Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 18, No. 2, 2025

Prediction method of coal seam porosity based on BP neural network

Ping Zhou [1, 2], Xinsi Hu [1, 2] and Danni Wei [1, 2]

[1] School of College of Earth Science and Engineering, Xi'an Shiyou University, Xi 'an 710065, China
[2] Shaanxi Provincia Key Laboratory of Petroleum Accumulation Geology, Xi'an Shiyou University, Xi'an 710065, China

Abstract: Coal seam porosity is a key parameter for evaluating the physical properties of coalbed methane (CBM) reservoirs, and its accurate prediction is crucial for CBM exploration and development. Traditional log interpretation methods exhibit limited accuracy when dealing with multifactor nonlinear relationships. This paper proposes a coal seam porosity prediction model based on an error back propagation (BP) neural network, which establishes the nonlinear relationship between optimized conventional logging curves (AC, CAL, CNL, DEN, GR, RT) and porosity (POR). Input variables were screened using Spearman correlation analysis, and a three-layer BP neural network structure was constructed. The network weights and thresholds were optimized via the back-propagation algorithm. Experimental results demonstrate that the model effectively learns the complex mapping relationship between logging parameters and porosity. The predicted results show strong agreement with measured values, with an average relative error of less than 8%, outperforming traditional regression methods. This approach provides a reliable technical means for predicting coal seam porosity.

Keywords: BP neural network; Coal seam porosity; Logging curve; Predictive model; Machine learning.

1. Introduction

Coal seam porosity is a key physical parameter that affects the adsorption, storage, and permeability of coalbed methane. Its accurate prediction is crucial for the evaluation of coalbed methane resources and the formulation of development plans [1].As a complex dual medium of pores and fractures, coal seams have strong heterogeneity in their pore structure. Traditional prediction methods based on linear regression or empirical formulas are difficult to fully capture the nonlinear and high-dimensional coupling relationship between logging response and porosity [2]. With the deepening of coalbed methane exploration and development, higher requirements have been put forward for the accuracy of porosity prediction.

In recent years, with the development of artificial intelligence technology, neural networks have been widely used in geophysical parameter prediction due to their powerful nonlinear fitting ability [3]. In the study of predicting coal seam porosity, multiple artificial logging parameters have provided solid evidence for the correlation between porosity and logging parameters. Shao Xianjie et al. [4] established a logging interpretation method for coal rock parameters through the case study of Hancheng mining area, providing an important foundation for the study of the relationship between logging parameters and porosity. Shen Yanjun et al. [5] conducted a correlation analysis on the pore structure of oil-rich coal in the Yushenfu mining area of northern Shaanxi, and found that the total porosity, macropores, mesopores, etc. were negatively correlated with compressive strength, while the micropore porosity was positively correlated with compressive strength. Zheng Guiqiang et al. [6] conducted a logging evaluation study on the characteristics of deep coal reservoirs in the Qinshui Basin, revealing that density (DEN) is the most critical factor affecting porosity. Zhang Chao et al. [7] conducted a study on coal structure identification based on well logging curves and found a significant positive correlation between acoustic time difference (AC) and porosity in multiple coalfields; In addition, studies have shown that neutron porosity (CNL) and natural gamma (GR) also have good indicative significance for porosity interpretation in certain coal seams. Cai et al. used experimental methods to study the effect of liquid nitrogen cooling on the pore structure of rocks and found that temperature changes have a significant impact on the pore structure. Foreign scholars have also conducted relevant research. Salmachi et al. [9] conducted experiments on the effect of CO_2 injection on the pore structure of coal in the Bowen Basin, Australia, and revealed that CO_2 injection can significantly reduce the effective porosity of coal seams. These studies not only reveal the correlation between specific logging parameters and porosity, but also confirm the universality of nonlinear relationships in predicting coal seam porosity.

With the deepening of coalbed methane exploration and development, higher requirements have been put forward for porosity prediction accuracy, and in recent years, the development of artificial intelligence technology has provided new ways to handle such complex relationships [10]. Neural networks have been widely used in predicting geophysical parameters due to their powerful nonlinear fitting ability, demonstrating promising application prospects. Liu Zhenming et al. [11] established a coal structure prediction model based on well logging data using BP neural network, providing important reference for the application of neural network in coal reservoir parameter prediction. Especially BP neural network (BackPropagation Neural Network), as a multi-layer feedforward neural network, can gradually approximate complex nonlinear function relationships by adjusting network parameters through error backpropagation algorithm [12], and has shown good application prospects in the field of geophysical parameter prediction [13].

This article aims to construct a coal seam porosity prediction model based on BP neural network, using various conventional logging data as inputs. By analyzing the correlation between logging parameters and porosity, optimizing the input feature variables, high-precision prediction of porosity can be achieved, providing new ideas and methods for coal seam gas logging interpretation. The main innovation of this study lies in: ① using Spearman correlation analysis method to scientifically screen input features; ② Optimize the structural parameters of BP neural network; ③ An intelligent model suitable for predicting coal seam porosity has been established.

2. Back Propagation (BP) neural network

2.1. The Structure of BP Neural Network

The structure of BP neural network includes input layer, output layer, and several hidden layers, each layer has several nodes, and the link status between nodes in each layer is reflected by weights. Among them, the number of nodes in the input layer is equal to the input dimension, the number of nodes in the output layer is equal to the output dimension, and the number of nodes in the hidden layer can be set according to the actual situation. In a BP neural network, each node contains a perceptron (i.e. a separate neuron) that includes input terms, weights, biases, activation functions, and outputs. In the forward propagation process, the input data is processed by the perceptron node and the activation function to obtain the output result; In the process of backpropagation, the results are compared with the expected results, and the weights of each node on the network are continuously adjusted through multiple iterations. A typical three-layer neural network is shown in Figure 1.

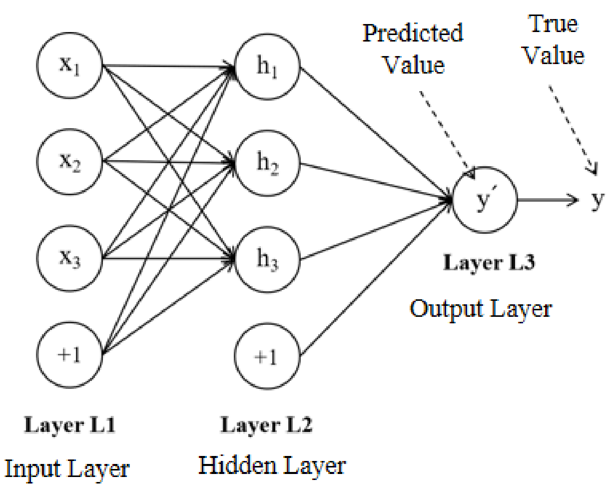

Figure 1. BP neural network structure

The calculation formula for each layer is:

$$y = T(WX + b) \tag{1}$$

Among them, T represents the activation function, b represents the activation threshold, and W represents the connection weight.

For multilayer networks, the feedforward propagation method is used to calculate, that is, each layer is calculated according to the above formula until the last output layer.

2.2. Mathematical Expression of BP Neural Network

After the structure of BP neural network is determined, the expression of the model is also determined. Further, we need to determine W,b(weight threshold) in the model so that the network prediction is the most accurate, that is, the error is minimal. Regarding the error of BP neural network, the mean square error function of BP neural network is as follows:

$$E(W,b) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{k}\sum_{j=1}^{k}\left(\hat{y}_{ij} - y_{ij}\right)^2 \tag{2}$$

Among them, m: the number of training samples; k: the number of outputs; $\hat{y}_{ij}$:the predicted value of the kth output of the ith sample; y_{ij} :he true value of the kth output of the i-th sample.

The error function is a function about W,b. If different weights and thresholds are used, there will be different errors.

A set of data enters the hidden layer through weighted and nonlinear transformation of the input layer, and then finally enters the output layer through the same operation. The output result is our predicted value. At this time, it is necessary to compare the difference between the predicted value and the real value, adjust the weight and bias through partial derivative calculation, and continuously iterate to obtain the predicted value similar to the real value. The mathematical expression of BP neural network is as follows:

$$\begin{aligned} y = {} & w_{11}^{(3,2)} * \text{tansig}\left(w_{11}^{(2,1)}x_1 + w_{12}^{(2,1)}x_2 + b_1^{(2)}\right) \\ & + w_{12}^{(3,2)} * \text{tansig}\left(w_{21}^{(2,1)}x_1 + w_{22}^{(2,1)}x_2 + b_2^{(2)}\right) \\ & + w_{13}^{(3,2)} * \text{tansig}\left(w_{31}^{(2,1)}x_1 + w_{32}^{(2,1)}x_2 + b_3^{(2)}\right) \\ & + b_1^{(3)} \end{aligned} \tag{3}$$

There are many parameters in the expression, but there are actually only two types of parameters: weight w and threshold b. Where $w_{12}^{(3,2)}$ represents that this weight is the weight of the second node in the second layer to the first node in the third layer; $b_2^{(2)}$ 2b represents that this threshold is the threshold of the first node of the second layer. The superscript is used to indicate which layer is, and the subscript is used to indicate which neuron is the layer.

2.3. Training of BP neural network

The training of BP neural network uses the backpropagation algorithm (Backpropagation). Backpropagation is an optimization algorithm. By continuously adjusting the connection weights between each neuron in the network, the neural network can learn the mapping relationship between input and output. Specifically, the backpropagation algorithm advances the calculation error from the last layer by calculating the state and activation values of each layer, and updates the parameters to minimize the error between the predicted output and the actual output of the network. This process will iterate until the stop criterion is met (such as the difference in errors between adjacent iterations is small).

In general, every step iterates, the error is reduced by a small step, and finally a local optimal weight and threshold is obtained. The training algorithm process of BP: 1. Initialize weights and thresholds; 2. Calculate the gradient of weights and thresholds; 3. Iterate the weights and thresholds in the direction of negative gradients; 4. Check whether the termination condition is terminated, otherwise repeat 2.3. The flow chart of BP neural network training is as follows:

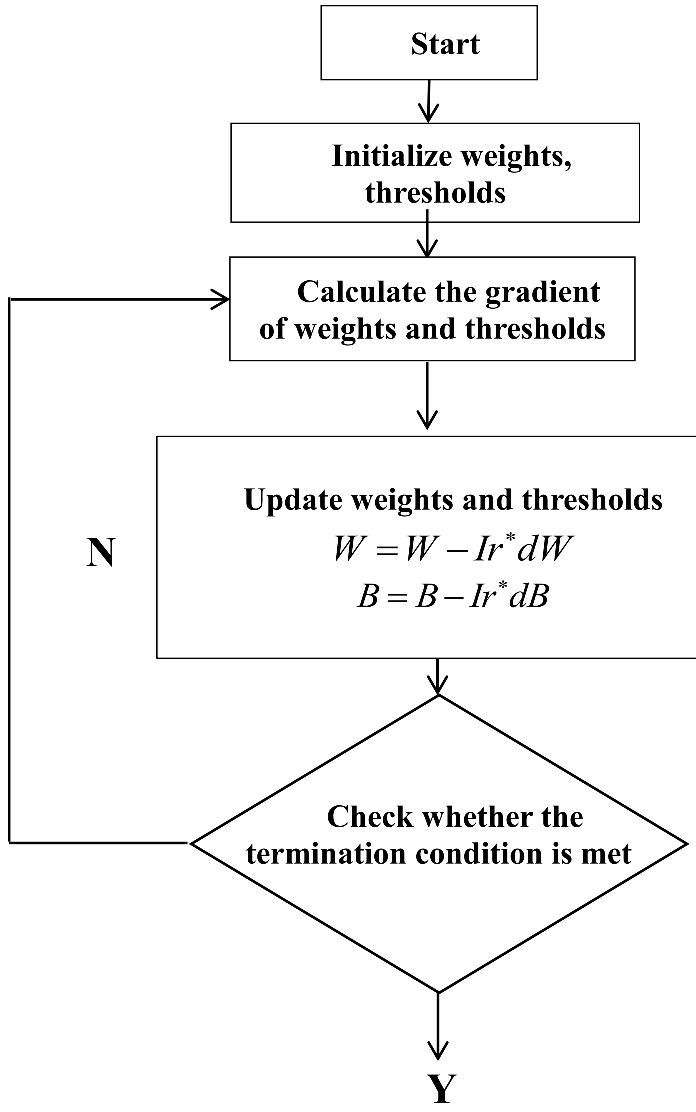

Figure 2. BP neural network training flow chart

3. Data preparation and neural network modeling

3.1. Data source and preprocessing

This study uses well logging data from a certain coalbed methane block, including AC (sonic time difference), CAL (well diameter), CNL (neutron porosity), DEN (density), GR (natural gamma), RT (resistivity) and POR (porosity).There are often problems such as missing values or inconsistent dimensions in the original data, so systematic data cleaning and standardization are required.

First, missing values were filled using the method of interpolation with adjacent data, and obvious outliers were eliminated. Subsequently, to eliminate the dimensional differences among various parameters, all input features were standardized using the Z-score normalization method [14]:

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{4}$$

Where x is the original data, μ is the mean of the data, and σ is the standard deviation of the data. The standardized data mean is 0 and the standard deviation is 1, which is conducive to improving the model training efficiency and convergence speed.

3.2. Feature Screening Based on Spearman Correlation Analysis

In correlation analysis, Pearson's correlation coefficient is more common, and it is mainly used to measure linear relationships between variables. However, compared with the Pearson correlation coefficient, Spearman rank correlation coefficients show stronger robustness and applicability when dealing with data with high nonlinearity, monotonous but not necessarily linear correlation, and can more effectively identify the correlation trends between variables. In addition, Spearman correlation coefficient is a non-parametric statistical method, and has no requirements for the overall distribution of variables and is not affected by variable dimensions. Therefore, it is suitable for a variety of data types including continuous variables, ordered categorical variables and even some disordered categorical variables [15][16]. Its specific expression is:

$$\rho = 1 - \frac{6\sum d_i^{\,2}}{n\left(n^2 - 1\right)} \tag{5}$$

where d_i is the rank difference between the two variables, n is the number of samples, and ρ represents the Spearman relationship.

The analysis results are shown in the figure (Figure 3). The correlation coefficients of each parameter and porosity are: AC (0.80), CAL (0.26), CNL (0.29), DEN (-0.89), GR (-0.46), RT (0.17), and SP (0.036).

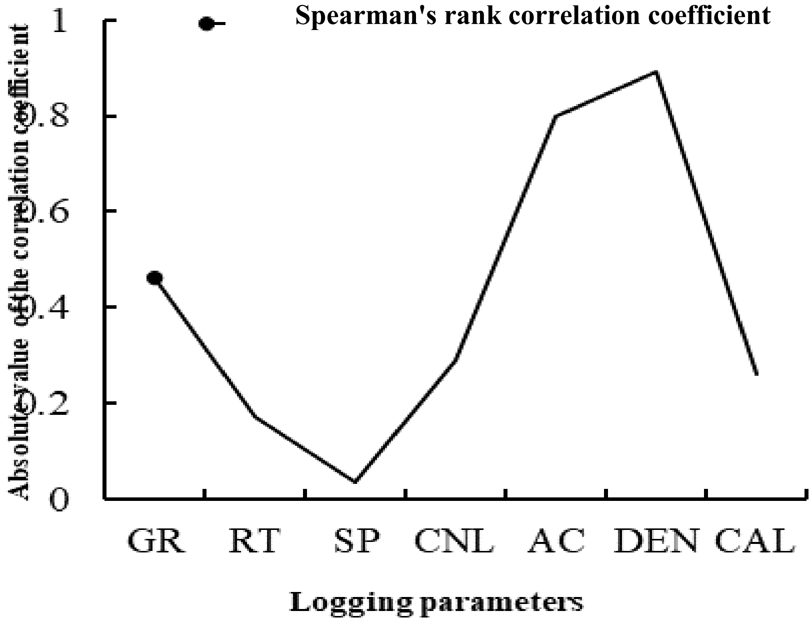

Figure 3. Linear graph of Spearman correlation coefficient

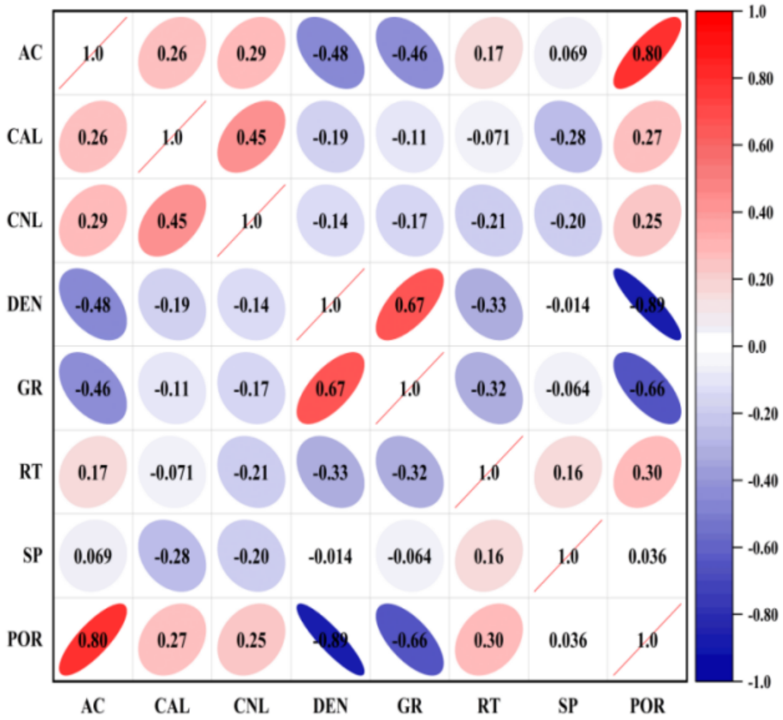

Figure 4. Spearman correlation coefficient heat map

The analysis results show that the Spearman correlation coefficients of each well logging parameter and porosity (POR) are as follows: AC (0.80), CAL (0.26), CNL (0.29), DEN (-0.89), GR (-0.46), RT (0.17), and SP (0.036).The results show that DEN is strongly negatively correlated with POR, and AC is strongly positively correlated with POR, which is consistent with the physical mechanism of coal seam: the greater the density, the smaller the porosity; the longer the sound wave propagation time, the greater the porosity. The correlation between SP and POR is extremely weak (0.036), so it is eliminated. Finally, six parameters such as AC, CAL, CNL, DEN, GR, and RT were selected as the model input features.

3.3. Neural network structure design and parameter settings

This study constructed a three-layer BP neural network containing 6 input nodes, 12 nodes in a single hidden layer and 1 output node. The hidden layer adopts the ReLU activation function to enhance the nonlinear fitting capability [17], and the output layer uses a linear activation function to adapt to the regression prediction task; Network training adopts the driving force gradient descent algorithm with a learning rate of 0.01 (Momentum=0.9) [18] to accelerate convergence and improve generalization performance. The training stop condition is set to have an error change of less than 10^{-5} or the maximum number of iterations reaches 10,000 times, so as to effectively control the risk of overfitting while ensuring accuracy.

3.4. Model training and verification

The preprocessed data set is randomly divided into the training set and the test set at a ratio of 7:3. The training set is used for model parameter learning, and the test set is used to evaluate the generalization ability of the model. During the training process, the mean square error (MSE) of each iteration is recorded and the error drop curve is drawn (as shown in Figure 5) to monitor whether the training process converges.

Mean Squared Error (MSE)

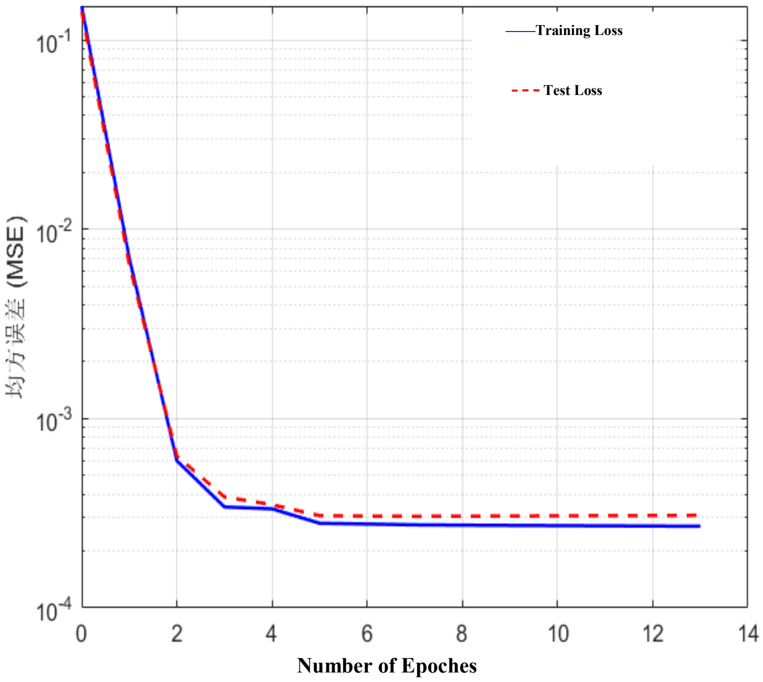

Figure 5. BP neural network training and testing loss

The error of the model continues to decline on the training set, and the performance is stable on the test set, and no obvious overfitting is seen. After the training is completed, the test set is used for prediction, and the prediction effect is quantitatively evaluated using three indicators: Decision coefficient (R^2), mean absolute error (MAE) and root mean square error (RMSE).

4. Results and Analysis

4.1. Predictive effect evaluation

In order to comprehensively evaluate the performance of BP neural network model in coal seam porosity prediction, this study used a variety of statistical indicators for comprehensive evaluation, including the determination coefficient (R^2), mean absolute error (MAE), and root mean square error (RMSE) [19]. These indicators reflect the prediction accuracy from multiple angles such as model goodness of fit, absolute error level and error distribution stability. The calculation formula is as follows:

$$R^2 = 1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y}_i)^2}, \quad MEA=\frac{1}{n}|y_i-\hat{y}_i|, \quad \mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{n}}\sum(y_i-\hat{y}_i)^2} \tag{6}$$

Among them, y_i is the actual measured porosity value, $\hat{y}_i$ is the model predicted value, $\bar{y}_i$ is the average value of the actual measured value, and n is the number of samples. As shown in Table 1, the R^2 of the BP neural network model reaches 0.91, the MAE is 0.062%, and the RMSE is 0.089%, indicating that the model has high interpretation variance ability and stability and has small prediction errors.

4.2. Comparative analysis

In order to objectively evaluate the superiority of BP neural network model, this paper compares it with two traditional machine learning methods: Multiple Linear Regression (MLR) and Support Vector Machine (SVM). As can be seen from the table below, BP neural networks are significantly better than the comparison method in all evaluation indicators: (Table 1).

Table 1. Error values for different models in the test set

method	R^2	MAE	RMSE
MLR	0.76	0.098	0.124
SVM	0.82	0.085	0.110
BP neural network	0.91	0.062	0.089

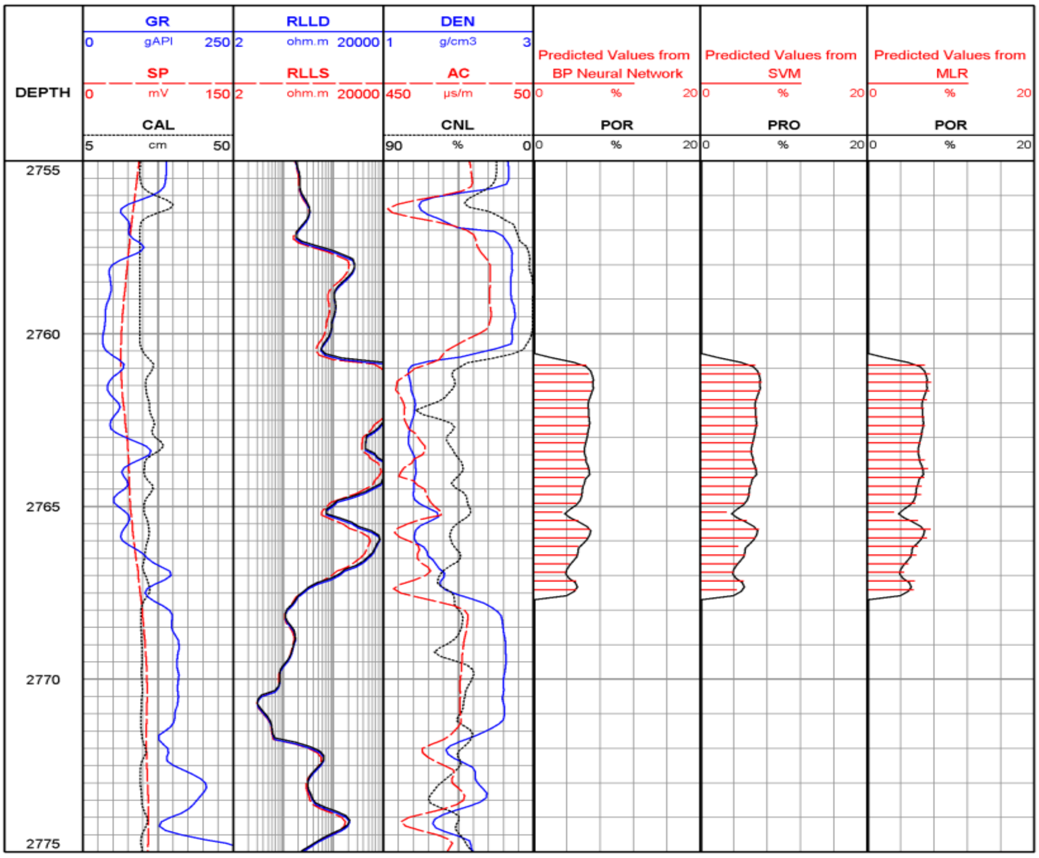

Figure 6. Porosity prediction results of different models (well A)

5. Conclusion

This paper constructs a coal seam porosity prediction model based on BP neural network, and optimizes input features through Spearman correlation analysis to achieve accurate prediction of porosity. The main conclusions are as follows:

(1) Spearman's correlation analysis shows that DEN (density) and AC (sonic time difference) are the two parameters with the strongest correlation with porosity, with correlation coefficients reaching -0.89 and 0.80 respectively; while the correlation between SP (natural potential) and porosity is only 0.036, which is eliminated.

(2) The optimal network structure was determined through the grid search method: 6 nodes in the input layer, 12 nodes in the single hidden layer, and 1 node in the output layer. This structure avoids overfitting while ensuring prediction accuracy.

(3) The neural network can effectively capture the complex nonlinear relationship between logging parameters and porosity, and the prediction accuracy is significantly higher than that of traditional methods (R^2=0.91, MAE=0.062%, RMSE=0.089%)

References

[1] Jin Zeliang, Xue Haifei, Gao Haibin, et al. Coalbed methane reservoir logging evaluation technology and application [J]. Coalfield Geology and Exploration, 2013, 41(02): 42-45.

[2] Yu Zhaolin, Wang Yanbin, Yu Yun, et al. Porosity logging prediction model of deep coal reservoir based on diversified regression [C]//The Petroleum Geology Committee of the China Petroleum Society, the Coal Methane Professional Committee of the China Coal Society, the Coal Methane Industry Technology Innovation Strategic Alliance. New Progress in Coal Methane Exploration and Development Technology - Proceedings of the 2018 National Coal Methane Academic Seminar. School of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), 2018:424-429.

[3] Yang Liuqing, Chen Wei, Cha Bei. Research and application of prediction of reservoir porosity using convolutional neural networks[J]. Progress in Geophysics, 2019, 34(04):1548-1555.

[4] Shao Xianjie, Sun Yubo, Sun Jingmin, et al. Explanation method of coal rock parameter logging - Taking Hancheng Mining Area as an example [J]. Petroleum Exploration and Development, 2013, 40(05):559-565.

[5] Shen Yanjun, Ma Wen, Wang Xu, et al. Research on the correlation between pore development and intensity characteristics of oil-rich coal in different coal seams[J]. Coal Mine Safety, 2023, 54(03):161-168.

[6] Zheng Guiqiang, Yang Defang, Li Xiaoming, et al. Research on the characteristics of deep coal reservoir logging in the Shizhuang North Block in Qinshui Basin [J]. Coal Science and Technology, 2019, 47(06): 178-186.

[7] Zhang Chao, Huang Huazhou, Xu Delin, et al. Coal structure identification based on logging curves[J]. Coal Science and Technology, 2017, 45(09): 47-51.

[8] CAI C Z, LI G S, HUANG Z W, et al Experimental study of the effect of liquid nitrogen cooling on rock pore structure[J]. Journal of Natural Gas Science and Engineering, 2014, 21: 507-517.

[9] SALMACHI A, ZEINIJAHROMI A, ALGARNI M S, et al Experimental study of the impact of CO_2 injection on the pore structure of coal: A case study from the Bowen Basin,

Australia[J]. International Journal of Coal Geology, 2023, 275: 104314.

[10] Hu Rui, Li Yong, Liu Yingtian, et al. Prediction method for logging reservoir parameters based on CL-Trans model [J]. Geophysical and chemical exploration computing technology, 2025, 47(03): 410-419.

[11] Liu Zhenming, Wang Yanbin, Han Wenlong, et al. Prediction of coal structure of BP neural network based on well logging data[J]. China Coal, 2018, 44(06): 38-41+55.

[12] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

[13] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281–305.

[14] Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. John Wiley & Sons.

[15] Hauke J, Kossowski T. 2011.Comparison of values of Pearson's andSpearman's correlation cefficients on the same sets of data. Quaestiones Geographicae, 30(2):87-93.

[16] Myers L, Sirois M J. 2014. Spearman correlation coefficients, differences. Between .// Wiey StatsRef:Statistics Reference Online.Washington: John Wiley & Sons.

[17] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 807–814).

[18] Qian, N. (1999). On the momentum term in gradient descent learning algorithms. Neural Networks, 12(1), 145–151.

[19] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.