Hindawi
Geofluids
Volume 2022, Article ID 3590764, 9 pages
https://doi.org/10.1155/2022/3590764

Research Article

A Prediction Model for Top-Coal Drawing Capability in Steep Seams Based on PCA–GRNN

Zhijie Zhu,[1,2] **Yin Hong**,[1] **and Zhuang Liang**[3]

[1]*School of Mining, Liaoning Technical University, Fuxin, China*
[2]*State Key Laboratory of Coal Mining and Clean Utilization, Beijing, China*
[3]*Research Centre, Ministry of Emergency Management, China*

Correspondence should be addressed to Zhijie Zhu; zhuzhijie@lntu.edu.cn

Received 22 March 2022; Accepted 26 April 2022; Published 14 May 2022

Academic Editor: Zhongguang Sun

To improve prediction accuracy of top-coal drawing capability in steep seams, principal component analysis (PCA) and the general regression neural network (GRNN) are combined (PCA–GRNN model) to predict top-coal drawing capability in steep seams. Nine commonly used influencing factors are selected to establish a predictive index system for top-coal drawing capability in steep seams. The PCA is used to eliminate correlation and reduce dimensions of various indices, thus obtaining three linearly uncorrelated principal components (PCs) y_1, y_2, and y_3, which form the input vectors of the GRNN. In this way, the factors that most affect the top-coal drawing capability in steep seams are found to be floor flatness, dip angle of the coal seam, and the hardness of the coal seam. The results show that the PCA–GRNN model outperforms the GRNN and random forest models in prediction results, which indicates that the PCA improves prediction accuracy of the GRNN model. It is feasible to predict top-coal drawing capability in steep seams by combining or even integrating different analytical models into one. The proposed PCA–GRNN model can be used to predict top-coal drawing capability in steep seams.

1. Introduction

Top-coal drawing capability refers to the difficulty of drawing of the top part of a coal seam under the mine pressure and its weight. The selection of an appropriate top-coal caving technology directly can determine the mining efficiency, and high top-coal drawing capability is the premise of applying the technology [1, 2]. In many top-coal caving faces, the top-coal drawing capability is not evaluated before mining, which leads to reduced economic benefits. Therefore, research into top-coal drawing capability cannot only improve the economic benefit but also provide an important theoretical basis for mine production [3–9].

At present, several methods have been proposed for identifying top-coal drawing capability or difficulty, such as the distance discrimination method, support vector machine (SVM), Fisher discriminant analysis, and neural network methods. According to the merits and demerits of these methods, Liu et al. [10] established a distance discrimination analysis model for distinguishing difficulty of top-coal caving in steep seams, which helped the popularization and application of the roadway caving method in steep seams. Liu et al. [11] built an SVM model for identifying difficulty of top-coal caving based on a radial basis function, which provides a new method for determining the top-coal drawing capability. Long-jun et al. [12] established a Fisher discriminant analysis model for judging top-coal drawing capability. They selected nine indices including mining depth of main roof, thickness of coal seam, and Protodyakonov coefficient as classification indices of the model, which allows accurate prediction of top-coal drawing capability in coal seams under different mining conditions. Wang et al. [13] evaluated the top-coal drawing capability in fully mechanized caving faces using the artificial neural network method. They not only assessed the extent of top-coal caving but also predicted the comprehensive technical and economic indicators of the working face. Each of these methods has their own merits and demerits. For example, the distance discrimination method regards that various indices or factors pertaining to samples are of equal importance when

determining the distance, while in fact, these indices or factors do not play roles of equal importance in determining the classification of samples. Therefore, if the importance of various indices or factors is not determined in advance, the distance discrimination method is likely to overstate the effect of some less important indices, leading to misjudgements in predictions [14]. The SVM is a machine learning method based on statistical learning theory and optimization theory that maximizes the geometric spacing for separating hyperplanes. By introducing kernel functions, the method can transform nonlinear classification problems into linear ones in a high-dimensional space. However, the method is limited to data decomposition itself and ignores the intrinsic structural compactness of data. This not only results in high complexity of the algorithm but is likely to lead to errors in processing noisy data, thus reducing classification accuracy. The Fisher discriminant analysis is a statistic analysis technique to identify newly obtained samples according to some existing quantitative characteristics of observation samples. The approach maps high-dimensional data points into a low-dimensional space, to render data points more dense. However, the matrix inversion and eigenvector calculation increase the computational workload, and the pairwise extraction and classification criterion need to be introduced for classification of multiple classes. The neural network method not only has the self-learning function but also the associative memory function. In the case of an onerous computational burden, the use of a feedback artificial neural network designed for a specific problem can give full play to the high-speed computing power available, so it may seek the optimal solution rapidly.

The general regression neural network (GRNN) has the advantages of simple structure, easy training, fast convergence, and strong fault tolerance and is mainly used in pattern classification problems such as fault diagnosis. In fact, most of the top-coal drawing capability prediction indicators have certain correlations, so the correlation between the prediction indicators should be eliminated before applying the GRNN network. Common methods to eliminate the correlation between indicators include limiting the number of indicators, separating overlapping elements, modifying indicator weights, principal component analysis, and factor analysis. Considering that there are many prediction indicators used in this paper, the principal component analysis (PCA) is used to preprocess the data of the prediction indicators of top-coal drawing capability, which can not only eliminate the correlation between the indicators but also reduce the dimensionality of the indicator data and improve the PNN. Based on the above theoretical analysis, in this paper, 25 groups of data on the influencing factors of top-coal drawing cavability are selected, and principal component analysis (PCA) is used to reduce the dimensionality of the indicators and convert multiple indicators into a few independent indicators. This eliminates correlation and realizes dimension reduction of indexes. Then, the general regression neural network (GRNN) is introduced. The PCA–GNRR prediction model for top-coal drawing capability in steep seams is established by combining the PCA and GRNN. In addition, a random forest model is also used to compare accuracy of prediction results of various models. In this way, the performance of the established model in predicting top-coal drawing capability in steep seams is evaluated, which provides a basis for improving accuracy of such predictions.

2. Factors Influencing Top-Coal Drawing Capability in Steep Seams

2.1. Roof Conditions. Roof conditions influence the top-coal drawing capability mainly through stability of the immediate roof and the main roof. If the immediate roof can cave following mining, it does not influence the top-coal drawing capability; if the immediate roof is very hard and does not cave over a large area, it will bring significantly affect the working face upon failure of the roof.

In the top-coal caving mining, stress on the main roof is relieved due to the buffering effect and absorption of the top coal for weighting, so the working face is less affected and suffers less damage during weighting, while the range of influence of the pressure is enlarged. Meanwhile, intense weighting on the main roof during slicing mining may induce spalling of the working face [15, 16].

2.2. Floor Conditions. Top-coal drawing capability of a coal seam is related to two interdependent factors, i.e., the stability and flatness of floors. That is, the flatter the floor, the more stable it is, and the better the top-coal drawing capability [17, 18].

2.3. Gas. Gas pressure is one of the causes of breakage of top coal. A high gas content can soften coal seams, which is conducive to breakage of top coal; however, the roadway caving method is not applicable to coal seams prone to gas outburst. Gas accumulation should be paid close attention to in coal seams with a high gas content [19–23].

2.4. Mining Depth. The mining depth directly influences the magnitudes of *in situ* stress and peak abutment pressure in surrounding rocks of a working face. The abutment pressure plays a decisive role in breakage of top coal. When ignoring the influences of the tectonic stress field, the greater the depth of occurrence of a coal seam, the more readily the critical failure condition of top coal is met and the higher the top-coal drawing capability, according to the Griffith strength criterion [24–26].

2.5. Dip Angle of Coal Seams. For coal seams with a large dip angle, the self-weight of a coal mass in the vertical direction is larger than the resultant of other forces acting thereon, so that coal mass is more likely to cave, which is favorable for drawing of top coal. If the dip angle reaches 90°, the force acting along the vertical direction only includes the weight of broken coal gangue in the goaf above the top coal, so the coal does not readily cave [27, 28].

2.6. Thickness of Coal Seams. Top coal is subjected to presplitting and damage in the mining process and stores significant amounts of energy. The energy is released during the migration of top coal, which breaks coal mass. If the top coal is too thin, it is difficult to ensure its caving at the tail of

support and causes advanced breakage of the immediate roof, which is drawn out together with top coal. As a result, a large amount of top coal is lost in the goaf. If the top coal is too thick, it is challenging to ensure sufficient looseness of top coal in the roof-control zone, so the top coal does not readily cave in the caving zone. In addition, there will not be enough room for caving if the top coal is too thick [29].

2.7. Hardness of Coal Seams. The hardness of coal seams is an important index for evaluating damage resistance of coal seams and directly influences the failure process and breakage degree of top coal under compressive stress. Therefore, it is inversely proportional to the top-coal drawing capability: the lower the hardness, the higher the capability. The hardness is represented by Protodyakonov's coefficient [30–32].

2.8. Dirt Bands. The extent of dirt bands is expressed as m/M, where m and M separately denote the total thickness of dirt bands in a coal seam and that of the coal seam. Influences of dirt bands on top-coal drawing capability are shown as follows: if dirt bands are weaker than the coal mass, they form a weak plane in coal seams and their presence is conducive to the breakage, caving, and drawing of top coal, which improves the top-coal drawing capability. The more, the thicker, and the softer the dirt bands are, the better the top-coal drawing capability, whereas dirt bands that are harder than coal mass are unfavorable for drawing of top coal [33].

3. Principles Underpinning the Methods

3.1. PCA. The PCA is a dimension reduction method in mathematics that uses orthogonal transform to convert a series of possibly linearly correlated variables into a set of linearly uncorrelated new variables or principal components (PCs). In this way, new variables are used to characterize data features in a lower dimension. These PCs are linear combination of original variables and their number is less than that of original ones. The combination is equivalent to generation of a new set of observations, which have different meanings with original data while contain most features of original data, show lower dimensions, and therefore are convenient for further analysis.

The PCA can be spatially interpreted as mapping original data into a new coordinate system. The first PC corresponds to the first coordinate axis, which represents the range of variation of the new variable transformed in a certain way from multiple variables in the original data; the second PC corresponds to the second coordinate axis and represents the range of variation of the second new variable transformed in a certain way from multiple variables in the original data. In a similar fashion, the difference in samples interpreted by original data is transformed into that interpreted by new variables. To remain interpretation of original data as far as possible, the maximum variance theory or minimum damage theory is generally used to ensure that the first PC has the largest variance (able to interpret difference in original data as much as possible). Each of subsequent PCs is orthogonal with the previous one and has the largest variance only second to the previous one. The PCs are derived as follows:

$$\begin{cases} y_1 = p_{11}x_1 + p_{12}x_2 + \cdots + p_{1m}x_m, \\ y_2 = p_{21}x_1 + p_{22}x_2 + \cdots + p_{2m}x_m, \\ \vdots \\ y_m = p_{m1}x_1 + p_{m2}x_2 + \cdots + p_{nm}x_m. \end{cases} \tag{1}$$

The PCs are calculated as follows:

(1) For the matrix X:

$$X = \begin{bmatrix} X_{11}\,X_{12} & \cdots & X_{1m} \\ X_{21} \quad X_{22} & \cdots & X_{2m} \\ \vdots \quad \vdots & \ddots & \vdots \\ X_{\mathrm{n1}} \quad X_{n2} & \cdots & X_{nm} \end{bmatrix}. \tag{2}$$

After standardization,

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}, \tag{3}$$

where

$$\bar{X}_j = \frac{1}{n}\sum_{n=1}^{n} X_{ij}, \tag{4}$$

$$S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2} \tag{5}$$

(2) The symmetric correlation coefficient matrix R of standardized variables is calculated as

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} \quad r_{22} & \cdots & r_{2m} \\ \vdots \quad \vdots & \ddots & \vdots \\ r_{n1} \quad r_{n2} & \cdots & r_{nm} \end{bmatrix}. \tag{6}$$

The correlation coefficient r_{jk} between variables is

$$r_{jk} = \frac{\sum_{k=1}^{n}\left(x'_{ki} - \bar{x}'_j\right)\left(x'_{kj} - \bar{x}'_j\right)}{\sum_{k=1}^{n}\left(x'_{ki} - \bar{x}'_i\right)^2 \sum_{k=1}^{n}\left(x'_{kj} - \bar{x}'_j\right)} \tag{7}$$

(3) The matrix R is subjected to eigenvalue decomposition, to calculate eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{m}}$ and eigenvectors $p_1, p_2, \ldots, p_m$

(4) PCs are calculated as $t_i : t_i = X \times p_i$

(5) Contribution v_s and cumulative contribution v_t of each PC are calculated as follows:

$$v_s = \frac{\lambda_i}{\sum_{k=1}^{m} \lambda_k} (i = 1, 2, \cdots, m), \tag{8}$$

$$v_t = \frac{\sum_{s=1}^{i} \lambda_s}{\sum_{k=1}^{m} \lambda_k} (i = 1, 2, \cdots, m) \tag{9}$$

In actual application, the first to the k^{th} ($k \le m$) PCs corresponding to eigenvalues whose cumulative contribution is greater than 85%, or PCs whose eigenvalues are greater than 1 are generally selected.

3.2. GRNN. The GRNN integrates the density estimation and Bayesian decision theory based on the radial basis function (RBF) neural network and substitutes the sigmoid activation function using an activation function deduced by the statistical method. The GRNN is also similar to the probabilistic neural network (PNN) in terms of structure, both comprising the input, model, summation, and output layers. The difference lies in that the GRNN has two types of neurons on the summation layer, allowing more comprehensive computation than the PNN.

The mapping relationship of the GRNN as shown in Figure 1 is established according to the following steps:

(1) The input layer is responsible for transferring input variables to the model layer via a linear function; neurons on the model layer correspond to different samples and the transfer function p_i is

$$p_i = \exp\left[-\frac{(U - U_i)^T (U - U_i)}{2\sigma^2}\right], \tag{10}$$

where p_i denotes the output of the i^{th} neuron on the hidden layer and U_i is the learning sample corresponding to the i^{th} neuron

(2) The output of the model layer is calculated through summation, in two ways: one is the arithmetic summation S_D:

$$S_D = \sum_{i=1}^{n} p_i \tag{11}$$

The other is weighted summation S_{Nj}:

$$S_{Nj} = \sum_{i=1}^{n} \gamma_{ij} p_i (j = 1, 2, \cdots, s), \tag{12}$$

where γ_{ij} is the connection weight and is valued as the j^{th} element in the i^{th} output sample and s represents the dimension of the output vector of the learning samples

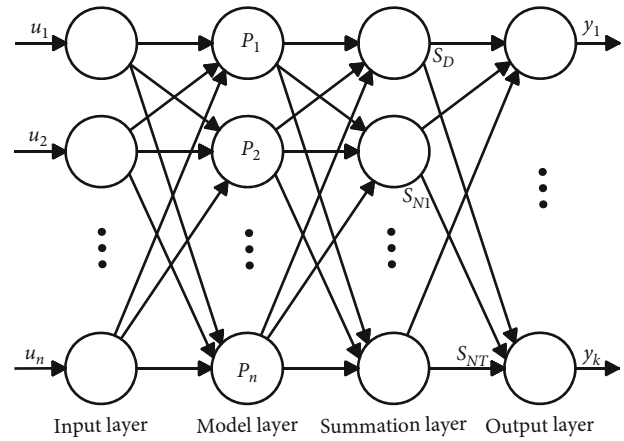

Figure 1: Structure of the GRNN model.

(3) By dividing outputs of the summation layer of each neuron, the output of each neuron on the output layer is obtained as

$$\gamma_j = \frac{S_{Nj}}{S_D} (j = 1, 2, \cdots, s) \tag{13}$$

3.3. PCA–GRNN Prediction Model for Top-Coal Drawing Capability in Steep Seams. The flowchart of the proposed PCA–GRNN prediction model for top-coal drawing capability in steep seams is shown in Figure 2. The main calculation steps are as follows.

Step 1. Selecting evaluation indices for factors influencing top-coal drawing capability.

Step 2. Collecting case data according to the indices.

Step 3. Zero-mean normalization of data about top-coal drawing capability to eliminate influences of different dimensions across indices on the test results, followed by correlation analysis of normalized data.

Step 4. Using PCA to eliminate correlation and reduce dimensions of indices, thus determining PCs.

Step 5. Establishing the GRNN model, in which the smooth factor is input. The model is trained with training samples until attaining satisfactory results.

Step 6. Inputting testing samples for predicting levels of top-coal drawing capability in the trained model and using the evaluation indices to evaluate and compare the prediction accuracy.

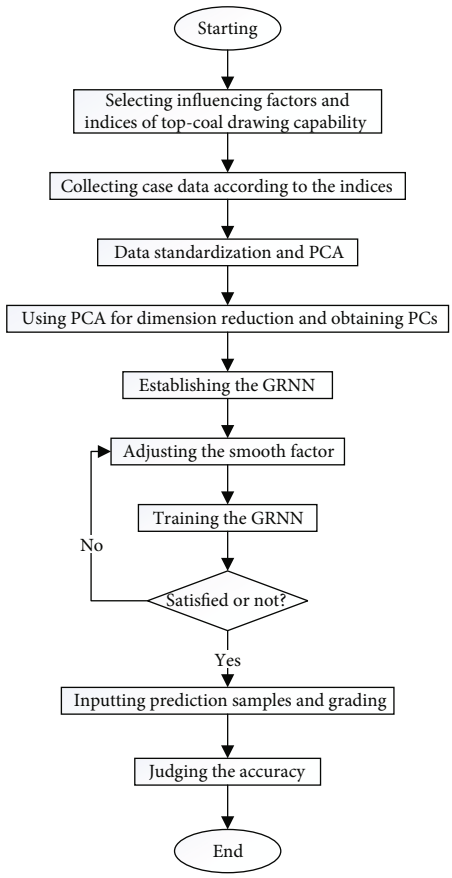

Figure 2: Flowchart of the prediction model for top-coal drawing capability in steep seams.

Table 1: Data about influencing factors of top-coal drawing capability.

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Actual level
1	4	4	1	28	280	75	4.5	1.2	0	A
2	3	3	1	36	385	58	2.3	1.3	0	A
3	2	2	2	36	425	54	2.1	1.4	0	A
4	3	2	1	39	340	59	2.0	1.1	0	A
5	4	2	1	31	310	70	6.5	1.1	0	A
6	3	1	1	4	340	67	4.5	1.8	0.1	B
7	2	2	2	38	627	78	7.9	1.8	0.05	B
8	2	1	1	38	649	76	5.8	1.8	0	B
9	2	2	2	30	546	72	6.4	1.9	0	B
10	3	2	1	33	527	79	6.8	1.5	0	B
11	3	2	1	3	280	65	3.5	1.7	0.08	C
12	4	3	1	5	302	68	3.6	1.7	0.1	C
13	3	3	2	5	322	63	3.2	1.8	0	C
14	2	2	1	4	328	63	3.8	1.6	0	C
15	4	4	1	12	320	68	2.5	1.5	0	C
16	4	3	1	3	180	85	5.2	2.3	0.1	D
17	4	4	2	4	200	89	4.9	2.0	0.15	D
18	4	3	2	6	168	82	4.8	2.0	0.18	D
19	3	3	2	5	170	83	5.0	2.0	0.23	D
20	4	3	1	5	180	87	2.9	1.8	0.35	D
21	3	2	2	36	405	60	2.4	1.3	0	A
22	3	2	2	25	430	2	3.8	1.1	0	A
23	2	2	1	24	352	54	2.2	1.2	0	B
24	3	3	1	8	368	70	3.0	1.3	0	B
25	3	3	2	10	340	54	2.0	1.3	0	C

4. Calculation Process and Results

4.1. Evaluation Indices and Data Pertaining to Top-Coal Drawing Capability in Steep Seams. Through comprehensive analysis, the type of main roof (X_1), the stability of the immediate roof (X_2), floor flatness (X_3), gas content (X_4), mining depth (X_5), dip angle of coal seams (X_6), thickness of coal seams (X_7), hardness of coal seams (X_8), and extent of dirt bands (X_9) are selected as discriminant factors affecting top-coal drawing capability. Therein, the type of main roof X_1 is graded into four levels as insignificant, significant, intense, and very intense according to the degree of first weighting. The stability of immediate roof X_2 is divided into four grades, that is, unstable, moderately stable, stable, and very stable, based on different first caving steps of the immediate roof: less than 8 m, 8–18 m, 18–28 m, and 28–50 m. The floor flatness X_3 is graded at four levels as flat, less flat, rough, and very rough. The gas content X_4, mining depth X_5, dip angle of coal seams X_6, and thickness of coal seams X_7 take corresponding values; the hardness of coal seams X_8 is represented by the Protodyakonov coefficient; and the extent of dirt bands X_9 is expressed as m/M, that is, the ratio of thickness of dirt bands to that of coal seams. The top-coal drawing capability is graded into four levels: very high (A), high (B), general (C), and poor (D). Table 1 lists 25 groups of data selected in the research.

4.2. PCA Preprocessing. To avoid influences induced by dimensional differences of sample data, data in the training set and the test set are standardized and then subjected to PCA. Correlation coefficients of various factors are listed in Table 2. The absolute value Q of the correlation coefficient reflects the degree of correlation between two factors. When $Q = 0$, $0 < Q \le 0.5$, $0.5 < Q \le 0.8$, $0.8 < Q \le 1$, and $Q = 1$, two factors are uncorrelated, slightly correlated, significantly correlated, extremely significantly correlated, and completely correlated, respectively. The absolute values of correlation coefficients between X_1 and X_2, X_1 and X_5, X_2 and X_5, X_4 and X_5, X_4 and X_8, X_4 and X_9, X_6 and X_8, and X_8 and X_9 are all greater than 0.5, so they are significantly correlated.

The total variances, variance percent, and cumulative variance of the initial eigenvalues, extraction sums of squared loadings, and rotation sums of squared loadings of each PC are listed in Table 3: the first three PCs have eigenvalues larger than 1 and initial cumulative contributions of 40.670%, 62.567%, and 75.280%, respectively, which contain the majority of information pertaining to the original factors. Therefore, the first three PCs are selected here as the comprehensive evaluation indices reflecting top-coal drawing capability.

Table 2: Correlations of various factors.

Factor		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Correlation	X_1	1.000	0.664	-0.198	-0.432	-0.705	0.295	-0.046	0.120	0.437
	X_2	0.664	1.000	0.081	-0.402	-0.592	0.285	-0.220	0.114	0.259
	X_3	-0.198	0.081	1.000	0.045	0.078	-0.171	0.092	0.197	0.068
	X_4	-0.432	-0.402	0.045	1.000	0.717	-0.252	0.151	-0.527	-0.547
	X_5	-0.705	-0.592	0.078	0.717	1.000	-0.250	0.339	-0.233	-0.601
	X_6	0.295	0.285	-0.171	-0.252	-0.250	1.000	0.420	0.633	0.498
	X_7	-0.046	-0.220	0.092	0.151	0.339	0.420	1.000	0.433	0.088
	X_8	0.120	0.114	0.197	-0.527	-0.233	0.633	0.433	1.000	0.595
	X_9	0.437	0.259	0.068	-0.547	-0.601	0.498	0.088	0.595	1.000

Table 3: Total variances.

PC	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	Variance percent	Cumulative variance (%)	Total	Variance percent	Cumulative variance (%)	Total	Variance percent	Cumulative variance (%)
1	3.660	40.670	40.670	3.660	40.670	40.670	3.306	36.737	36.737
2	1.971	21.897	62.567	1.971	21.897	62.567	2.294	25.492	62.229
3	1.144	12.713	75.280	1.144	12.713	75.280	1.175	13.052	75.280
4	0.870	9.662	84.942						
5	0.484	5.379	90.321						
6	0.444	4.932	95.253						
7	0.177	1.970	97.224						
8	0.155	1.719	98.943						
9	0.0195	1.057	100						

The scoring coefficients of components in each PC are listed in Table 4. On this basis, the formula for each PC is derived as

$$\begin{aligned} y_1 &= 0.231X_1 + 0.251X_2 + 0.022X_3 - 0.230X_4 - 0.301X_5 \\ &\quad - 0.007X_6 - 0.204X_7 + 0.013X_8 + 0.150X_9, \\ y_2 &= -0.017X_1 - 0.097X_2 - 0.042X_3 - 0.028X_4 + 0.084X_5 \\ &\quad + 0.368X_6 + 0.399X_7 + 0.350X_8 + 0.182X_9 \\ y_3 &= -0.251X_1 + 0.014X_2 + 0.799X_3 - 0.140X_4 - 0.013X_5 \\ &\quad - 0.260X_6 - 0.085X_7 + 0.222X_8 + 0.126X_9. \end{aligned}, \tag{14}$$

When reducing dimensions used the PCA in the proposed PCA–GRNN model, each of the original influencing factors contributes to different degrees. According to contributions of original influencing factors and those of PCs after dimension reduction, the floor flatness (X_3), dip angle of coal seams (X_6), and hardness of coal seams (X_8) are found to make the highest contributions (Table 5). Therefore, it is inferred that they have more obvious influences on top-coal drawing capability.

4.3. Parameter Determination of the GRNN. In order to demonstrate the superiority of the PCA in predicting top-coal drawing capability, the traditional GRNN model is compared with the model established in the present research. The PCA–GRNN model takes the comprehensive evaluation indices y_1, y_2, and y_3 for top-coal drawing capability as the input vectors, while the GRNN uses the selected indices X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, and X_9 for top-coal drawing capability as the input vectors. The research finds that the selection of the smooth factor plays a critical role in the model performance. If the smoothing factor is too

Table 4: Scoring coefficients of PCs.

	PCs		
	0.231	-0.017	-0.251
X_1	0.251	-0.097	0.014
X_2	0.022	-0.042	0.799
X_3	-0.230	-0.028	-0.140
X_4	-0.301	0.084	-0.013
X_5	-0.007	0.368	-0.260
X_6	-0.204	0.399	-0.085
X_7	0.013	0.350	0.222
X_8	0.150	0.182	0.126
X_9	0.231	-0.017	-0.251

Table 5: Calculated contributions of original influencing factors to the established model.

Influencing factors	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Contributions	0.584	0.324	0.724	0.537	0.375	0.794	0.640	0.974	0.674

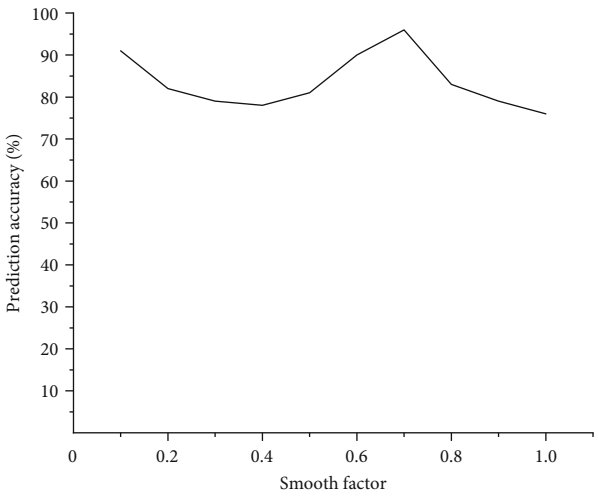

Figure 3: Relationship between the smoothing factor and prediction accuracy.

small, the network is likely to be overfitted, while if too large, a smoothing factor fails to distinguish between various details. The smooth factor here is selected through use of the construction method according to the following steps: inputting training samples to train the original GRNN model and then setting the smoothing factor to different values in the range of [0.1, 1.0], with an increment of 0.1. The relationship between prediction accuracy of the model for training samples and the value of the smoothing factor is illustrated in Figure 3. The value of the smoothing factor corresponding to the highest accuracy is selected.

4.4. Predicted Results. Apart from studying differences of the GRNN model and the PCA–GRNN model subject to dimension reduction with PCA in accuracy in the model training stage, the research also introduces the random forest model to compare test results of different models. At first, 20 groups (five groups at each level) of training samples are input as the learning samples to train each model, and then, five groups of samples to be judged are input to test performance of these models.

The test results are illustrated in Figure 4 and Table 6. For testing samples, the PCA–GRNN model always has accuracy higher than other models. Among learning samples, the PCA–GRNN model yields results different from the actual level in only one learning sample, while its results in testing samples are same as the actual level. It is evident that the PCA can improve the prediction accuracy of the model. The result indicates that the PCA–GRNN model provides a prediction method for accurately determining top-coal drawing capability.

When comparing with the random forest model and the GRNN model as shown in Table 7, the GRNN and random

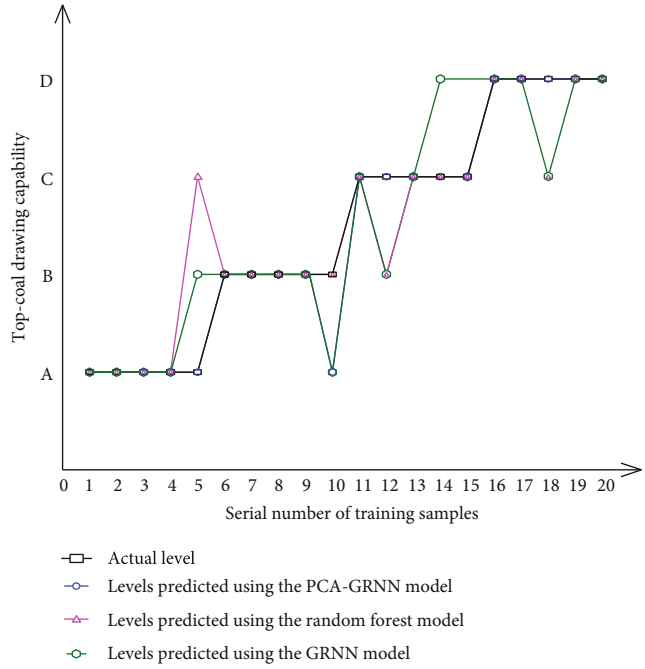

Figure 4: Comparison of levels predicted using the GRNN model, the random forest model, and the PCA–GRNN model.

Table 6: Performance test results of each model.

No.	Actual level	GRNN	Random forest	PCA–GRNN
21	A	A	A	A
22	A	B	A	A
23	B	B	C	B
24	B	B	B	B
25	C	C	C	C

Table 7: Comparison of the evaluation results of each model.

Parameters	GRNN	Random forest	PCA–GRNN
Average training accuracy (%)	92	85	98
Average prediction accuracy (%)	80	80	100
Time (s)	6.52	9.71	4.15

forest models have lower accuracy in the training samples and prediction samples. Also, PCA–GRNN model take less time to run. Therefore, the application of this model to accurately identify the cavability of coal seams is beneficial to reduce the blindness in the promotion work and improve the mining effect of top-coal caving in steeply inclined coal seams.

In view of the different level of top-coal drawing capability, necessary top-coal weakening measures will be taken to ensure the top-coal discharge rate and effectively prevent the occurrence of roof disasters.

5. Conclusions

(1) The PCA is used to process data related to factors influencing top-coal drawing capability, thus transforming nine influencing factors into three PCs. This reduces dimensions of data, simplifies the prediction model, and finally improves prediction efficiency and accuracy

(2) PCA and GRNN are combined to establish the PCA–GRNN prediction model for top-coal drawing capability in steep seams, and an appropriate smooth factor is selected. Comparisons with the GRNN and random forest models prove that the accuracy of the PCA–GRNN model is higher

(3) According to analysis of contributions of each influencing factor to PCs and PCs to the model in the PCA process, importance of each influencing factor is calculated and predicted. In this way, the floor flatness (X_3), dip angle of coal seams (X_6), and hardness of coal seams (X_8) are found to make the largest contributions, so it is inferred that they exert significant influences on top-coal drawing capability

(4) The combined model PCA–GRNN proposed in the research can improve the prediction accuracy, while the prediction results can be further optimized. In future research, the number of samples will be increase to improve the generalization ability and prediction accuracy of the model. At the same time, considering the influence of multicoal seam mining, groundwater, and other factors on the cavitation of top coal, the prediction model will be further improved

Data Availability

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the State Key Laboratory of Coal Mining and Clean Utilization (2021-CMCU-KF016) and the Basic Scientific Research Projects of Universities in Liaoning Province (LJKZ0343).

References

[1] Z. Zhu, Y. Wu, and J. Han, "A prediction method of coal burst based on analytic hierarchy process and fuzzy comprehensive evaluation," *Frontiers in Earth Science*, vol. 9, 2022.

[2] X. Li, Z. Cao, and Y. Xu, "Characteristics and trends of coal mine safety development," *Energy Sources Part a-Recovery Utilization and Environmental Effects*, vol. 2020, pp. 1–19, 2021.

[3] A. Celik and Y. Ozcelik, "Investigation of the efficiency of longwall top coal caving method applied by forming a face in horizontal thickness of the seam in steeply inclined thick coal seams by using a physical model," *International Journal of Rock Mechanics and Mining Sciences*, vol. 148, p. 104917, 2021.

[4] X. L. Li, S. J. Chen, S. M. Liu, and Z. H. Li, "AE waveform characteristics of rock mass under uniaxial loading based on Hilbert-Huang transform," *Journal of Central South University*, vol. 28, no. 6, pp. 1843–1856, 2021.

[5] T. Ai, S. Wu, R. Zhang et al., "Changes in the structure and mechanical properties of a typical coal induced by water immersion," *International Journal of Rock Mechanics and Mining Sciences*, vol. 138, p. 104597, 2021.

[6] X. Wang, Z. Wen, Y. Jiang, and H. Huang, "Experimental study on mechanical and acoustic emission characteristics of rock-like material under non-uniformly distributed loads," *Rock Mechanics and Rock Engineering*, vol. 51, no. 3, pp. 729–745, 2018.

[7] C. Fan, L. Yang, G. Wang, Q. Huang, X. Fu, and H. Wen, "Investigation on coal skeleton deformation in CO2 injection enhanced CH4 drainage from underground coal seam," *Frontiers in Earth Science*, vol. 9, 2021.

[8] B. Zhao, G. Wen, J. Nian et al., "Numerical simulation study on the multi-physical field response to underground coal and gas outburst under high geo-stress conditions," *Minerals*, vol. 12, no. 2, p. 151, 2022.

[9] Z. Wen, E. Xing, S. Shi, and Y. Jiang, "Overlying strata structural modeling and support applicability analysis for large mining-height stopes," *Journal of Loss Prevention in the Process Industries*, vol. 57, pp. 94–100, 2019.

[10] J. H. Liu, T. Feng, W. J. Wang, and D. H. Xie, "Distance discriminant analysis method for distinguishing the difficulty degree of top-coal caving in steep seam and its application," *Journal of China Coal Society*, vol. 33, no. 6, pp. 601–605, 2008.

[11] N. Liu, H. Wang, and Z. Yuan, "Support vector machine model for distinguishing the difficulty degree of top-coal caving in steep seam," *Journal of China Coal Society*, vol. 35, no. 11, pp. 1859–1862, 2010.

[12] D. Long-jun, L. I. Xi-bing, and B. A. I. Yun-fei, "A Fisher discriminant analysis model for classifying top coal cavability of the steep seam," *Journal of China Coal Society*, vol. 34, no. 1, pp. 58–63, 2009.

[13] M. Wang, C. Tai, Q. Zhang, Z. Yang, J. Li, and K. Shen, "Application of improved and optimized fuzzy neural network in classification evaluation of top coal cavability," *Scientific Reports*, vol. 11, no. 1, p. 11(1), 2021.

[14] T. D. Le, R. Mitra, J. Oh, and B. Hebblewhite, "A review of cavability evaluation in longwall top coal caving," *International Journal of Mining Science and Technology*, vol. 27, no. 6, pp. 907–915, 2017.

[15] Z. He, H. Xie, M. Gao, G. Deng, G. Peng, and C. Li, "The fracturing models of hard roofs and spatiotemporal law of mining-induced stress in a top coal caving face with an extra-thick coal seam," *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, vol. 7, no. 1, pp. 1–15, 2021.

[16] D. Kong, Y. Lou, S. Zheng, and S. Pu, "The characteristics of roof breaking and the law of ground pressure behavior in fully mechanized top-coal caving face with large mining height," *Geotechnical and Geological Engineering*, vol. 39, no. 1, pp. 285–297, 2021.

[17] H. Jangara and C. A. Ozturk, "Longwall top coal caving design for thick coal seam in very poor strength surrounding strata," *International Journal of Coal Science & Technology*, vol. 8, no. 4, pp. 641–658, 2021.

[18] D. Zhu, Z. Chen, W. du, L. Zhang, and Z. Zhou, "Caving mechanisms of loose top-coal in longwall top-coal caving mining based on stochastic medium theory," *Arabian Journal of Geosciences*, vol. 11, no. 20, p. 11(20), 2018.

[19] H. Gao, X. Gong, X. Cheng, and R. Yu, "Dynamic prediction model of multisource gas emissions in a fully mechanized top coal caving based on the coal particle size distribution," *Shock and Vibration*, vol. 2021, 13 pages, 2021.

[20] W. Wang, F. Wang, B. Zhao, and G. Li, "Gas distribution law for the fully mechanized top-coal caving face in a gassy extra-thick coal seam," *Advances in Civil Engineering*, vol. 2018, 8 pages, 2018.

[21] C. Fan, H. Wen, S. Li, G. Bai, and L. Zhou, "Coal seam gas extraction by integrated drillings and punchings from the floor roadway considering hydraulic-mechanical coupling effect," *Geofluids*, vol. 2022, 10 pages, 2022.

[22] B. Zhao, G. Wen, Q. Ma, H. Sun, F. Yan, and J. Nian, "Distribution characteristics of pulverized coal and stress-gas pressure-temperature response laws in coal and gas outburst under deep mining conditions," *Energy Science & Engineering*, vol. 2022, 2022.

[23] Z. Wen, S. Jing, Y. Jiang et al., "Study of the fracture law of overlying strata under water based on the flow-stress-damage model," *Geofluids*, vol. 2019, 12 pages, 2019.

[24] C. Li, J. Xie, Z. He, G. Deng, B. Yang, and M. Yang, "Case study of the mining-induced stress and fracture network evolution in longwall top coal caving," *Geomechanics and Engineering*, vol. 22, no. 2, pp. 133–142, 2020.

[25] Z. Jia, H. Xie, R. Zhang et al., "Acoustic emission characteristics and damage evolution of coal at different depths under triaxial compression," *Rock Mechanics and Rock Engineering*, vol. 53, no. 5, pp. 2063–2076, 2020.

[26] K. Xiao, Z. Zhang, R. Zhang et al., "Anisotropy of the effective porosity and stress sensitivity of coal permeability considering natural fractures," *Energy Reports*, vol. 7, pp. 3898–3910, 2021.

[27] D. Lang, X. Wu, Y. Wu, H. Lin, and S. Luo, "Boundary distribution of top-coal limit-equilibrium zone in fully mechanized caving in steeply dipping coal seams," *Geomatics Natural Hazards & Risk*, vol. 12, no. 1, pp. 2561–2589, 2021.

[28] J. Wang, W. Wei, and J. Zhang, "Theoretical description of drawing body shape in an inclined seam with longwall top coal caving mining," *International Journal of Coal Science & Technology*, vol. 7, no. 1, pp. 182–195, 2020.

[29] M. T. Bui, V. Nguyen, P. H. Nguyen, N. D. Vo, and H. H. Do, "Front stress distribution under the impact of cutting height to caving height ratio in extra-thick longwall top coal caving technology," *Inzynieria Mineralna-Journal of the Polish Mineral Engineering Society*, vol. 1, no. 2, pp. 123–129, 2020.

[30] X. Li, S. Chen, S. Wang, M. Zhao, and H. Liu, "Study on in situ stress distribution law of the deep mine: taking Linyi mining area as an example," *Advances in Materials Science and Engineering*, vol. 2021, 11 pages, 2021.

[31] X. Li, S. Chen, Q. Zhang, X. Gao, and F. Feng, "Research on theory, simulation and measurement of stress behavior under regenerated roof condition," *Geomechanics and Engineering*, vol. 26, no. 1, pp. 49–61, 2021.

[32] S. Liu, X. Li, D. Wang, and D. Zhang, "Investigations on the mechanism of the microstructural evolution of different coal ranks under liquid nitrogen cold soaking," *Energy Sources Part a-Recovery Utilization and Environmental Effects*, pp. 1–17, 2020.

[33] C. Hu, W. Liu, and Y. Wang, "Study on the influence of the location of dirt band on top coal caving property in extra-thick coal seam," *Geotechnical and Geological Engineering*, vol. 38, no. 6, pp. 6221–6230, 2020.