

Article

# A New Method for Calculating Prediction Parameters of Surface Deformation in the Mining Area

Shenshen Chi <sup>1,2,\*</sup>, Lei Wang <sup>2,\*</sup> and Xuexiang Yu <sup>2</sup>

<sup>1</sup> Coal Industry Engineering Research Center of Mining Area Environmental and Disaster Cooperative Monitoring, Anhui University of Science and Technology, Huainan 232001, China

<sup>2</sup> School of Geomatics, Anhui University of Science and Technology, Huainan 232001, China; xxyu@aust.edu.cn

\* Correspondence: austcss@163.com (S.C.); lwang@aust.edu.cn (L.W.)

**Abstract:** The accurate calculation of mining-induced surface deformation has important guiding significance for efficient and safe production in mining areas. The probability integral method (PIM) is a main prediction method in China, and the selection of its parameters is directly related to the prediction accuracy of surface deformation in mining areas. To overcome shortcomings of PIM and other methods, this paper proposed a prediction model of the parameters of PIM combining a multiple regression model and an extreme learning machine. In this paper, the Huainan mining area was selected as the research object, the influence factors of PIM parameters were analyzed and the accuracy of the model was verified. The influence of the number of hidden layer nodes, the selection of activation function and the proportion of training set and test set in the model were analyzed. The conclusions suggest that the PIM parameters calculated in this paper could be used to predict mining subsidence and obtain surface movement and deformation data. The research results provide an effective method for the selection of surface deformation prediction parameters of new working faces or faces lacking measured data.

**Keywords:** mining subsidence; PIM; surface deformation prediction; multiple regression model; ELM



**Citation:** Chi, S.; Wang, L.; Yu, X. A New Method for Calculating Prediction Parameters of Surface Deformation in the Mining Area. *Appl. Sci.* **2023**, *13*, 8030. <https://doi.org/10.3390/app13148030>

Academic Editor: Nikolaos Koukoulas

Received: 8 June 2023

Revised: 3 July 2023

Accepted: 6 July 2023

Published: 10 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mining-induced surface deformation is a worldwide issue [1–4]. After coal mining, the initial stress state of rock in the goaf will be destroyed, and the overlying strata will reach a new equilibrium state through bending, caving, spalling and other forms. In this process, the surface will deform and eventually form collapse pits, which will affect surface water circulation and destroy the surface ecological environment of the mining area [5–8]. Therefore, accurately predicting surface movement and deformation is meaningful to the exploitation of coal resources and can support rational relocation, restoration and rebuilding of the ecological environment in mining areas [9,10].

At present, the prediction methods of mining-induced surface deformation mainly include the empirical method, the theoretical model method and the influence function method based on measured data [11–14]. Among them, the PIM has high prediction accuracy and simple prediction degree, and has been widely used in China [15–17]. According to the literature [18–20], the prediction error of the PIM mainly comes from the model error and calculation parameter error. Model error mainly comes from the disparity between assumptions derived by the PIM and complex geological and mining conditions, and it is often difficult to improve. Therefore, an appropriate method is used to determine the uncertainties in the model parameters of the deformation, which are critical.

The prediction errors of the PIM mainly come from model errors and parameter errors. Model errors mainly come from the inconsistency between the assumptions of the PIM and the actual complex geological mining conditions, which are often difficult to improve. Therefore, reducing parameter errors of PIM is an important means to improve the prediction accuracy.

The model parameters of PIM can be obtained by data-based methods and analogy methods [21,22]. The former use the periodic monitoring data of surface deformation to invert the surface prediction parameters and predict the surface deformation. These methods are of high accuracy and reliability. Kwinta [23] proposed the least squares method based on the measured data to obtain the parameters of the predicted model through the linearization of the vertical deformation. Wang et al. [24] used the improved fireworks algorithm to invert deformation parameters for the observation data of a mine in Huainan over 20 months. However, the measured parameters have a lag in guiding the prediction of the observed working face. Generally, the observation time is long (2–3 years), which needs substantial energy to construct and maintain the monitoring station, and it is impossible to effectively guide the new mining face. The analogy methods are to select statistical analysis results of parameters to directly determine parameters. The most common methods are the statistical model method and the neural network method. Hejmanowski [25] used orthogonal analysis to obtain the horizontal movement value and verified it. Wang et al. [26] used the measured data to comprehensively study the methods of analogy theory and fuzzy recognition theory, and a comprehensive analogy method for surface prediction parameters suitable for the Chongqing mining area was finally proposed. Polanin [27] obtained the estimated surface parameter values based on the measured deformation values and geological and mining conditions. The essence of the statistical model is a more intuitive empirical model, simple and intuitive. It is commonly used to summarize the characteristics of data rules. However, it considers fewer influencing factors. In some cases, it cannot obtain satisfactory fitting results, and it lacks generalization ability [28]. As for the neural network method, although it has high accuracy, it often requires more monitoring data, and the model is prone to falling into local minima, affecting fitting accuracy [29].

In this paper, accuracy of regression analysis and easy-to-fall-into-local-extrema neural networks are enhanced, taking the Huainan mining area as an example. The influencing factors of the parameters of PIM will be analyzed. On this basis, the relationships between actual conditions and parameters of the PIM are established. The ELM neural network method is used to predict the residuals of the statistical model, and then a prediction model of the relevant parameters of the surface movement basin is constructed by combining the multiple regression model and ELM. The accuracy and reliability of the model are verified.

Section 2 presents an overview of the study area. Section 3 introduces the main methods of fusion of multiple regression model (M) and ELM, and discusses the application steps of the M-CM-GA-ELM integrated method. Section 4 analyzes reliability of the model. Section 5 verifies accuracy of parameter estimation with examples. Section 6 discusses the parameters of the ELM neural network, and Section 7 draws the conclusions and prospects.

## 2. Overview of Huainan Mining Area

The Huainan mining area is at  $115^{\circ}50' \sim 117^{\circ}45'$  E longitude and  $32^{\circ}25' \sim 33^{\circ}10'$  N latitude. It is mainly distributed in the north of  $27^{\circ}40'$  N latitude. It starts from the Tanlu fault zone in the east, ends at Fuyang fault in the west, connects with Minglong Mountain and Shangyao in the north, and Shungengshan and Bagong Mountain in the south. It adjoins with Huainan, east of Fuyang, Bozhou and other counties and cities. The mining area is about 180 km long in EW direction, 15–25 km wide in N-S direction and has a total area of about 3600 km<sup>2</sup> [30–32].

The Huainan mining area lies in the Huaihe alluvial plain; the ground elevation is about +20~+30 m. It is divided into the old mining area on the south bank of the Huaihe River and the Panxie mining area on the north bank of the Huaihe River. The mines in the old area are distributed in the Huainan arc-shaped tectonic belt protruding southward from the Bagongshan area in the south wing of the Huainan synclinal structure. There are many faults, the dip angle of the coalfield is generally larger than  $20^{\circ}$ , and the geological and mining conditions are relatively poor. The main mine field of the Panxie mining area is located in the south wing of Chenqiao anticline and Panji anticline in the Huainan synclinal structure. The coalfield has a small dip angle of generally less than  $15^{\circ}$ , simple



deformation by the superposition principle. It was proposed by Liu et al. [33] and has been extensively used in China. The qualitative random model of granular medium is shown in Figure 2. The prediction formula of unit basin movement and deformation in unit mining could be obtained through the random medium theory, and the prediction formula of deformation at any point on the surface could be further deduced. There are five parameters in the PIM parameter model, namely subsidence coefficient  $q$ , horizontal movement coefficient  $b$ , tangent of major influence angle  $\tan\beta$ , offset of inflection point  $s$  and propagation angle of extraction  $\theta_0$ .

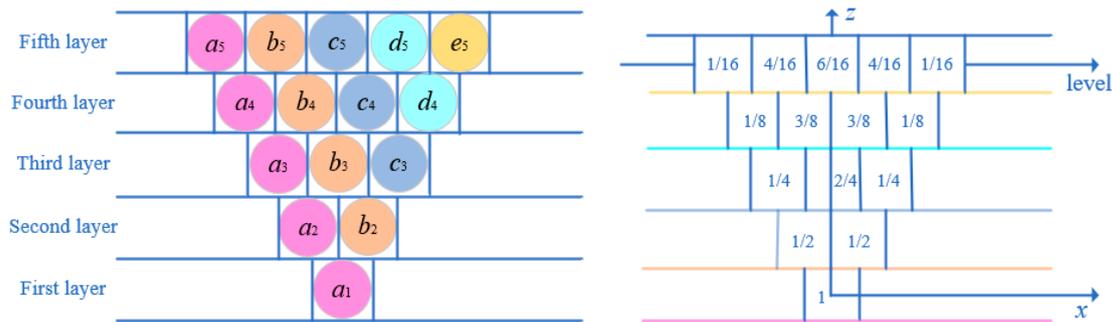


Figure 2. Random model of particle medium movement.

### 3.2. Analysis of Influencing Factors of PIM in Huainan Mining Area

Parameters of PIM are mainly affected by geological and mining conditions such as the structure and lithology of the overlying rock stratum, stratum dip angle, coal mining method and roof management method. These specific factors include average mining depth ( $H_0$ ), the thickness of the loose layer ( $h_s$ ), coal seam dip angle ( $\alpha$ ), etc. If the relationship between the geological mining conditions and parameters of PIM can be established according to the existing observation data, parameters of PIM can be obtained with unknown geological mining conditions. Then, the range of surface mining can be obtained before mining. The Huainan mining area is taken as an example for the following research. It was found that under the corresponding geological and mining conditions in the Huainan mining area, the offset of the inflection point generally fluctuated within a certain range without obvious rules. Relevant data show that its value is generally maintained at 0.1 times the mining depth to meet the requirements. Therefore, this paper ignores it.

#### 3.2.1. Subsidence Coefficient

Many studies have proved that the subsidence coefficient can be affected by the mining height, the treatment method of the goaf and the nature of the overlying strata. Based on the surface subsidence coefficients, mining depths and loose layer thickness of some working faces in the Huainan mining area, the regression formula between subsidence coefficient and bedrock and mining depth ratio  $(H_0 - h_s)/H_0$  can be obtained through regression analysis, As shown in Equation (1) and Figure 3,

$$q = -1.0305(H_0 - h_s)/H_0 + 1.3489 \tag{1}$$

Goodness of fit  $R^2 = 0.7744$ , showing that the subsidence coefficient is closely related to the bedrock and mining depth and decreases with the increase in the proportion of bedrock in the mining depth.

#### 3.2.2. Horizontal Movement Coefficient

The horizontal movement coefficient refers to the ratio between the maximum horizontal surface movement value and the maximum surface subsidence value under the critical mining of horizontal or nearly horizontal coal seams. It mainly depends on the dip angle of the coal seam and the thickness of the loose layer. Based on the collected data,

mining depth and the thickness of the loose layer of some typical working faces in the Huainan mining area, the regression formula between horizontal movement coefficient  $b$  and loose layer and mining depth ratio  $h_s/H_0$  could be expressed as Equation (2):

$$b = -0.2521h_s/H_0 + 0.4785 \tag{2}$$

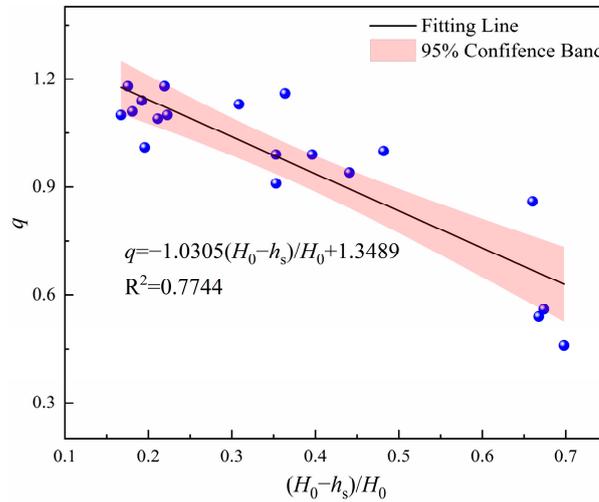


Figure 3. Fitting curve of subsidence coefficient and  $(H_0-h_s)/H_0$ .

The fitting results are shown in Figure 4, The goodness of fit  $R^2 = 0.7318$ . The fitting results show that the horizontal movement coefficient is closely related to thickness of unconsolidated layer and mining depth, and decreases with increase of the proportion of unconsolidated layer in the mining depth.

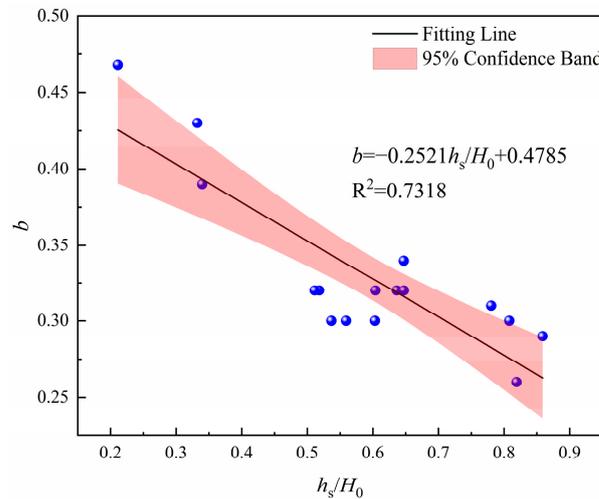


Figure 4. Fitting curve of horizontal movement coefficient and  $h_s/H_0$ .

### 3.2.3. Tangent of Major Influence Angle

The main influence angle tangent is used to represent the major influence radius at different mining depths, which is generally expressed as the ratio of mining depth to major influence radius. It is related to the roof and overlying strata, mining depth and thickness. According to the collected information in the Huainan mining area, the regression formula between the tangent of major influence angle and the unconsolidated layer thickness could be written as Equation (3):

$$\tan \beta = 0.0044h_s + 0.0932 \tag{3}$$

The fitting results are shown in Figure 5, The goodness of fit  $R^2 = 0.6546$ , showing that the size of the tangent of major influence angle is proportional to the thickness of the unconsolidated layer.

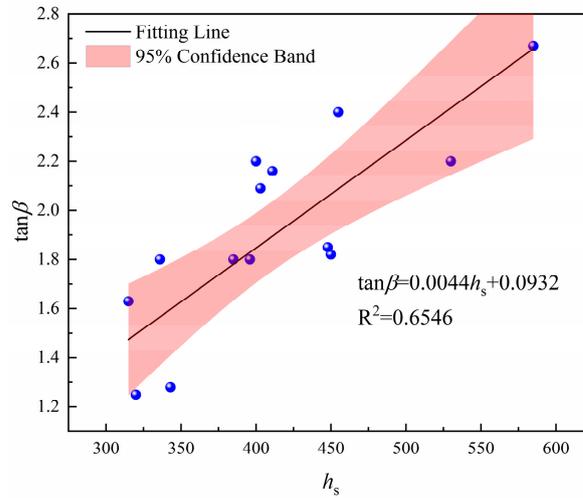


Figure 5. Fitting curve of tangent of major influence angle and thickness of unconsolidated layer.

### 3.2.4. Propagation Angle of Extraction

The propagation angle of extraction is an important parameter in the prediction model of the probability integration method, and its size is affected by the dip angle of the coal seam, which determines the degree to which the surface mobile basin deviates from the goaf. According to collected data, the regression formula between  $\theta_0$  and  $\alpha$  could be expressed as Equation (4) through regression analysis in Figure 6:

$$\theta_0 = -0.5927\alpha + 90.566 \tag{4}$$

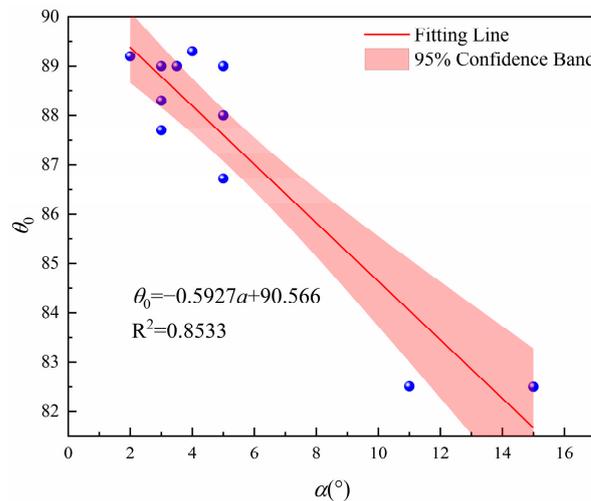


Figure 6. Fitting curve of propagation angle and coal seam dip angle.

The goodness of fit  $R^2 = 0.8533$ . The fitting results show that the propagation angle of extraction decreases with the increase in dip angle.

### 3.3. Construction of Combined Model

#### 3.3.1. ELM Neural Network

The ELM neural network is a supervised machine learning algorithm, a single hidden layer feed-forward neural network which overcomes the defect that the BP neural network gradient algorithm is prone to falling into local extrema, and has the characteristics of high

classification and regression efficiency and easy programming [34,35]. Based on advantages of the ELM neural network in regression analysis, the geological and mining conditions can be used as input and parameters of PIM as output to establish a regression analysis model. Figure 7 shows the network structure of the ELM neural network, which is similar to that of the BP neural network. Assuming that there are  $n$  neurons in the input layer,  $m$  neurons in the output layer, and one neuron in the hidden layer, and the connection weight between the input layer and the hidden layer is  $w$ , the mathematical model can be expressed as Equation (5):

$$\sum_{i=1}^l \beta_i g(w_i x_j + b_i) = t_j, j = 1, 2, \dots, n \tag{5}$$

where  $w_i$  is the input weight connecting the  $i$ th hidden layer node and the input layer node;  $\beta_i$  is the output weight connecting the  $i$ th hidden layer node and the output layer node;  $b_i$  is the offset of the  $i$ th hidden layer node;  $x_j$  and  $t_j$  represent the input eigenvector and output eigenvector of the  $j$ th sample, respectively.

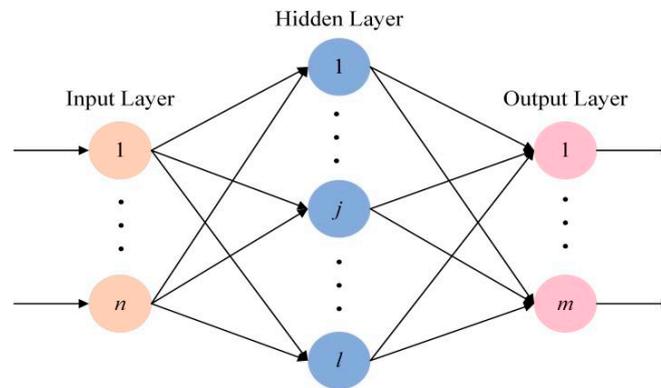


Figure 7. Topological structure of ELM neural network.

Let  $\beta = [\beta_1, \beta_2, \dots, \beta_l]^T, T = [t_1, t_2 \dots t_n]^T$  and

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) & \dots & g(w_l x_1 + b_l) \\ \vdots & \vdots & \vdots \\ g(w_1 x_n + b_1) & \dots & g(w_l x_n + b_l) \end{bmatrix}_{n \times l} \tag{6}$$

Then, Equation (5) could be simplified as Equation (7):

$$H\beta = T \tag{7}$$

where  $H$  is the output matrix of the hidden layer, and the least square solution of the output weight could be easily obtained through the linear Equation (7):

$$\hat{\beta} = H^+ T \tag{8}$$

wherein  $H^+$  is the Moore-Penrose pseudoinverse of the hidden layer output matrix  $H$ .

The above analysis indicates that the network randomly generates weights and thresholds before training, so only the number of hidden layers and activation functions need to be determined, and the output  $\hat{\beta}$  can be calculated.

### 3.3.2. GA-Optimized ELM Neural Network

The ELM neural network has a strong nonlinear fitting ability, but due to the random generation of the weight matrix between the input layer and hidden layer and that of the threshold matrix of hidden layer neurons, the prediction fitting ability and prediction stability of the network model are insufficient. To select a proper ELM simulation model,

this paper proposed a GA-optimized ELM neural network. The specific modeling process is as follows:

1. Determine the relevant parameters of the GA algorithm;
2. Carry out coding and population generation, randomly generate weights and thresholds of the ELM neural network, and generate the original population by binary code;
3. Calculate the fitness function value of each individual, calculate RMSE of each individual's test set, and take it as the individual fitness value;
4. Population evolution;
5. Training and prediction of ELM network: Decode the final population after iterative optimization to obtain the optimized weights and thresholds, and assign to the ELM; train the ELM network by training samples, and calculate output layer weight  $\hat{B}' = H'^+ T'$  by the least squares method; finally, bring the test samples into the ELM model for prediction.

### 3.3.3. Construction of CM-GA-ELM Model

The initial population of the standard GA is randomly generated, and the coverage of the generated population is large, which may lead to the initial population not necessarily containing the optimal solution. If the initial population does not contain the optimal solution and the genetic algorithm cannot cover the global optimal solution in a certain algebra, it is bound to cause premature convergence and affect the regression accuracy of the model. In order to eliminate the outliers caused by model parameters in the GA-ELM prediction model and improve the accuracy and reliability of the prediction model, this paper proposes a linear-weighted combination prediction method (CM), which is simple and can solve the above problems. The solution is as follows:

1. Elimination of abnormal predicted values

Assuming that  $n$  times of GA-ELM prediction models were operated and  $n$  groups of GA-ELM prediction models were established, different solutions were obtained. The  $i$ th prediction value was  $x_i$ , and the average value of  $n$  times of prediction value was  $\bar{x}$ . Equation (9) was used for calculation. To improve the reliability of prediction results, abnormal values in the results were eliminated. First, Equation (10) was used to calculate the mean square error  $R$  of multiple prediction results, and the results greater than 3 times the mean square error were eliminated.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_i + \cdots + x_n}{n} \quad (9)$$

$$R = \pm \sqrt{\frac{[v_i v_i]}{n-1}} \quad (10)$$

where  $v_i = x - x_i$ .

2. Combination of predicted values

Assuming that  $M$  predicted values were excluded, the average of the remaining  $J$  ( $J = n - M$ ) predicted values could be calculated, and the final predicted value could be obtained by Equation (11):

$$\bar{x} = \frac{1}{J} \sum_{i=1}^J x_i \quad (i = 1, 2, \cdots, J) \quad (11)$$

Above all, the CM-GA-ELM combined prediction model could be used to predict the subsidence coefficient of the surface movement basin.

### 3.3.4. Construction of M-CM-GA-ELM Neural Network Integrated Model

Figure 8 shows the calculation process of the M-CM-GA-ELM model, and the following are the specific steps:

1. The multiple linear regression model of the relationship between the relevant parameters of the surface movement basin and conditions of geological and mining (such as mining height, dip, mining degree, advance speed, rock stratum lithology, thickness of unconsolidated layer, etc.) was constructed, and the trend term and residual term of the prediction model were calculated;
2. The GA-ELM neural network prediction model was constructed with geological and mining conditions as the input layer and the residual term as the output layer;
3. Errors of the GA-ELM prediction model obtained from prediction were eliminated, and the CM-GA-ELM prediction model was built;
4. At the same time, the multiple linear regression model and the CM-GA-ELM prediction model were used for prediction, and the final predicted value was obtained.

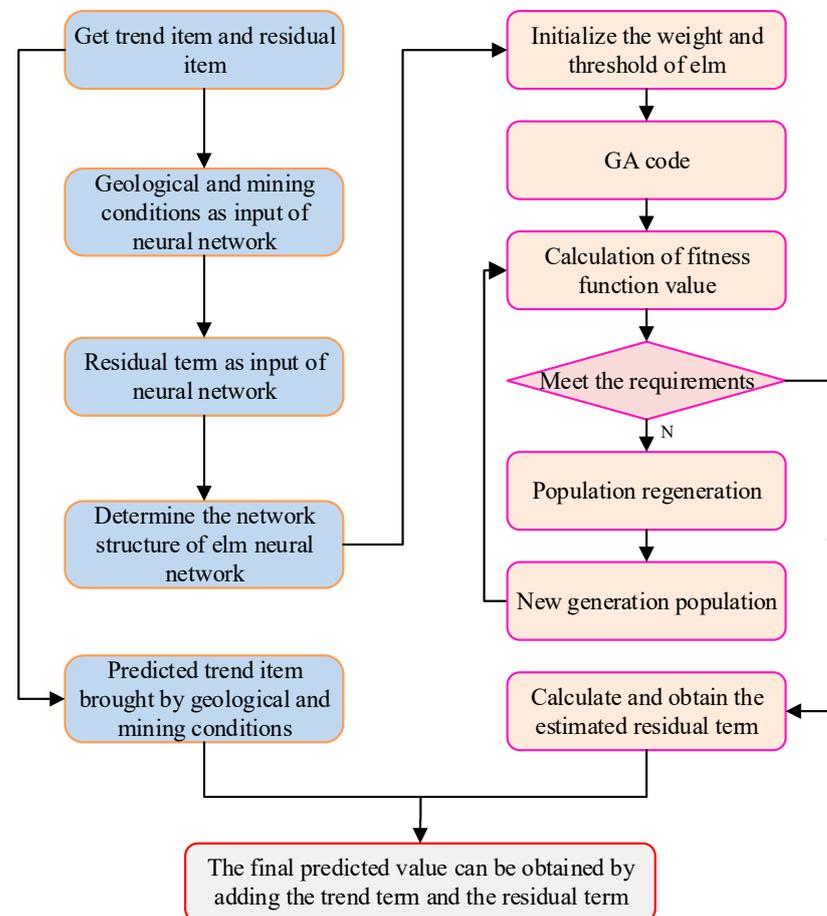


Figure 8. Calculation process of M-CM-GA-ELM model.

In summary, the essence of the fusion model is to use the prediction method of the ELM neural network to compensate errors of the linear regression model. After the error compensation, M-CM-GA-ELM has both the experience of linear regression and the strong nonlinear mapping ability of neural networks.

## 4. Model Verification

To verify the generalization performance of the M-CM-GA-ELM model, the subsidence coefficient  $q$  was taken as an example for analysis. There were few measured data (nearly 20) collected in the Huainan mining area, so it was difficult to meet the data requirements of

the neural network. In this paper, some measured data were collected from Reference [36]. In addition, there are 70 sets of subsidence coefficient values from observation stations in the Huainan mining area. According to the analysis of Section 3.2, the factors affecting the subsidence are mining depth, thickness of bedrock and thickness of unconsolidated layer. Similar to Section 3.2, this paper constructed a multivariate linear fitting model based on 70 sets of data from observation stations. The fitting results are shown in Figure 9.

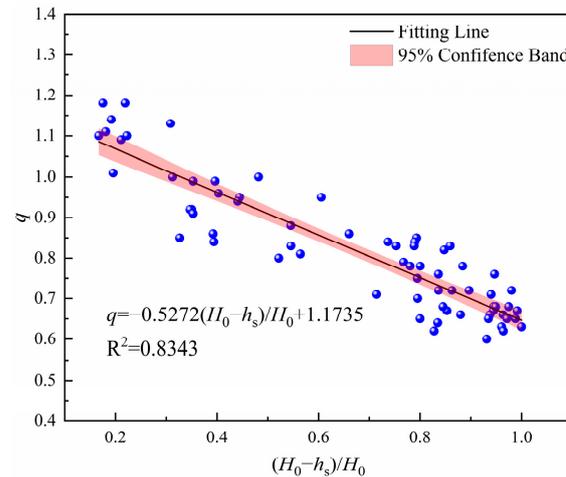


Figure 9. Relationship between subsidence coefficient and mining depth ratio of bedrock.

In order to establish the fusion model, 70 groups of data were divided into two groups in this paper. Eighty percent of the measured data (56 groups) were used for modeling, and twenty percent of the data (14 groups) were used to verify the prediction performance of the model.

Relevant parameters of the M-CM-GA-ELM model were set as follows. The number of input layer neurons was 2 ( $H_0$  and  $h_s$ ), the number of hidden layers was 12, the number of output layers was 1, and the activation function adopted the sig function. To reduce the running time, the running times were set to 50 in the CM model. To verify the prediction effect of the M-CM-GA-ELM model, the regression formula of the subsidence coefficient  $q$  with  $H_0$  and  $h_s$  was  $q = -0.5114(H_0 - h_s)/H_0 + 1.174$ , and the CM-GA-ELM model, CM-ELM model, ELM model, and GA-ELM model were applied for comparison. The results are shown in Figure 10.

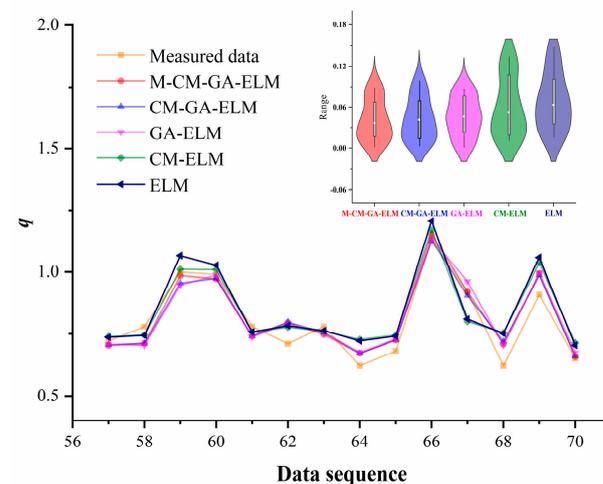


Figure 10. Comparison of prediction results.

In Figure 10, the violin chart is used to compare the accuracy of five prediction models. It is usually used to show the distribution state and probability density of multiple groups of

data, mainly to show the distribution shapes of data. The white dot on the chart represents the median, the black bar represents the range of quartiles, the thin black line represents the 95% confidence interval, and the width of the violin represents the value taking probability of the observed value. The violin chart can show the overall distribution of data in addition to the above statistics. Therefore, Figure 10 can explain different prediction accuracy of the five models. The M-CM-GA-ELM model has the smallest black bar and black line, and the center is around 0, indicating that the error range between the predicted value and the actual value is the smallest. On the other hand, the width of the M-CM-GA-ELM model is relatively wide near 0, where most of the errors are distributed. To sum up, the M-CM-GA-ELM combined model has the highest prediction accuracy.

In order to quantitatively analyze differences between the prediction performance of the improved ELM neural network prediction model and that of the conventional ELM neural network model, evaluation indexes of the prediction of the ELM prediction model, the CM-ELM prediction model, the GA-ELM prediction model and the CM-GA-ELM prediction model were calculated, respectively. In this paper, the average relative error and the RMSE of the prediction results were taken as evaluation indexes:

$$\text{MeaRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x'_i - x_i}{x_i} \right| \times 100\% \tag{12}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x'_i - x_i)^2} \tag{13}$$

where  $x'_i$  is the  $i$ th predicted value;  $x_i$  is the  $i$ th measured value;  $n$  is the number of predicted samples.

Among them, MeaRE represents the stability of the prediction results. The smaller the value, the higher the stability of the prediction model. RMSE can effectively show the degree of deviation between the predicted value of the model and the true value. The smaller the RMSE value, the closer the predicted value will be to the true value. Table 1 shows the accuracy index:

**Table 1.** Accuracy of prediction model (unit: mm).

Prediction Model	M-CM-GA-ELM	CM-GA-ELM	GA-ELM	CM-ELM	ELM
MeaRE	5.509	5.719	6.066	8.131	8.752
RMSE	0.050	0.052	0.054	0.074	0.079

In Figure 10, the M-CM-GA-ELM prediction model has good generalization performance. Although some points have slight fluctuations, the overall prediction accuracy is high. Table 1 shows that the prediction results are M-CM-GA-ELM > CM-GA-ELM > GA-ELM > CM-ELM > ELM. Compared with M-CM-GA-ELM, CM-GA-ELM, GA-ELM and CM-ELM, the GA algorithm has greater advantages in improving accuracy. The M-CM-GA-ELM and CM-GA-ELM precision comparison shows that the precision of the combined model is significantly enhanced.

### 5. Engineering Application

The 1613 (1) working face of the Guqiao Coal Mine in Huainan is flat. The comprehensive mechanized coal mining method was adopted. The average mining height was 2.9 m, the working face was 1528 m long, the width was 251 m and the average mining speed was 5.56 m/d. The average dip angle of coal seam was 3°, which belonged to a nearly horizontal coal seam. The average mining depth was 668 m, and the average thickness of unconsolidated layer was 420 m. Figure 11 shows the layout of the monitoring points, and half a strike line and a tendency line were set up. The surface movement observation time was from 2 May 2017 to 7 January 2019. The subsidence and horizontal movement values above the 1613 (1) working face were obtained by leveling and RTK surveying.

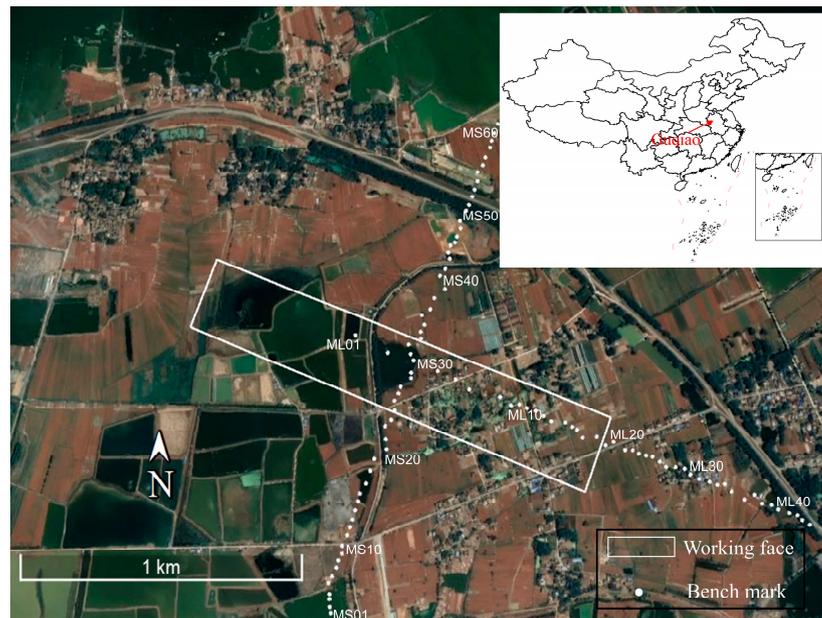


Figure 11. Observation line layout for 1613 (1) working face.

The model constructed in this paper was used to calculate parameter values of the PIM of the 1613 working face, and the results are shown in Table 2. GA was used to calculate the measured PIM of the surface [17]. The values were compared. Based on the absolute errors in Table 2, the difference between the measured PIM parameters and the predicted PIM parameters is small. They were applied to predict mining subsidence, respectively. It can be seen from Figure 12 that the subsidence and horizontal movement values calculated by the predicted parameters are in good agreement with the actual surface deformation values. The actual surface deformation values calculated by the parameters are in good agreement. Therefore, from Table 2 and Figure 12, the model constructed in this paper provides a novel way to predict surface deformation in the Huainan mining area.

Table 2. Comparison of parameters of PIM.

Parameter Values	$q$	$b$	$\tan\beta$	$\theta_0$
Measured parameters	1	0.32	1.76	85
Calculated parameters	0.97	0.31	1.84	87.5
Absolute error	0.03	0.01	0.08	2.5

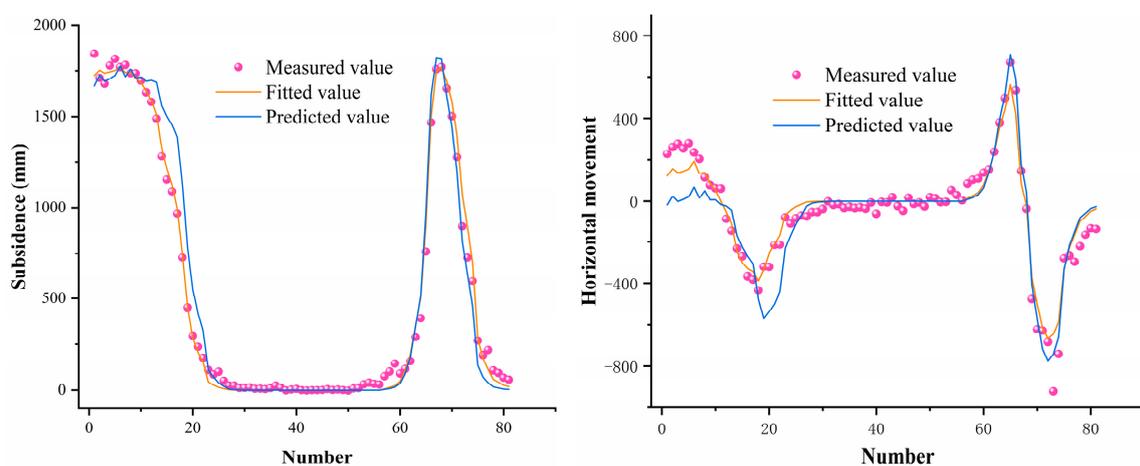


Figure 12. Comparison between measured values and predicted values.

In order to further illustrate the method of predicting mining subsidence by using the PIM parameters calculated by the M-CM-GA-ELM model, the surface movement and deformation values of the 1613 working face after mining were calculated by using the predicted PIM parameters in Table 2. The surface subsidence contour line is shown in Figure 13. The same method could be used to obtain other movement and deformation contours.

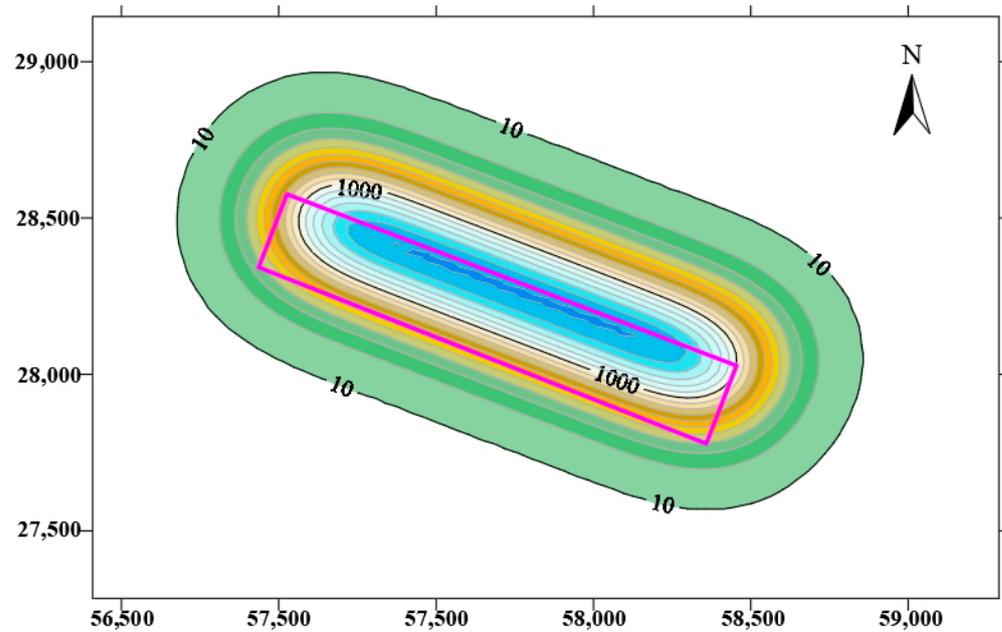


Figure 13. Contour map of surface subsidence of 1613 working face.

## 6. Discussions

### 6.1. Influence of Hidden Layer Nodes on the Prediction of Subsidence Coefficient

The number of hidden layer neurons plays a key role in the prediction accuracy of the M-CM-GA-ELM neural network model. Too many or too few hidden layers will directly affect the prediction results. Repeated experiments and their own experience are generally used to determine the number of hidden layer neurons. The experiment found that when nodes of the hidden layer are greater than 30, the error has a large mutation. The range of nodes selected in this paper was 1 to 30. The number of hidden layer nodes was judged by the size of the predicted average relative error MeaRE value of the test set. The prediction results are given in Figure 14.

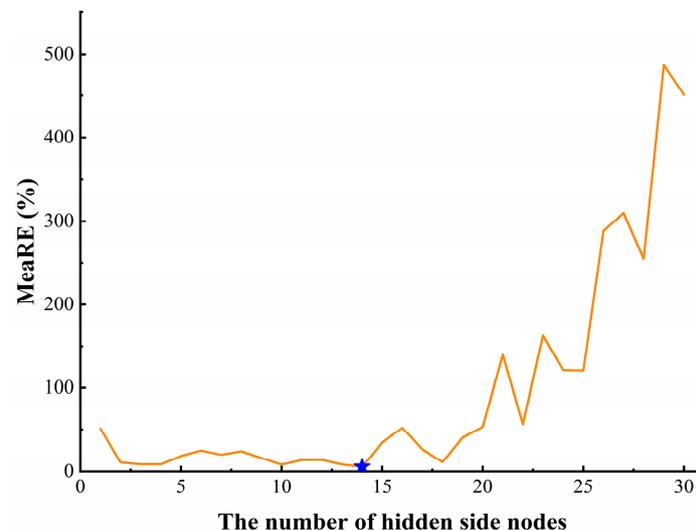


Figure 14. Selection of the number of hidden layer nodes.

In Figure 14, the prediction accuracy of the test set varies greatly with the number of hidden layers. Through multiple experiments, when the number of hidden layers was 14, the average relative mean square error of the test set reached the minimum of 5.699%, and the accuracy was the highest. Therefore, the number of hidden layers was 14 in this paper.

6.2. Influence of Activation Function Selection on Predicted Results

Common activation functions are Sigmoid (sig), Sin (sin) and hardlim, in the forms of

(1) sig function

$$f(a, b, x) = \frac{1}{1 + e^{-(ax+b)}} \tag{14}$$

(2) sin function

$$f(a, b, x) = \sin(ax + b) \tag{15}$$

(3) hardlim function

$$f(a, b, x) = \begin{cases} 1 & \text{if } ax - b \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Three kinds of functions were used to construct the prediction model, and other parameters were constant. The M-CM-GA-ELM prediction model was used to predict subsidence coefficient. Figure 15 and Table 3 show the prediction results.

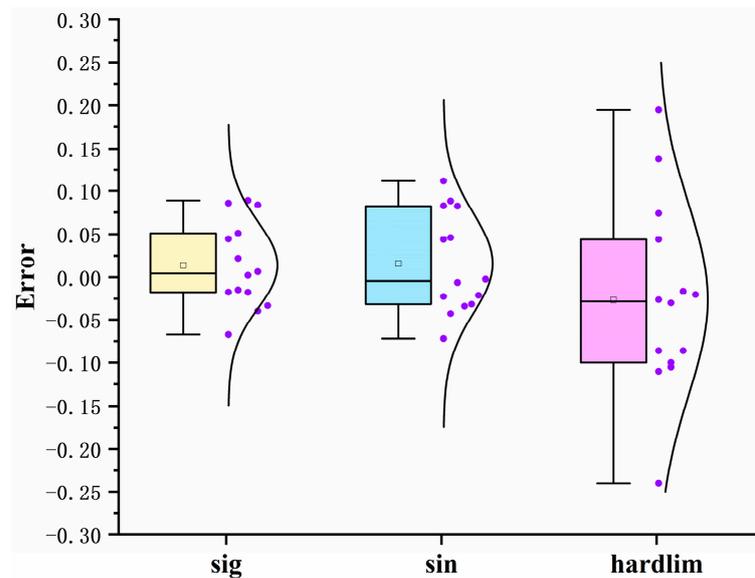


Figure 15. Comparison of prediction results of different activation functions.

Table 3. The accuracy of the prediction model (unit:mm).

Activation Function		sig	sin	Hardlim
Subsidence factor	MeaRE	5.509	6.321	11.296
	RMSE	0.050	0.059	0.110

In Figure 15, the box graph with normal distribution is used to describe errors. When the activation function was sig, the error was between  $-0.075$  and  $0.1$ . When the activation function was sin, the error was between  $-0.075$  and  $0.125$ . When the activation function was hardlim, the error was mainly distributed between  $-0.25$  and  $0.2$ . The sig function had the

highest accuracy. The same conclusion could be drawn from the probability curve and box graph in Figure 15. At the same time, according to MeaRE and RMSE values corresponding to different activation functions in Table 3, when the activation function was a sig function, the model had the best generalization performance, followed by the sin function, and the hardlim function was the worst. This shows that when the activation function of the M-CM-GA-ELM model is sig, it can achieve better generalization performance during the prediction of subsidence coefficient.

### 6.3. Influence of the Number of Test Sets on Training Results

Reasonable allocation of the proportion of training set and test set has a certain influence on the results. This paper discusses how to set the proportion of test set and training set. The training set and test set are usually divided into 7:3; with a validation set, they are divided into 6:2:2. When the data magnitude is below ten thousand, or for a specific problem to be solved, appropriately changing the ratio of the test set and the training set can improve the accuracy of the prediction model to some extent.

Taking the M-CM-GA-ELM solution model as an example, the number of test samples was set to 14 (20% of the total sample), 21 (30% of the total sample), 28 (40% of the total sample), and 35 (50% of the total sample). The prediction results are in Table 4.

**Table 4.** Prediction results (unit:mm)

Parameters		Number of Test Samples			
		14	21	28	35
Subsidence coefficient	MeaRE	5.509	5.588	6.236	7.824
	RMSE	0.050	0.051	0.059	0.072

When the subsidence coefficient was estimated, under a certain data level, the MeaRE value and RMSE value showed an increasing trend with the decrease in learning samples, indicating that the prediction performance was reduced, which shows that at a limited data level under the circumstances, increasing the number of learning samples can greatly improve the expected effect.

In general, it takes 1–2 years of observation to obtain the surface subsidence coefficient, so it is difficult to obtain more training samples. If a better distribution ratio can be obtained, it is of great significance to improve the prediction efficiency.

## 7. Conclusions

The selection of parameters of subsidence prediction lacks theoretical basis in mining areas without measured data or with special geological and mining conditions. In this paper, the Huainan mining area was chosen as the object of research. The mechanism of the influence of geological and mining conditions on parameters of PIM was deeply analyzed. Then, a combined prediction model was constructed by integrating the parameters of the multiple regression model and the extreme learning machine with the parameters of PIM. The measured surface subsidence values in mining areas were taken as the learning, training, and testing samples of the network to verify the accuracy and reliability of the model. The effects of the number of nodes in the hidden layer, the selection of activation function, and the ratio of training set to test set on the results were illustrated. It provides an effective way to select related parameters of the surface movement basin of new working faces.

It needs to be explained that the verification of the CM-GA-ELM model in this paper is mainly aimed at the Huainan mining area, so the results are better applicable to this mining area; further verification is still needed for other thick-, loose-layered mining areas. The gaps between the rocks left after coal mining will be gradually compacted by the overlying rocks, and the surface will slowly deform during the compaction process. The probability integral method model assumes that the overlying rocks have been compacted. The parameters were obtained under compaction conditions, but with the passage of time

they will still cause a slight subsidence of the surface. and we will carry out related research in the future.

**Author Contributions:** Conceptualization, S.C. and X.Y.; methodology, L.W.; validation, S.C., X.Y. and L.W.; formal analysis, S.C.; investigation, S.C.; resources, X.Y.; data curation, X.Y.; writing—original draft preparation, S.C.; writing—review and editing, S.C.; visualization, L.W.; supervision, X.Y.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Natural Science Foundation of Anhui Colleges (Grant NO.KJ2021A0422), the Coal Industry Engineering Research Center of Mining Area Environmental and Disaster Cooperative Monitoring (Grant NO.KSXTJC202201), the Introduction of Talent Research Start-up Fund of Anhui University of Science and Technology (Grant No. 2021yjrc43), the Major Science and Technology Projects of Anhui Province (Grant NO. 202103a05020026), the Key Research and Development Program of Anhui Province (Grant NO. 202104a07020014), and the Natural Science Foundation of China (Grant No. 52074010).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this article is not public.

**Acknowledgments:** We gratefully acknowledge many important contributions from the authors of all the reports cited in our paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Saeidi, A.; Deck, O.; Seifaddini, M.; Heib Marwan, A.L.; Verdel, T. An improved methodology for applying the influence function for subsidence hazard prediction. *Georisk* **2022**, *16*, 347–359. [[CrossRef](#)]
2. Smolinski, A.; Malashkevych, D.; Petlovanyi, M.; Rysbekov, K.; Lozynskiy, V.; Sai, K. Research into Impact of Leaving Waste Rocks in the Mined-Out Space on the Geomechanical State of the Rock Mass Surrounding the Longwall Face. *Energies* **2022**, *15*, 9522. [[CrossRef](#)]
3. Kwinta, A.; Gradka, R. Analysis of the damage influence range generated by underground mining. *Int. J. Rock. Mech. Min. Sci.* **2020**, *128*, 104263. [[CrossRef](#)]
4. Modeste, G.; Doubre, C.; Masson, F. Time evolution of mining-related residual subsidence monitored over a 24-year period using InSAR in southern Alsace, France. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102392. [[CrossRef](#)]
5. Guzy, A.; Witkowski, W. Land Subsidence Estimation for Aquifer Drainage Induced by Underground Mining. *Energies* **2021**, *14*, 4658. [[CrossRef](#)]
6. Malashkevych, D.; Petlovanyi, M.; Sai, K.; Zubko, S. Research into the coal quality with a new selective mining technology of the waste rock accumulation in the mined-out area. *Min. Miner. Depos.* **2022**, *16*, 103–114. [[CrossRef](#)]
7. Kapusta, L.; Szojda, L. The role of expansion joints for traditional buildings affected by the curvature of the mining area. *Eng. Fail. Anal.* **2021**, *128*, 105598. [[CrossRef](#)]
8. Adelson, E.; Iannacchione, A.; Winn, R. Investigations on longwall mining subsidence impacts on Pennsylvania highway alignments. *Int. J. Min. Sci. Technol.* **2020**, *30*, 85–92. [[CrossRef](#)]
9. Xiao, W.; Chen, W.Q.; He, T.T.; Zhao, Y.L.; Hu, Z.Q. Remote sensing monitoring and impact assessment of mining disturbance in mining area with high underground water level. *J. China Coal Soc.* **2021**, *47*, 922–933.
10. Tichavsky, R.; Jirankova, E.; Fabianova, A. Dating of mining-induced subsidence based on a combination of dendro geomorphic methods and in situ monitoring. *Eng. Geol.* **2020**, *272*, 105650. [[CrossRef](#)]
11. Xu, Z.H.; Zhu, R.S.; He, W.R.; Li, X.B. Subsidence model of large working face in shallow buried coal seam with thick loose layer. *J. Min. Saf. Eng.* **2020**, *37*, 264–271.
12. Knothe, S. Effect of time on formation of basin subsidence. *Arch. Min. Steel Ind.* **1953**, *1*, 1–7.
13. Cui, X.M.; Che, Y.H.; Malinowska, A.; Zhao, Y.L.; Li, P.X. Method and problems for subsidence prediction in entire process induced by underground mining. *J. China Coal Soc.* **2022**, *47*, 2170–2181.
14. Ghabraie, B.; Ren, G.; Barbato, G.; Smith, J.V. A predictive methodology for multi-seam mining induced subsidence. *Int. J. Rock. Mech. Min. Sci.* **2017**, *93*, 280–294. [[CrossRef](#)]
15. Liu, B.C.; Dai, H.Y. Research Development and Origin of Probability Integral Method. *Coal Min. Technol.* **2016**, *21*, 1–3.
16. Chen, Y.; Tao, Q.X.; Liu, G.L.; Wang, L.Y.; Wang, F.Y. Detailed mining subsidence monitoring combined with InSAR and probability integral method. *Chin. J. Geophys. Chin. Ed.* **2021**, *64*, 3554–3566.

17. Zha, J.F.; Feng, W.K.; Zhu, X.J. Research on Parameters Inversion in Probability Integral Method by Genetic Algorithm. *J. Min. Saf. Eng.* **2011**, *28*, 655–659.
18. Niedojadlo, Z.; Gruszczynski, W. The Impact of the Estimation of the Parameters Values on the Accuracy of Predicting the Impacts of Mining Exploitation. *Arch. Min. Sci.* **2015**, *60*, 173–193.
19. Li, P.X.; Tan, Z.X.; Yan, L.L. A Shaft Pillar Mining Subsidence Calculation Using Both Probability Integral Method and Numerical Simulation. *CMES-Comp. Model. Eng. Sci.* **2022**, *117*, 231–249. [[CrossRef](#)]
20. Gruszczynski, W.; Niedojadlo, Z.; Mrochen, D. Influence of Model Parameter Uncertainties on Forecasted Subsidence. *Acta Geodyn. Geomater.* **2018**, *15*, 211–228.
21. Zou, Y.F. The determining method of predication parameters on mining subsidence. *J. Henan Polytech. Univ. Nat. Sci.* **2001**, *20*, 253–257.
22. Guo, W.B.; Deng, K.Z.; Zou, Y.F. Artificial Neural Network Model for Predicting Parameters of Probability-Integral Method. *J. China Univ. Min. Technol.* **2001**, *33*, 88–92.
23. Kwinta, A. Application of the Least Squares Method in Determination of the Knothe Deformation Prediction Theory Parameters. *Arch. Min. Sci.* **2011**, *56*, 319–329.
24. Wang, L.; Jiang, K.G.; Wei, T.; Jiang, C.; Zha, J.F. Estimation of parameters of probability integral method model based on improved fireworks algorithm. *Surv. Rev.* **2022**, *53*, 366–382. [[CrossRef](#)]
25. Hejmanowski, R.; Kwinta, A. Determining the Coefficient of Horizontal Displacements with the Use of Orthogonal Polynomials. *Arch. Min. Sci.* **2009**, *54*, 441–454.
26. Wang, Q.C.; Li, J.; Guo, G.L. Study on analogy analysis on prediction of surface subsidence parameters in mining area of Chongqing City. *Coal Sci. Technol.* **2018**, *46*, 196–202.
27. Polanin, P. Application of two parameter groups of the Knothe-Budryk theory in subsidence prediction. *J. Sustain. Min.* **2015**, *14*, 67–75. [[CrossRef](#)]
28. Zhang, X.Y.; Tian, Y.P.; Zhang, X.F.; Bai, M.D.; Zhang, Z.T. Use of multiple regression models for predicting the formation of bromoform and dibromochloromethane during ballast water treatment based on an advanced oxidation process. *Environ. Pollut.* **2019**, *254*, 113028. [[CrossRef](#)] [[PubMed](#)]
29. Bian, H.F.; Yang, H.C.; Zhang, S.B. Research on intelligent optimization for predicting parameters of probability-integral method. *J. Min. Saf. Eng.* **2013**, *30*, 385–389.
30. Huang, H.Z. Study on Geological Theories of Stress Relief Coalbed Methane Drainage From the Distant Protected Seam by Vertical Surface Wells and Their Application in Huainan Coal Mine Area. Ph.D. Thesis, China University of Mining and Technology, Xuzhou, China, 2010.
31. Shi, X.C.; Zhang, J.X.; Li, G.Q. Characteristics of in situ stress field in the Huainan mining area, China and its control factors. *Environ. Earth Sci.* **2021**, *80*, 682. [[CrossRef](#)]
32. Liu, W.Q.; Chen, G.Z. Evolution of Ecosystem Service Value and Ecological Storage Estimation in Huainan Coal Mining Area. *Bull. Environ. Contam. Toxicol.* **2021**, *107*, 1243–1249.
33. Liu, B.C.; Zhang, J.S. A stochastic media method for surface subsidence induced by near-surface excavation. *Chin. J. Rock. Mech. Eng.* **1995**, *14*, 289–296.
34. Tejada, A.T.; Ella, V.B.; Lampayan, R.M.; Reaño, C.E. Modeling Reference Crop Evapotranspiration Using Support Vector Machine (SVM) and Extreme Learning Machine (ELM) in Region IV-A, Philippines. *Water* **2022**, *14*, 754. [[CrossRef](#)]
35. Yu, C.X.; Koopialipoor, M.; Murlidhar, B.R.; Mohammed, A.S.; Armaghani, D.J. Optimal ELM-Harris Hawks Optimization and ELM-Grasshopper Optimization Models to Forecast Peak Particle Velocity Resulting from Mine Blasting. *Nat. Resour. Res.* **2021**, *30*, 2647–2662. [[CrossRef](#)]
36. Hu, B.N.; Zhang, H.X.; Shen, B.H. *Guide for Coal Pillar Reservation and Coal Pressure Mining in Buildings, Water Bodies, Railways and Main Shafts and Roadways*; Coal Industry Press: Beijing, China, 2018; pp. 100–158.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.