify the performance of PSO-GBDT, this study analyzed its training and testing time, as well as the accuracy of the algorithm in lithology recognition. The selected comparison algorithms include SVM [20], probabilistic neural network (PNN) [21], random forest (RF) [22], and adaptive gradient boosting (AdaBoost) [23].

Design of logic-CNN algorithm for 3D lithology recognition

To further improve the accuracy of lithology identification, a logic-CNN model was designed. The structure of the logic-CNN lithology recognition model mainly included 3D seismic, 3D filter, 3D acoustic, 3D lithology, logging lithology, logging acoustic, convolution operation, weight coefficient, bias coefficient, and adaptive moment estimation optimizer (Figure 5). The logic-CNN model not only combined artificial interpretation of lithology logic and CNN, but also integrated geophysical data such as seismic and logging data. The reservoir lithology exhibited 3D spatial heterogeneity characteristics. Through 3D seismic exploration and corresponding information, it could provide good information support for 3D imaging of lithology. As one of the representative algorithms of machine learning, CNN has the ability of representation learning. The main content of CNN algorithm includes three points including local connections, weight sharing, and pooling. The nonlinear output in CNN is mainly achieved through activation functions. The nonlinear relationship between input features and output targets was in equation (11) [24].

$$\chi = \eta\left(W\varphi + U\right) \tag{11}$$

where φ was the input matrix and χ was the output matrix. η was the nonlinear activation function. W was the weight filter, also known as the weight coefficients. U was the bias coefficient. By using the size of the loss function and the random gradient descent method, the weights in the neural network model could be automatically adjusted and used for predicting the 3D lithology distribution. The definition of the loss function was shown in equation (12).

$$\mathrm{LOSS} = -\sum_{\theta=1}^{G} \hat{\delta}_\theta \log \delta_\theta + \left(1-\hat{\delta}_\theta\right)\log\left(1-\delta_\theta\right) \tag{12}$$

where G was the amount of output data. θ was the serial number of the output data. δ_θ was the measured data. $\hat{\delta}_\theta$ was the model output data. The error calculation between the generated value and the true value of acoustic data was as below.

$$\mathrm{LOSS}' = \frac{1}{\varpi + S}\sum_{\mathrm{D},\Psi=1}^{\varpi,S}\left(\delta_{\mathrm{D},\Psi} - \hat{\delta}_{\mathrm{D},\Psi}\right)^2 \tag{13}$$

where ϖ was the number of boreholes with acoustic logging. S was the recorded data of acoustic time difference along each borehole. D was the serial number of the logging. Ψ was the serial number of the depth segment. $\delta_{\mathrm{D},\Psi}$ and $\hat{\delta}_{\mathrm{D},\Psi}$ were the predicted and actual logging acoustic data for the D-th well and the Ψ-th depth segment, respectively. To minimize the error, it was necessary to update the weights and biases of the neural network as shown in equation (14).

$$\left(W,U\right)^{\varepsilon} = \left(W,U\right)^{\varepsilon-1} - \phi \times \frac{\sigma_\varepsilon / \left(1-\psi_1^{\varepsilon}\right)}{\sqrt{\Omega_\varepsilon / \left(1-\psi_2^{\varepsilon}\right)} + \xi} \tag{14}$$

where ε was the training duration. ϕ was the learning rate. σ_ε and Ω_ε were the weight and bias coefficient updated at the first moment. ψ was exponential decay. ξ was constant. By using 3D acoustic data and convolution operations, it was possible to ultimately generate a 3D lithology distribution. The error between the final generated 3D lithology distribution and the real lithology data was represented by equation (15).

$$\mathrm{LOSS}'' = -\frac{1}{\varpi + S}\sum_{\mathrm{D},\Psi=1}^{\varpi,S}\left[\delta_{\mathrm{D},\Psi}\log\left(\delta_{\mathrm{D},\Psi}\right) + \left(1-\delta_{\mathrm{D},\Psi}\right)\log\left(1-\delta_{\mathrm{D},\Psi}\right)\right] \tag{15}$$

The hyper-parameters of 3D CNN structures involved network depth, network width, and the size of 3D filters. The designed 3D CNN had a network depth of 4, a width of 21, a filter size of 6 × 6 × 6, and a number of 3 filters. To verify the performance of the logic-CNN algorithm, this study analyzed the loss function, lithology prediction results, and accuracy of the model. The selected comparison algorithms include RF, eXtreme gradient boosting (XGBoost) [25], and AdaBoost.

Results and discussion

Performance verification of SMOTE algorithm

Before using SMOTE algorithm, the recognition rates of the seven minority rock types were 92.21%, 92.52%, 91.76%, 93.76%, 97.87%, 100%, and 84.68%, respectively (Figure 6a). After using SMOTE to balance the logging data, the classification model achieved recognition rates of 91.89%, 93.82%, 94.79%, 95.66%, 98.90%, 100%, and 96.77% on seven minority lithology types, respectively (Figure 6b). The results showed that the recognition rate of 5 minority rock types had been improved to a certain extent after using the SMOTE algorithm. Among them, tuff had the best improvement effect with an improvement ratio of 12.09%. The results confirmed that SMOTE could improve the recognition performance of classification models on minority rock types. To analyze the recognition effect of SMOTE on the overall lithology, the corresponding GBDT classification models were constructed and compared with other data balance algorithms including random over sampling (ROS), SVM-

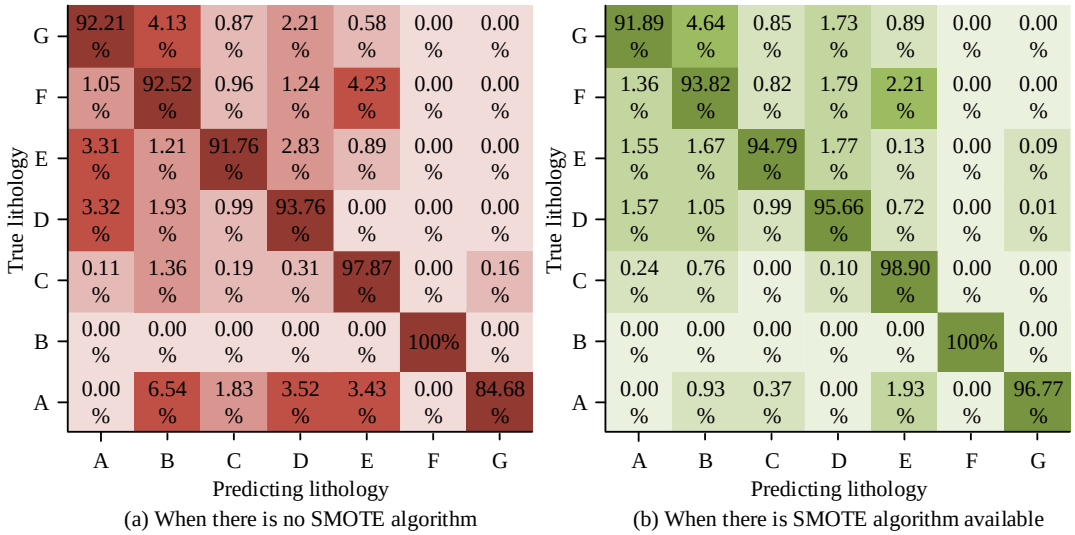


Figure 6. Comparison of confusion matrices in the recognition results of classification models before and after using the SMOTE. A: sandstone. B: mudstone. C: argillaceous siltstone. D: siltstone. E: volcanic rock. F: granodiorite. G: tuff.

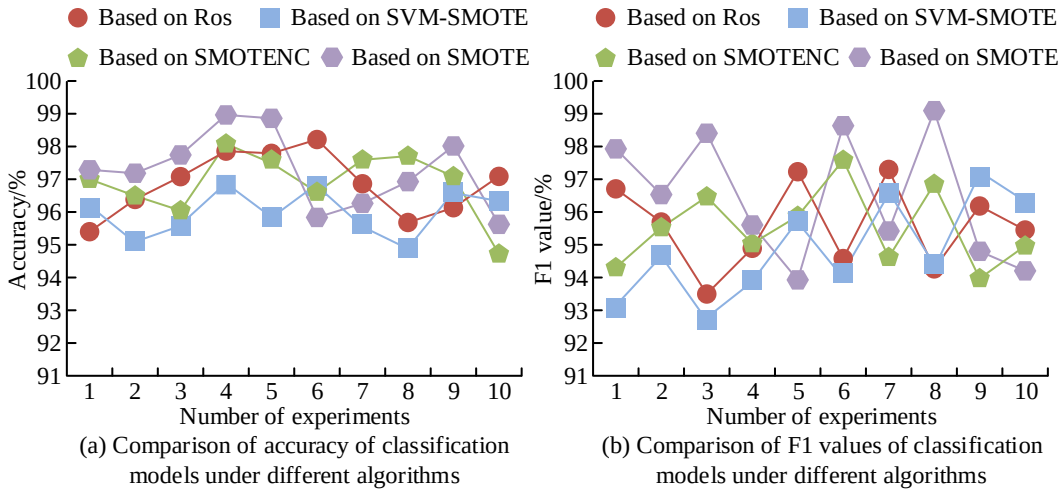
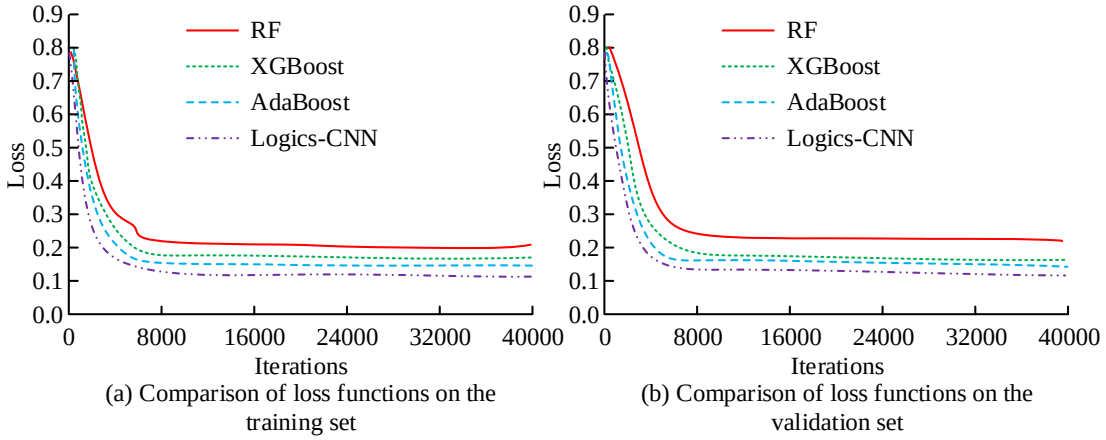


Figure 7. Comparison of accuracy and F1 value of classification models under different data balancing algorithms.

SMOTE combined with SVM and SMOTE, and SMOTENC [26]. After repeating the analysis 10 times, the accuracy and F1 value of classification models were compared under different data balancing algorithms. The results showed that the average classification accuracies based on ROS, SVM-SMOTE, SMOTENC, and SMOTE models were 96.25%, 95.57%, 96.27%, and 96.44%, respectively (Figure 7a). The average F1 values of the four models were 95.96%, 94.81%, 95.81%, and 96.11%, respectively (Figure 7b). The performance of classification model based on SMOTE was significantly better than that of other models in terms of accuracy and F1 value, which indicated that, after SMOTE processing, the model had a better recognition effect on the overall lithology.

Performance verification of PSO-GBDT algorithm

The optimal value for the hyper-parameter of the maximum number of decision trees generated in



(a) Comparison of loss functions on the training set

(b) Comparison of loss functions on the validation set

Figure 8. Comparison of loss functions for different algorithms.

PSO-GBDT was 580, and the maximum growth depth of the decision tree was 7. The learning rate was 0.02, and the required sample ratio for training a single decision tree was 0.9. Different algorithms were also used to compare the root mean square error (RMSE) and recognition accuracy. The results showed that the maximum RMSE values of SVM, PNN, RF, AdaBoost, and PSO-GBDT algorithms were 3.17%, 2.87%, 2.58%, 2.29%, and 1.89%, respectively, while the minimum values were 2.86%, 2.54%, 2.35%, 1.98%, and 1.57%, respectively. The maximum accuracy values of the five algorithms for lithology recognition were 91.37%, 93.46%, 95.87%, 97.17%, and 98.95%, respectively, while the minimum values were 88.51%, 92.75%, 94.26%, 96.88%, and 97.01%, respectively. Therefore, the PSO-GBDT algorithm was significantly better than the comparison algorithms in terms of RMSE and recognition accuracy, indicating its better performance and recognition effect.

Performance verification of logic-CNN algorithm

The lithology of the study area mainly included mudstone and sandstone, and there were 10 boreholes. The comparison of loss functions for different algorithms demonstrated that, on the training set, RF, XGBoost, AdaBoost, and logic-CNN algorithms all underwent 40,000 iterations, and after nearly 5,900, 5,600, 5,400, and 5,000 iterations, they tended to flatten out. The minimum values of the loss functions for the four algorithms were 0.20, 0.18, 0.14, and 0.11, respectively (Figure 8a). On the validation set, the RF, XGBoost, AdaBoost, and logic-CNN algorithms iterated nearly 6,000, 5,750, 5,500, and 5,100 times, respectively, and then tended to flatten out. The minimum values of the loss functions for the four algorithms were 0.21, 0.17, 0.15, and 0.12, respectively (Figure 8b). The results confirmed that the logic-CNN performed better than other algorithms. The prediction performance of the lithology recognition model based on the logic-CNN showed that, among the 10 drilling wells, the predicted lithology of sandstone and mudstone was basically consistent with the actual lithology (Figure 9). Among them, the predicted lithology and actual lithology of drilling with serial numbers A, B, and D were completely consistent, while the differences between the predicted lithology and actual lithology of other drilling were also relatively small. Overall, the prediction accuracy of all drilling lithology was above 75%, indicating that the lithology recognition model based on logic-CNN algorithm had good recognition effect and performance. The comparison results of recognition effects on sandstone of different algorithms demonstrated that the F1 values of RF, XGBoost, AdaBoost, SVM, and logic-CNN algorithms were 82.57%, 83.13%, 85.72%, 82.67%, and 90.84%, respectively. The accuracy rates of the five algorithms were 86.26%, 87.94%,

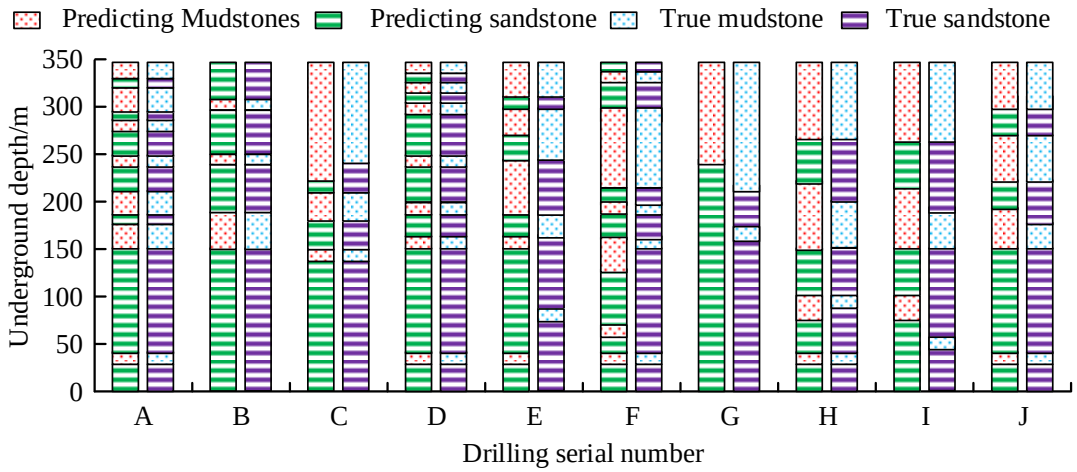


Figure 9. Prediction performance of lithology recognition model based on Logics-CNN.

89.76%, 85.72%, and 90.23%, respectively, and the recall rates were 77.98%, 78.84%, 80.95%, 76.13%, and 90.92%, respectively. The precision rates of the five algorithms were 94.37%, 95.76%, 97.75%, 90.94%, and 98.99%, respectively, with RMSE values of 2.54%, 2.98%, 2.57%, 3.26%, and 1.37%, respectively. The logic-CNN algorithm performed significantly better than the comparison algorithms in terms of F1 value, accuracy, recall, precision, and RMSE, indicating that the algorithm had superior performance.

Conclusion

In response to the common issue of imbalanced logging data in lithology identification models, this study designed SMOTE algorithm and conceived the PSO-GBDT algorithm and logic-CNN lithology identification models. The results validated that, after using SMOTE, the recognition rates of the classification model on seven minority rock types were 91.89%, 93.82%, 94.79%, 95.66%, 98.90%, 100%, and 96.77%, respectively. Among them, the recognition rates of mudstone, argillaceous siltstone, siltstone, volcanic rock, and tuff increased by 1.30%, 3.03%, 1.90%, 1.03%, and 12.09%, respectively. The training times for SVM, PNN, RF, AdaBoost, and PSO-GBDT algorithms were 3 s, 0 s, 13 s, 20 s, and 43 s, respectively. The testing times for the five algorithms were 398 ms, 1,718 ms, 819 ms, 318 ms, and 279 ms, respectively. The lithology recognition model based on logic-CNN algorithm had a prediction accuracy of over 75% in drilling lithology. The F1 value, accuracy, recall, precision, and RMSE of the logic-CNN algorithm were 90.84%, 90.23%, 90.92%, 98.99%, and 1.37%, respectively. This study gave less consideration to the generalization ability of the model, and in the future, it could be improved from multiple perspectives. The designed SMOTE algorithm had a high dependence on lithology labels, and this problem could be avoided from the perspective of unsupervised learning methods in the future.

References

1. Chen C, Zeng W, Hou S, Ding Y. 2023. Design of an unattended ore grading measurement system in a uranium mine. Kerntechnik. 88(1):71-79.
2. Wong WK, Chen KP, Lin JW. 2020. Real-time data logging and online curve fitting using raspberry pi in physics laboratories. Int J Dist Educ Tech. 18(3):57-77.
3. Brown GM. 2021. Lithological and paleocommunity variation on a mississippian (tournaisian) carbonate ramp, Montana, USA. Int J Soc Econ Paleont Miner. 36(3):95-114.
4. Smith MP, Rasmussen JA. 2020. The geology of the Centrums area of Kronprins Christian Land, northeast Greenland, and lithological constraints on speleogenesis. J Cave Karst Stud. 47(2):60-65.

5. Bajwa RS, Ahsan N, Ahmad SR. 2020. A review of landsat false color composite images for lithological mapping of Pre-Cambrian to recent rocks: A case study of pail/padhrar area in Punjab Province, Pakistan. J Indian Soc Remote Sens. 48(5):721-728.
6. Podrezov DR. 2020. Issues of improving control and increasing efficiency of production blocks at an ISL uranium mine. Int J Min Sci Techno. 5(2):131-153.
7. Chen S, Pei H, Pisonero J, Yang S, Fan Q, Wang X. 2022. Simultaneous determination of lithology and major elements in rocks using laser-induced breakdown spectroscopy (LIBS) coupled with a deep convolutional neural network. J Anal Atom Spectrom. 37(3):508-516.
8. Yan B, Zheng L, Wang X, Li Q. 2023. Deep learning-based recognition method of red bed soft rock image. Geol J. 58(6):2418-2426.
9. Hossain TM, Watada J, Aziz IA, Hermana M, Sakai H. 2021. Lithology prediction using well logs: A granular computing approach. Int J Innov Comput I. 17(1):225-244.
10. Hembram S, Gupta SD. 2021. Significance of suitable wavelet estimation to the analysis of spectral decomposition method to detect channel feature: a case study in the Jaisalmer sub-basin, India. J Seism Explor. 30(4):381-404.
11. Han L, Li X, Liu Z, Guo W, Cui Y, Qian C. 2023. Study on rock mechanics characteristics of deep shale in Luzhou block and the influence on reservoir fracturing. Energy Sci Eng. 11(1):4-21.
12. Sun Z, Chen T, Zhu L, Lu J, Zhang S, Pan Z. 2022. Analysis of the upper and lower boundaries of permeability evolution during shale rock shear deformation. Energ Fuel. 36(4):2007-2022.
13. Nohl T, Wetterich J, Fobbe N, Munnecke A. 2020. Lithological dependence of aragonite preservation in monospecific gastropod deposits of the Miocene Mainz Basin: Implications for the (dia-)genesis of limestone-marl alternations. J Sediment Res. 90(11):1500-1509.
14. Silik A, Noori M, Altabey WA, Dang J, Ghiasi R, Wu Z. 2022. Optimum wavelet selection for nonparametric analysis toward structural health monitoring for processing big data from sensor network: A comparative study. Struct Health Monit. 21(3):803-825.
15. Johal SK, Mohana R. 2020. Effectiveness of normalization over processing of textual data using hybrid approach sentiment analysis. Int J Grid High Perfor Comput. 12(3):43-56.
16. Groumpos PP. 2023. A critical historic overview of artificial intelligence: issues, challenges, opportunities, and threats. AIA. 1(4):197-213.
17. Lin L, Guo H, Lv Y, Liu J, Tong C, Yang S. 2022. A machine learning method for soil conditioning automated decision-making of EPBM: hybrid GBDT and Random Forest Algorithm. Eksploat Niezawodn. 24(2):237-247.
18. Arumugam P, Kuppan V. 2020. A GBDT-SOA approach for the system modelling of optimal energy management in grid-connected micro-grid system. Int J Energ Res. 45(5):6765-6783.
19. Resma KS, Sharvani GS, Somula R. 2021. Optimization of cloud load balancing using fitness function and duopoly theory. Int J Intell Comput Cyber. 14(2):198-217.
20. Shi Q, Zhang H. 2020. Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets. IEEE J Ind Electron. 68(7):6248-6256.
21. Wu C, Jiang H, Wang P. 2020. Education quality detection method based on the probabilistic neural network algorithm. Diagnostyka. 21(4):79-86.
22. Zhang W, He Y, Wang L, Liu S, Meng X. 2023. Landslide Susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie, Chongqing. Geol J. 58(6):2372-2387.
23. Li H, Wang S, Islam M, Bobobee ED, Zou C, Fernandez C. 2022. A novel state of charge estimation method of lithium-ion batteries based on the IWOA-AdaBoost-Elman algorithm. Int J Energ Res. 46(4):5134-5151.
24. Li J, Zhang D, Meng B, Li Y, Luo L. 2023. FIMF score-CAM: Fast score-CAM based on local multi-feature integration for visual interpretation of CNNS. IET Image Process. 17(3):761-772.
25. Guo Z, Lin L. 2023. Cognitive physiological data analysis based on the XGBoost algorithm to realize positive perceptual sample classification. J Intell Fuzzy Syst. 44(4):6525-6543.
26. Ahirwar A, Sharma N, Bano A. 2020. Enhanced SMOTE & Fast Random Forest Techniques for credit card fraud detection. Solid State Technol. 63(6):4721-4733.