Research Article

Parameter Acquisition Study of Mining-Induced Surface Subsidence Probability Integral Method Based on RF-AGA-ENN Model

Jinman Zhang, Liangji Xu, Jiewei Li, Yueguan Yan, and Ruirui Xu

1College of Geoscience and Surveying Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China
2State Key Laboratory Coal Resources and Safe Mining (Beijing), Beijing 100083, China
3School of Spatial Information and Geomatics Engineering, Anhui University of Science and Technology, Huainan 232001, China
4State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Huainan 232001, China
5Geological Exploration Bureau of Zhejiang Province, Hangzhou 310052, China
6Anhui Provincial Bureau of Coal Geology, Hefei 230088, China

Correspondence should be addressed to Liangji Xu; ljxuaust@163.com

Received 23 November 2021; Revised 10 September 2022; Accepted 5 October 2022; Published 25 October 2022

Abstract

The mining of underground coal resources can trigger geological hazards such as subsidence basins, cave-in pits, and step cracks. In China, the probability integral method (PIM), the most popular method for predicting surface movement deformation caused by coal resource mining, has a prediction accuracy that is mainly influenced by both the measurement data (i.e., quantity and quality) from ground movement observatories and the parameter inversion method. To obtain more accurate PIM parameters in the absence of observational data, we propose a combined machine learning model (RF-AGA-ENN)—random forest (RF) extracts the best combination of features as the input layer of Elman neural network (ENN); ant colony algorithm (ACO) and genetic algorithm (GA) are combined (called AGA) for the weights and thresholds of ENN optimization. The results of the study show that (1) the RF-AGA-ENN model is used to obtain PIM values with MAXRE values between 1.94% and 9.18%, AVERY values between 0.98% and 3.98%, and RMSE values between 0.0050 and 0.9632. (2) Compared with the PIM parameters obtained from BP neural network, RF-ENN, RF-ACO-ENN, and RF-GA-ENN models, the PIM parameters obtained from the RF-AGA-ENN model have better stability and accuracy. (3) According to the PIM parameters obtained by the RF-AGA-ENN model, the predicted and measured values of surface settlement at the 11111 working face have a high degree of agreement. In summary, the RF-AGA-ENN model to obtain the PIM parameters has good application value.

1. Introduction

As the cornerstone of China’s healthy and sustainable economic development, coal resources have greatly contributed to the rapid development of China’s economy in the past decades [1]. Due to the current energy structure of more coal, less gas, and poor oil, the dominant position of coal as China’s energy source will remain unchanged for a long time. Coal as the “stabilizer” and “ballast” of the national energy supply will continue to bear the burden of national energy security. As the “stabilizer” and “ballast” of national energy supply, coal will continue to bear the heavy responsibility of national energy security and sustainable economic development [2].

The mining of underground coal resources leads to a series of geological hazards, such as subsidence basins, collapse pits, and step cracks, which bring a series of hazards to the product life and ecological environment of people in mining areas [3–5]. To minimize the geological hazards in mining areas caused by mining subsidence so that reason-
Figure 1: ENN network structure diagram.

Figure 2: RF feature selection flow chart.
able preventive measures can be taken in advance, it is crucial to accurately anticipate the ground movement deformation caused by coal resources mining [6–8]. Based on the actual measured settlement data in the mine area, Zhang et al. obtained the rock mechanical parameters through orthogonal experiments and numerical simulation inverse analysis and then used numerical simulation to predict the surface settlement caused by two-layer coal mining [9]. Zhu et al. proposed a superposition prediction model for infill strip mining based on the traditional probabilistic integral method to accurately predict the surface movement deformation caused by infill strip mining [10]. Zhou et al. constructed a combined prediction model for mining subsidence by using alluvium and bedrock as two different media in thick alluvial mining areas [11]. As the most widely used and mature method in the field of mining subsidence in China, scholars have conducted a lot of research in recent decades on how to improve the accuracy of the parameters of the probabilistic integral method of the estimation model. In recent years, the research on the acquisition of PIM parameters is mainly focused on the following two aspects: one is to combine PIM and intelligent optimization algorithms (such as genetic algorithm and ant colony algorithm) [12–14] based on the measurement data of mobile surface observatory to invert PIM parameters, because the establishment of the observatory requires a lot of human, material, and financial resources and has an obvious lag, so this method has limitations when facing the lack of another method is to use machine learning algorithms (such as support vector machine and BP neural network) [15–17] to establish the nonlinear relationship between PIM parameters and geological mining conditions based on the existing geological mining conditions and the corresponding PIM parameters data set, which is a good solution to the nonlinear relationship between PIM and geological mining conditions. This method is a good solution to the difficulty of

Figure 3: The flow chart of the AGA algorithm.
expressing the nonlinear relationship between PIM and geological mining conditions by mechanical and mathematical formulas. The method has a high application value.

Since the 20th century, more than 200 surface movement observation stations have been established in typical mining areas in China to study the surface movement deformation during coal mining activities, which has accumulated a large amount of observation data to guide the production activities in mining areas [18]. For these observations, scholars have established empirical relationships between PIM parameters and geological mining conditions in typical mining areas [19–21]. Still, in many cases, these empirical relationships are difficult to accurately express the complex nonlinear relationships between PIM parameters and geological mining conditions. For this reason, Guo et al. proposed constructing an optimized neural network model for PIM parameter acquisition by using an improved BP neural network model to learn and train a large number of surface mobile observatory real data [22]. Zhao et al. built a random forest regression prediction model for finding the surface subsidence coefficient based on a large number of mobile surface observatory measured data [23]. Wang et al. used the GA algorithm to search for the optimal smooth factor of GRNN and constructed a GA-GRNN model to predict the surface subsidence coefficient based on the observation data of a typical mining area in China [24]. Li constructed a support vector machine prediction model based on the ACO algorithm by optimally selecting the parameters of the support vector machine using the ACO algorithm [25]. Chi et al. used the MIV and GP algorithms to optimize the BP neural network to construct a PIM parameter prediction model [26]. These prediction models achieved better accuracy and provided a good idea to obtain PIM parameters for working faces lacking observation data.

ENN networks (Figure 1) have an extra memory layer than BP neural networks, giving the whole network a richer dynamic, and making the various network more robust. To address the problem that the initial weights and thresholds of ENN networks are challenging to determine, Wang and Jiang proposed to use the GA algorithm to search for the weights and thresholds of ENN networks, but GA has insufficient late search capability [27]. Wang and Zhao proposed to use GA for search in the early stage and ACO for investigation in the later stage to complement each other’s advantages [28]. Zhang and Jiang constructed the AGA algorithm using the GA algorithm to improve the ACO algorithm from the perspective of solving the distribution scheduling of intelligent unmanned trucks in mines and constructed the optimal distribution scheduling model of unmanned trucks with the combined AGA algorithm [29]. Xiong combined and optimized the GA algorithm and ACO algorithm to construct the AGA algorithm and used the AGA algorithm to optimize the support vector machine parameters to construct the subsidence prediction model for slope instability prediction [30]. To study the slope stability under the influence of ore body mining, Shi et al. combined the GA algorithm and the ACO algorithm to construct the AGA algorithm and used the AGA algorithm to search for the critical sliding surface to determine the safety factor of the slope [31, 32]. Liu et al. proposed to use RF for feature extraction to effectively reduce the dimensionality of the input layer of the LSTM model and establish the RF-LSTM prediction model [33]. However, there is no relevant research in the field of mining subsidence. In this paper, RF is used to simplify the complexity of the ENN network, and the AGA algorithm is used to search for the optimal smooth factor of the slope [31, 32]. Liu et al. proposed to use RF for feature extraction to effectively reduce the dimensionality of the input layer of the LSTM model and establish the RF-LSTM prediction model [33]. However, there is no relevant research in the field of mining subsidence. In this paper, RF is used to simplify the complexity of the ENN network, and the AGA algorithm is used to search for the optimal smooth factor of the slope [31, 32].

This paper collects 70 sets of actual measurement data of coal mining faces as experimental data. Firstly, the RF algorithm is used to calculate the OOB error to obtain the optimal data set as the input layer of the ENN network to simplify the complexity of the ENN network. Then, the AGA algorithm is used to optimize the weights and thresholds of the ENN network and establish the RF-AGA-ENN model for the prediction of PIM parameters.

2. Construction of RF-AGA-ENN Model

2.1. Theory of RF. RF algorithm, proposed by Breiman in 2001, is a variant of the Bagging algorithm, a new supervised
machine learning algorithm. RF is a classifier that uses multiple decision trees to learn and integrate predictions on samples. It uses the Bootstrap resampling technique to construct multiple samples randomly from samples, then uses the random splitting technique of nodes for each resampled sample to construct multiple decision trees, and finally, combines the multiple decision trees to arrive at the final prediction result by voting. The core idea of the RF algorithm is to randomly draw \( N \) samples from the original training set in a put-back manner. That is, for these \( N \) samples, \( N \) samples are randomly selected \( N \) times; each time, one sample is selected from the \( N \) samples and then "replicated"; in the next sampling, the sample set is still \( N \). Since the sampling process is put back, some samples may be selected several times and appear in the same training set several times, while others may not be selected once; these ignored samples are called "out-of-bag data (OOB)."

RF algorithms often use the Gini index as the division function or OOB error as the generalization error to measure the importance of features. OOB error can not only evaluate the critical attribute of each feature but also evaluate the generalization error of RF, so this paper uses OOB error to measure the importance of features. The main steps of RF feature selection are as follows: (1) dividing the sample data into training and test sets, Bootstrap resampling of the training set; (2) using a classification regression tree (CART) to build a decision tree as a base classifier and calculate the OOB error for each training subset; (3) classifying the test set by a strong classifier composed of a large number of base classifiers to determine the final diagnostic result through a voting mechanism (Figure 2).

2.2. Model of Optimized ENN

2.2.1. Theory of AGA. The ACO algorithm is a global optimization intelligent bionic algorithm using distributed parallel computing, which has the advantages of strong robustness and easy integration with other methods. The AGA algorithm overcomes the shortcomings of the ACO algorithm and the GA algorithm to achieve the purpose of complementary advantages.
Figure 5: Continued.
Figure 5: Continued.
Firstly, the GA algorithm is used to obtain a better solution. Then, the initial pheromone of the ACO algorithm is set according to the better solution to guide the ACO algorithm to continue the search for the best solution.

The operation flow of the AGA algorithm is shown in Figure 3, and the specific steps are given as follows:

1. Initialize chromosomes: set the number of populations $I$, the hybridization probability $P_c$, the mutation probability $P_m$, etc. Keep the current as of the global optimum

2. Perform selection, crossover, and mutation operations with ants with low fitness. Calculate fitness and update parent and offspring: update pheromone and current global optimum. Determine whether the number of iterations reaches the preset value, and if so, output the better solution to step 3; otherwise, repeat step 2. The initial pheromone can be obtained from

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \sum_{k=1}^{\text{cell}(I/2)} \Delta \tau_{ij}^k,$$  \hspace{1cm} (1)$$

$$\Delta \tau_{ij}^k = \begin{cases} 2Q/L_k, & k \text{ is the optimal individual}, \\ Q/L_k, & k \text{ is a non-optimal individual}, \\ 0, & \text{others}, \end{cases}$$  \hspace{1cm} (2)$$

where $I$ is the number of ant populations, $Q$ is a constant, and $L_k$ is the path taken by individual $k$, $\tau_{ij}(t)$ is the pheromone concentration from position $i$ to position $j$ at time $t$.

3. The ant colony parameters are set according to the optimal solution obtained by the GA algorithm: ant population size $I$, maximum evolutionary generation $G$, transfer probability coefficient $P_0$, etc. Each ant will randomly select the next state point and store the record when constructing the path, and the formula for calculating the probability selection is shown in

$$P_{ij}^k = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}(t)]^\beta}{\sum_{u \in v_k} [\tau_{iu}(t)]^\alpha [\eta_{iu}(t)]^\beta}, & j \in v_k, \\ 0, & \text{others}, \end{cases}$$  \hspace{1cm} (3)$$

where $P_{ij}^k$ is the probability of ant $k$ choosing position $j$ at position $i$, $\eta_{ij}(t)$ is the heuristic, $v_k$ is the set of all possible positions chosen by ant $k$ at position $i$, $\alpha$ is the important factor, and $\beta$ is the important factor of the heuristic end, etc.

4. The pheromone is updated, and the global optimum is recorded according to the individual ant’s merit-seeking process, and the pheromone update is shown in

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \Delta \tau_{ij}(t),$$  \hspace{1cm} (4)$$

Figure 5: RW smoothing results.
Figure 6: Comparison of ENN model prediction values before and after AGA optimization.
Table 4: Prediction result accuracy index values.

| PIM parameters | Accuracy index | RF-ENN | RF-AGA-ENN |
|---------------|---------------|--------|------------|
|               | MAXRE         | AVERE  | RMSE       |
| \( q \)       | 7.69          | 3.62   | 0.0423     |
|               | 6.93          | 3.55   | 0.0097     |
| \( b \)       | 5.98          | 4.08   | 0.0816     |
| \( \tan \beta \) | 9.87          | 6.25   | 2.57       |
|               | 8.97          | 2.85   | 0.544      |
| \( s/h \)     | 3.98          | 1.72   | 0.98       |
|               | 9.18          | 1.5242 | 0.9632     |

\[
\Delta r_{ij}(t) = \sum_{k=1}^{j} r_{ij}^k(t) \theta, \quad (5)
\]

where \( \rho \) is the volatility coefficient of pheromone; other parameters are the same as the above formula parameter meaning.

(1) Set the number of nodes of the input layer, implicit layer, takeover layer, and output layer of ENN according to the actual problem, and initialize the configuration of ENN network parameters.

(2) Set the relevant parameters of the AGA algorithm, the settings of hybridization probability and mutation probability, initial population size, the maximum number of evolutionary generations, pheromone volatility coefficient, etc.

(3) According to the initialization parameters of the AGA algorithm and the iterative determination criterion, the iterative operation is performed until the conditions for the algorithm to stop iterating are satisfied, the optimal individual is obtained, and then, the optimal weight and threshold of ENN are obtained.

(4) The optimal weights and thresholds obtained by the AGA algorithm are used as the initial weights and thresholds for ENN network training.

(5) The ENN network is continuously trained and learned until the result satisfies the training stopping condition and the optimal ENN network is obtained.

3. Example Application and Result Analysis

3.1. Data Dimensionality Reduction by RF Algorithm

3.1.1. Experimental Data. The literature [34] counted 408 sets of relatively complete measured surface movement actual data of coal mining workings in all major coal mines in China. In this paper, 70 sets of measured coal mining face data are selected as experimental data, of which 50 sets of coal mining face data are used as training sets to train the RF-AGA-ENN model, and the other 20 sets of coal mining face data are used as test data to verify the expected accuracy of the RF-AGA-ENN model.

The experimental data consists of two parts, one is the geological mining conditions, which are used as the input data for the training and learning of the RF-AGA-ENN model, and the other is the PIM parameters, which is used as the output of the training and learning of the RF-AGA-ENN model. The geological mining conditions include the following: mining thickness \( M \), coal seam inclination \( \alpha \), mining depth \( h \), the ratio of the depth of extraction to the thickness of extraction \( h/M \), area mined \( A \), the ratio of working face inclination length to mining depth \( D/h \), workforce advance speed \( v \), the ratio of bedrock to loose layer thickness \( h/h_s \), loose layer thickness \( h_s \), and overlying rock compressive strength \( R_s \). The PIM parameters include the following: sink factor \( q \), horizontal movement factor \( b \), main influence angle tangent \( \beta \), the ratio of inflection point offset distance to mining depth \( s/h \), and influence propagation angle \( \theta \). Only some of the data are presented in the paper, and the data are shown in Table 1. The complete experimental data have been given in the supplementary material (available here).

3.1.2. Selection of Input Variables. The RF-based feature selection in this thesis is implemented using python programming. To obtain more accurate PIM parameters, RF classifiers for each parameter of PIM are constructed.
separately. Based on the constructed RF classifier, the OOB error of each geological mining condition is calculated, and the smaller the OOB error indicates, the greater importance of the results. Accordingly, the significance of each geological mining condition and other variables for each PIM parameter is derived. The OOB errors of each geological mining condition and each PIM parameter are shown in Table 2.

The OOB errors for the effects of each geological mining condition on the PIM parameters are listed in order from smallest to largest, as shown in Table 3. The smaller the OOB error indicates, the greater importance of the corresponding geological mining conditions. The five influencing factors with the greatest importance are selected as the input layer of the model, and the complexity of the input layer network is simplified to improve the model’s prediction accuracy.

3.2. Experimental Data Noise Reduction Processing. Since the field observation environment is often very complex, there is inevitably a certain amount of error in the data collection process. In addition, the training sample data and the test sample data often have a large difference in matching, which can greatly impact the prediction results. RW is a noise

| PIM parameters | $q$ | $b$ | $\tan \beta$ | $s/h$ | $\theta$ |
|----------------|-----|-----|--------------|------|--------|
| Values         | 1.38| 0.16| 2.7          | 0.086| 85.94  |

Table 5: RF-AGA-ENN model-derived PIM parameters of working face 11111.
A reduction method based on weighted least squares polynomial fitting of discrete data, which uses a robust fitting process to prevent deviations from distorting smooth data points. In this paper, RW is used to perform noise reduction on the measured data, and the noise reduction processed data and the measured data are shown in Figures 5(a)–5(j).

From Figures 5(a) to 5(j), it can be seen that the up and down fluctuation range of the measured data is significantly reduced after the RW noise reduction treatment, and the data curve is smoother after the treatment. Noise reduction was applied to some data points with large fluctuation ranges to reduce the effects of data acquisition errors and significant differences in matching between the training sample data and the test sample data.

In mining subsidence, intelligent optimization algorithms are often used to learn data from similar geological

![Figure 9: Comparison of predicted and measured values.](image)
Figure 10: The value of PIM parameters obtained by different methods.
independent models are established with each of the parameters, this paper adopts a many-to-one network structure to improve the accuracy of solving PIM parameters for the intelligent optimization algorithm.

To simplify the ENN network structure, the relevant parameters of the model are selected as follows: training times 5000, learning rate 0.01, and training error 0.000001. The number of neurons in the hidden layer is \( k = \sqrt{m + n + a} \), where \( m \) is the number of input neurons, \( n \) is the number of output neurons, and \( a \) is a constant, taking values from 0 to 10 and traversing the constant \( a \) value to determine the optimal grid; at this point, the corresponding value is the value of \( a \). The tansig function is used as the transfer function for the input and hidden layers, the purelin function is used as the transfer function for the hidden and output layers, and the trainlm function is used as the training function for the grid. In the ACO algorithm, the initial population size is 30; the maximum evolutionary generation is 50; the volatility coefficient of pheromone \( \rho \) is 0.9. The importance coefficient \( \alpha \) is 0.3; the importance coefficient of heuristic \( \beta \) is 0.5. In the GA algorithm, the population is 20, the maximum evolutionary generation is 20, the hybridization probability \( P_c \) is 0.8, and the mutation probability \( P_m \) is 0.2.

The data sets 1-50 in Table 1 are used as the training set, the data sets 51-70 are used as the test set, and the prediction results are shown in Figures 6(a)–6(e) using the RF-ENN model and the RF-AGA-ENN model, respectively.

From the predicted results (Figures 6(a)–6(e)), it can be intuitively seen that the optimized ENN network of the AGA algorithm is more adaptable to the intrinsic connection between geological mining conditions and PIM parameters. The predicted results of the AGA optimized ENN grid are in better agreement with the true values of experimental data.

To quantitatively describe the ENN network structure optimized using the AGA algorithm is expected to have better adaptability to PIM parameters. Considering that the values of the five parameters of PIM vary greatly, it is difficult to evaluate the prediction accuracy of a single measure, so MAXRE, AVERE, and RMSE indicators can be used to comprehensively evaluate the accuracy of the prediction results of PIM parameters. The accuracy index values of the RF-AGA-ENN model and RF-ENN model for PIM parameter estimation are calculated separately, and three values of maximum relative error (MAXRE), average relative error (AVERE), and root mean square error (RMSE) are used in this thesis as the indexes to evaluate the accuracy of the estimation results. The accuracy index values of the predicted results are shown in Table 4.

### Table 4: RF feature selection for input layer variables.

| Type | \( M \) | \( a \) | \( h \) | \( h/M \) | \( A \) | \( D_i/h \) | \( V \) | \( h_i/h \) | \( R_i \) |
|------|--------|--------|--------|---------|------|--------|------|--------|-------|
| Type RF | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 1 | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 2 | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 3 | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type RF | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 1 | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 2 | √      | √      | √      | √       | √    | √      | √    | √      | √     |
| Type 3 | √      | √      | √      | √       | √    | √      | √    | √      | √     |

### 4. Engineering Application

The “three zones” theory is commonly used to explain the subsidence of landmarks caused by mining. The collapse zone, crack zone, and bending zone are from bottom to top. A diagram of the “three zones” theory is shown in Figure 7. The PIM parameters of similar mines are often used as the empirical parameters of this mine in the production practice process. This paper proposes the RF-AGA-ENN model to obtain the PIM parameters. This paper verifies the feasibility of the RF-AGA-ENN model to obtain...
Figure 11: PIM parameter prediction results under different input layer conditions.
The values of L02 and L03 near the railway sign and the protective pillars are left during the coal mining process. The existence of L02 and L03 is near the railroad, and the protective pillars provide safety for the railway line. Therefore, we use the RF-AGA-ENN model to predict the subsidence values of the monitoring points. The predicted values are shown in Figure 5.

The geological mining conditions of each model input layer were determined according to Table 6, and the AGA-ENN model was used for the prediction of PIM parameters. The predicted results are shown in Figure 11 and Table 7. From Figure 11 and Table 7, it can be seen that there is some variability in the PIM parameters solved by the AGA-ENN model under different input layer conditions, among which the PIM parameters are obtained by using the optimal feature set obtained by RF as the input layer of the model is closest to the true values (the values of its accuracy assessment indexes MAXRE, AVERE, and RMSE are smaller), indicating that the optimal feature set selected by RF can simplify the ENN network complexity to improve the prediction accuracy. This also shows the indispensable role of RF feature selection.

5. Discussion

5.1. Comparison with BP Neural Network for PIM Parameters. The accuracy of obtaining PIM parameters directly affects the accuracy of subsidence prediction. Since PIM parameters are closely related to geological mining conditions. With the development of science and technology, machine learning has become the primary method of obtaining PIM parameters. In the field of mining subsidence, commonly used machine learning methods include the RF-AGA-ENN model proposed in this paper to obtain PIM parameters, the RF-AGA-ENN model and BP neural network are used to obtain PIM parameters, respectively. The obtained subsidence factor and horizontal movement coefficient are shown in Figure 10.

Table 7: Accuracy index values of PIM parameters under different input layer conditions.

| PIM parameters | Accuracy index | Type 1 | Type 2 | Type 3 |
|----------------|----------------|--------|--------|--------|
| MAXRE          | 4.9            | 8.71   | 7.58   | 8.53   |
| AVERE          | 2.44           | 3.68   | 3.39   | 3.90   |
| RMSE           | 0.0241         | 0.0401 | 0.0412 | 0.0431 |
| MAXRE          | 5.04           | 6.67   | 6.95   | 7      |
| AVERE          | 2.03           | 2.87   | 2.99   | 3.02   |
| RMSE           | 0.0069         | 0.0098 | 0.0091 | 0.0099 |

PIM parameters by using a mining face 11111 in Anhui Province, China, as an example.

The main information on the working face of the study area is as follows: The overburden lithology is 33.5 MPa, and the working face advances at a rate of 86 m per month. The 11111 working face has an inclination length of 145 m and a strike length of 411 m. The average mining depth of the working face is 394 m, the thickness of the loose seam is 336 m, the inclination of the coal seam is 6°, and the mining thickness is 4.8 m. Half of the strike observation line and half of the inclination observation line are set on the surface above the 11111 working face. There are 24 monitoring points in the strike observation line and 26 monitoring points in the incline direction. The diagram of monitoring points on the working face is shown in Figure 8.

The information of the 11111 working face is used as the input layer of the RF-AGA-ENN model to obtain PIM parameters. Then, the prediction of the subsidence value of working face 11111 was carried out based on the obtained PIM parameters. The prediction of the subsidence value of 11111 working face is carried out according to the obtained PIM parameters (as shown in Table 5). The predicted values and the measured values are shown in Figure 5.

As seen from Figure 9, the predicted and measured values of the monitoring points are in good agreement except for the strike monitoring points L02 and L03. After the actual field research, we know that monitoring points L02 and L03 are located near the railroad, and protective coal pillars are left during the coal mining process. The existence of the protective coal pillars makes the predicted values of L02 and L03 near the railroad significantly larger than the measured values. Therefore, we use the RF-AGA-ENN model to find the PIM parameters. Using the obtained PIM parameters for engineering practice should be combined with the actual site conditions to make a comprehensive judgment. In summary, the RF-AGA-ENN model for PIM parameters has good engineering application value.

5.2. The Role of RF Feature Selection. To verify the role of RF feature selection, five geological mining factors were selected as input layers, and the AGA-ENN model was used for PIM parameter prediction. Among them, type RF is the control group, and the selected geological mining factors are the most optimal feature set consisting of the five factors with the highest importance selected by RF. Type 1, type 2, and type 3 are the experimental groups. The selected geological mining factors are the nonoptimal feature set. The subsidence and horizontal movement factors were selected as the output layer and modeled separately. The specific geological mining factors selected for each projected model are shown in Table 6.

The geological mining conditions of each model input layer were determined according to Table 6, and the AGA-ENN model was used for the prediction of PIM parameters. The predicted results are shown in Figure 11 and Table 7. From Figure 11 and Table 7, it can be seen that there is some variability in the PIM parameters solved by the AGA-ENN model under different input layer conditions, among which the PIM parameters are obtained by using the optimal feature set obtained by RF as the input layer of the model is closest to the true values (the values of its accuracy assessment indexes MAXRE, AVERE, and RMSE are smaller), indicating that the optimal feature set selected by RF can simplify the ENN network complexity to improve the prediction accuracy. This also shows the indispensable role of RF feature selection.

5.3. The Role of the AGA Algorithm in ENN Networks. To verify the role of the AGA algorithm in the ENN network, the ACO algorithm, the GA algorithm, and the AGA algorithm are used to optimize the ENN network, respectively, and the relevant algorithm parameters are used with the parameter values described previously. The prediction of the subsidence factor and horizontal movement coefficient, the prediction results, and the prediction accuracy are shown in Figure 12 and Table 8. From Figure 12 and Table 8, it can
be seen that the sink factor and horizontal movement coefficient predicted by the RF-AGA-ENN model are closer to the true values, and the prediction accuracy index values of the RF-AGA-ENN model are significantly smaller than those of the RF-ACO-ENN model and the RF-GA-ENN model.

The RF-AGA and RF-AGA-ENN models were used to obtain the PIM parameters, respectively, and the number of iterations of each model is shown in Table 9.

As shown in Table 9, the number of iterations for the RF-ENN model to obtain the PIM parameters fluctuates between 244 and 319, and the number of iterations for the RF-AGA-ENN model to obtain the PIM parameters fluctuates between 107 and 135. The computational efficiency of the RF-AGA-ENN model to obtain the PIM parameters is significantly improved compared to the RF-ENN model.

6. Conclusions

Determining PIM parameters has always been an issue of great concern to scholars engaged in mining subsidence research. To solve the problem that the low accuracy of ENN solving PIM parameters is difficult to meet the needs of production practice, this paper proposes the RF-AGA-ENN model for PIM parameter prediction, which first uses RF to obtain the optimal set of special features as the input layer. It then uses the ENN network optimized by the AGA algorithm for the prediction of PIM parameters. Finally, the feasibility of the RF-AGA-ENN model to obtain PIM parameters is verified by engineering applications. The main findings are as follows:

(1) Seventy sets of measured geological mining conditions and PIM parameter data of coal mining face were used as experimental data. The input layer of the ENN network was optimized by RW smoothing, noise reduction processing, and RF feature selection. The prediction results significantly improved the prediction accuracy of the optimized ENN network.

(2) The ACO and GA algorithms optimize the weights and thresholds of the ENN network, the RF algorithm is used to simplify the complexity of the

Table 8: Accuracy index values of PIM parameters under different input layer conditions.

| PIM parameters | Accuracy index | RF-ACO-ENN | RF-GA-ENN | RF-AGA-ENN |
|----------------|----------------|------------|-----------|------------|
| q MAXRE        | 7.92           | 7.47       | 4.9       |
| q AVERAGE      | 3.38           | 3.55       | 2.44      |
| q RMSE         | 0.0355         | 0.0392     | 0.0241    |
| b MAXRE        | 5.71           | 6.57       | 5.04      |
| b AVERAGE      | 2.95           | 3.27       | 2.03      |
| b RMSE         | 0.0091         | 0.0096     | 0.0069    |

Table 9: The number of iterations of PIM parameters model solving.

| Predictive models | q | b | tan β | s/h | θ |
|-------------------|---|---|-------|-----|---|
| RF-ENN            | 319| 272| 281   | 278 | 244|
| RF-AGA-ENN        | 110| 121| 135   | 129 | 107|
ENN network, and the RF-AGA-ENN model is established for the prediction of PIM parameters. The predicted PIM parameters’ accuracy indicators were as follows: MAXRE values between 1.94% and 9.18, AVERAGE values between 0.98% and 3.98, and RMSE values between 0.0050 and 0.9632.

(3) The PIM parameters derived from the RF-AGA-ENN model were used for subsidence prediction at the 11111 working face of a mine in Anhui Province. The results show that the model has good application value. However, the application of the PIM parameters from the RF-AGA-ENN model for subsidence prediction should be analyzed on a case-by-case basis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research work was supported by the National Natural Science Foundation of China (41472323 and 51404272) and the Foreign Science and Technology Cooperation Program of Anhui Province (201904311020015). All those financial supports are gratefully acknowledged. In addition, I would like to thank Dr. Chuanwu Zhao for his great help in the revision process of the thesis. We would like to thank Dr. Weitao Yan for his answers during the revision of the paper. Thanks to Dr. Zhou Bang for his help during the code debugging process.

Supplementary Materials

Table 1: experimental data. (Supplementary Materials)

References

[1] H. P. Xie, L. X. Wu, and D. Z. Zheng, “Prediction on the energy consumption and coal demand of China in 2025,” Journal of China Coal Society, vol. 44, no. 7, pp. 1949–1960, 2019.

[2] L. Yuan, Y. D. Jiang, K. Wang, Y. X. Zhao, X. J. Hao, and C. Xu, “Precision exploitation and utilization of closed/abandoned mine resources in China,” Journal of China Coal Society, vol. 43, no. 1, pp. 14–20, 2018.

[3] Y. L. Bi, L. Xiao, C. Guo, and P. Christie, “Revegetation type drives rhizosphere arbuscular mycorrhizal fungi and soil organic carbon fractions in the mining subsidence area of northwest China,” Catena, vol. 195, article 104791, 2020.

[4] J. H. Dong, L. R. Meng, Z. F. Bian, and A. M. Fang, “Investigating the characteristics, evolution and restoration modes of mining area ecosystems,” Polish Journal of Environmental Studies, vol. 28, no. 5, pp. 3539–3549, 2019.

[5] X. M. Cui, Y. G. Gao, and D. B. Yuan, “Sudden surface collapse disasters caused by shallow partial mining in Datong coalfield, China,” Polish Journal of Environmental Studies, vol. 74, no. 2, pp. 911–929, 2014.

[6] H. J. Yin, G. L. Guo, H. Z. Li, T. N. Wang, and Y. F. Yuan, “Prediction method and research on characteristics of surface subsidence due to mining deeply buried Jurassic coal seams,” Bulletin of Engineering Geology and the Environment, vol. 81, no. 10, p. 449, 2022.

[7] Z. F. Yang, Z. W. Li, J. J. Zhu, H. W. Yi, J. Hu, and G. C. Feng, “Deriving dynamic subsidence of coal mining areas using InSAR and logistic model,” Remote Sensing, vol. 9, no. 2, pp. 125–144, 2017.

[8] J. Roser, D. Potocnik, and M. Vulic, “Analysis of dynamic surface subsidence at the underground coal mining site in Velenje, Slovenia through modified sigmoidal function,” Minerals, vol. 8, no. 2, p. 74, 2018.

[9] B. Zhang, J. C. Ye, Z. J. Zhang, L. Xu, and N. X. Xu, “A comprehensive method for subsidence prediction on two-seam longwall mining,” Energies, vol. 12, no. 16, p. 3139, 2019.

[10] X. J. Zhu, G. L. Guo, H. Liu, and X. Y. Yang, “Surface subsidence prediction method of backfill-strip mining in coal mining,” Bulletin of Engineering Geology and the Environment, vol. 78, no. 8, pp. 6235–6248, 2019.

[11] D. W. Zhou, K. Wu, X. X. Miao, and L. Li, “Combined prediction model for mining subsidence in coal mining areas covered with thick alluvial soil layer,” Bulletin of Engineering Geology and the Environment, vol. 77, no. 1, pp. 283–304, 2018.

[12] X. G. Jia, “Calculation parameters inversion of mining subsidence based on ant colony algorithm,” Industry and Mine Automation, vol. 41, no. 6, pp. 10–13, 2015.

[13] L. Wang, X. N. Zhang, and S. S. Chi, “Study of probability integration method parameter inversion by the genetic algorithm,” International Journal of Mining Science and Technology, vol. 27, no. 6, pp. 1073–1079, 2017.

[14] P. X. Li, D. Peng, Z. X. Tan, and K. Z. Deng, “Parameters inversion of probability integral method based on multi-population genetic algorithm,” Geomatics and Information Science of Wuhan University, vol. 51, no. 11, pp. 50–54, 2021.

[15] P. X. Li, Z. X. Tan, L. L. Yan, and K. Z. Deng, “Time series forecasting of probability integration method parameters based on support vector machine,” Journal of China Coal Society, vol. 36, no. S2, pp. 380–385, 2011.

[16] Z. M. Zhao, T. W. Shi, W. Dong, and Y. L. Liu, “The prediction of probability-integral method parameters based on grey relational analysis and BP neural network,” Science of surveying and Mapping, vol. 42, no. 7, pp. 36–40, 2017.

[17] Z. Shen, L. J. Xu, X. P. Liu, C. C. Qin, and Z. B. Wang, “Probability integral method parameters prediction model optimization based on machine learning methods,” Bulletin of Surveying and Mapping, vol. 10, pp. 35–380, 2016.

[18] State Bureau of Coal Industry, Buildings, Water Bodies, Railroads and Main Shaft Coal Pillar Retention and Coal Compression Mining Regulations, Coal Industry Press, Beijing, 2000.

[19] W. M. Zhou, H. X. Zhang, and W. L. Zhong, “Study of getting surface subsidence parameters under the effect of multi working face mining,” Journal of China Coal Society, vol. 25, no. 3, pp. 260–263, 2000.

[20] G. L. Guo and Y. J. Wang, “Study of robust determining parameters model for probability-integral method and its
application,” Acta Geodaetica et Cartographica Sinica, vol. 29, no. 2, pp. 162–165+171, 2000.

[21] P. X. Li, Z. X. Tan, and K. Z. Deng, “Related factors analysis of calculation parameters of probability integral method for surface movement,” Journal of Mining And Strata Control Engineering, vol. 16, no. 6, pp. 14-15+5, 2011.

[22] W. B. Guo, K. Z. Deng, and Y. F. Zou, “Artificial neural network model for predicting parameters of probability-integral method,” Journal of China University of Mining and Technology, vol. 33, no. 3, pp. 322–326, 2004.

[23] B. C. Zhao, Z. X. Tan, and K. Z. Deng, “Calculation of surface subsidence factor based on random forest,” Journal of Hefei University of Technology(Natural Science), vol. 39, no. 8, pp. 1123–1126, 2016.

[24] F. X. Wang, Z. X. Tan, and K. Z. Deng, “Study on prediction method of surface subsidence coefficient based on GA-GRNN,” Coal Engineering, vol. 46, no. 7, pp. 94–96, 2014.

[25] G. Y. Li, Prediction of Mining Surface Settlement Based on Ant Colony Optimization Algorithm for Support Vector Machine, Jiang Xi University of Science and Technology, 2015.

[26] S. S. Chi, X. X. Yu, and L. Wang, “Calculation method of probability integration method parameters based on MIV-GP-BP model,” Technical Gazette, vol. 28, no. 1, pp. 160–168, 2021.

[27] Y. Wang and Y. N. Jiang, “Gray Elman neural network prediction model based on genetic algorithm,” The Practice and Understanding of Mathematics, vol. 47, no. 22, pp. 315–320, 2017.

[28] Z. G. Wang and C. X. Zhao, “Task scheduling of fog computing based on genetic ant colony algorithm,” Computer Applications and Software, vol. 38, no. 8, pp. 291–297, 2021.

[29] C. Zhang and S. Jiang, “Intelligent dispatching of unmanned truck in open pit mine based on improved ant colony algorithm,” Journal of Anhui University of Technology(Natural Science), vol. 37, no. 3, pp. 267–275, 2020.

[30] Q. Xiong, Slope Displacement Prediction Based on Genetic Ant Colony Combination Algorithm for Optimizing SVM Model, Jiangxi University of Science and Technology, 2019.

[31] L. Shi, X. C. Li, W. Ren, and Z. M. Fang, “Hybrid of ant colony algorithm and genetic algorithm and its application to searching critical slope slip surface,” Rock and Soil Mechanics, vol. 30, no. 11, pp. 3486–3492, 2009.

[32] L. Shi, X. C. Li, G. H. Sun, and B. Bai, “Research on stability of open-mining slope under underground mining,” Rock and Soil Mechanics, vol. 33, no. 3, pp. 812–820, 2012.

[33] Y. M. Liu, Y. Li, and Z. Y. Zhao, “RF-LSTM model based on feature selection for component stock price trend prediction,” Statistics & Decisions, vol. 37, no. 1, pp. 157–160, 2021.

[34] State Administration of Work Safety, Guide to Coal Pillar Retention and Coal Compression Mining for Buildings, Water Bodies, Railroads and Major Shafts, Coal Industry Press, Beijing, 2017.