

# Risk Assessment of Coal Mine Water Inrush Based On PCA-DBN

Zhang ye

China University of Mining and Technology

Tang Shoufeng (✉ [sieedksft@cumt.edu.cn](mailto:sieedksft@cumt.edu.cn))

China University of Mining and Technology

Shi Ke

China University of Mining and Technology

---

## Research Article

**Keywords:** deep belief network, principal component analysis, BP neural network, feature extraction

**Posted Date:** July 9th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-674596/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Risk assessment of coal mine water inrush based on *PCA-DBN*

Zhang ye, Tang Shoufeng<sup>1</sup>, Shi Ke

*School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221000, China*

## Abstract

To provide an effective risk assessment of water inrush for coal mine safety production, a BP neural network prediction method for water inrush based on principal component analysis and deep confidence network optimization was proposed. Because deep belief network (DBN) is disadvantaged by a long training time when establishing a high-dimensional data classification model, the principal component analysis (PCA) method is used to reduce the dimensionality of many factors affecting the water inrush of the coal seam floor, thus reducing the number of variables of the research object, redundancy and the difficulty of feature extraction and shortening the training time of the model. Then, a DBN network was used to extract secondary features from the processed nonlinear data, and a more abstract high-level representation was formed by combining low-level features to find the expression of the nonlinear relationship between the characteristics of water inbursts. Finally, a prediction model was established to predict the water inrush in coal mines. The superiority of this method was verified by comparing the prediction of the actual working face with the actual situation in typical mining areas of North China.

**Keywords** deep belief network, principal component analysis, BP neural network, feature extraction

## Introduction

Deep mining of the Carboniferous Taiyuan Formation system is carried out in most North China-type coal mines. Because of its proximity to Ordovician limestone, water inrush accidents often occur. Therefore, the prediction of coal mine water inrush is a necessary part of coal mine safety production. The research and development of coal mine water inrush prediction are based on research on the mechanism of coal seam water inrush. Through this research on the mechanism of water inrush, many scholars in China and abroad have determined an index system for water inrush and defined the related influencing factors of water inrush. Regression analysis (Shi and Han.2004; DU Chunlei, et al.2014; Liu Weitao et al. 2015; Liu Zaibin et al.2009), classification technology (Liu Zaibin et al.2011), geographic information systems (Li Fengjun and Zheng Jidong.2010), support vector machines (Ma. D et al.2018; Shi

---

<sup>1</sup> Shoufeng Tang

[sieedksft@cumt.edu.cn](mailto:sieedksft@cumt.edu.cn)

School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221000, China

L et al.2017; Cao Qingkui and Zhao Fei.2011; Yan Zhigang et al.2008), neural networks (Qiao Yufeng. 2011), extreme learning machine (Zhao Zuopeng and Hu Mengke.2014; Zhao Z et al.2013) and other data analysis algorithms are applied to the prediction of coal mine water inrush, and the probability of water inrush accidents is analysed and evaluated, which provides data support for coal mine safety production.

### **Analysis of the influencing factors of water inrush**

The occurrence of water inrush accidents in coal mines is the result of the joint action of many influencing factors. The interactions between the influencing factors that contribute to the water inrush accidents form a nonlinear system, which cannot be accurately expressed by the classic mathematical model. In China, the study of water inrush law began in the 1960s. The water inrush coefficient method was proposed by the Ministry of Coal, and the empirical formula of the water inrush coefficient was established (Shi Longqing 2012). Professor Jing Zigang of Shandong University of Science and Technology proposed the theory of "the next three belts" (Huang Hao 2015). Dr. Liu Tianquan and Zhang Jincai of the General Institute of Coal Mine proposed the "two-zone" model, which considered that floor rock masses were composed of mining-induced water-conducting fissure zones and floor water-isolating zones (Zhang Jincai and Liu Tianquan 1990). In the 21st century, Professor Shi Longqing of Shandong University of Science and Technology put forward the theory of the "Lower Four Belts" based on the theory of "Lower Three.

Belts" (Shi Longqing and Han Jin 2005). The Institute of Geology, Chinese Academy of Sciences, put forward the theory of a "strong seepage channel" in the 1990s, which believes that the presence of a water inrush channel is the key to the occurrence of water inrush (Duan Hongfei 2012). Qian Minggao, an academician at the China University of Mining and Technology, proposed the KS theory of key strata of stope floor rock according to the layered structure characteristics of floor rock (Qian Minggao et al 1995). It can be inferred from the above theory that the water inburst problem of the coal seam floor is mainly caused by the hydrogeological environment as the basic background and is caused by the joint actions of many factors, such as the condition of the waterproof layer, the structural condition, the condition of the aquifer and the mining method used in the coal mine. Based on the summary of previous studies, it is concluded that the main factors affecting the water inrush from the coal floor are five first-level indexes, namely, aquifer condition, water-barrier condition, coal seam condition, structure condition and mining condition, and twelve second-level indexes. The Influencing factors of coal seam water inrush are shown in table 1.

**Table 1** Influencing factors of coal seam water inrush

Aquifer conditions	Aquifer water pressure
	Working face distance
	aquifer thickness
Water barrier condition	Thickness of argillaceous rock
	Thickness of Sandy Rock
	Thickness of limestone section
Coal seam condition	Coal seam thickness
	Coal seam dip angle
Structure condition	Fault fall
Mining condition	Mining area
	Strike length
	Mining height

### The theory of methods

#### PCA

PCA is a dimension reduction algorithm. The principle is that the use of multiple indicators through linear transformation converts the comprehensive indicators of several unrelated indicators to each other, and according to certain rules to classify the integration of the comprehensive index, never reduces the dimension of the original data, extracts the main information in the original data, and minimizes the information loss in the process of the dimension reduction algorithm.

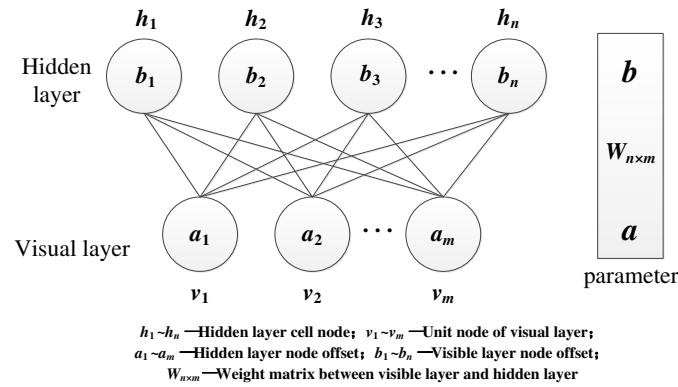
There is information overlap among the variables influencing the occurrence of water inrush, which will increase the cost and time of the classification prediction algorithm and reduce the success rate of its prediction. PCA is used to carry out dimensionality reduction processing on the original feature data, eliminate redundant information within the acceptable loss range, save the key evaluation index factors, and realize the dimensionality reduction of the evaluation index (Li Pei. 2014).

## DBN

DBN is a probabilistic generation model that is composed of a stack of several *Restricted Boltzmann machine (RBM)* and a classification or regression layer at the top. Through forward learning combined with the reverse fine-tuning mechanism of gradient descent, more accurate model training accuracy can be achieved.

## RBM

*RBM is a probabilistic abrupt model that can be explained by a stochastic neural network. In the classic RBM structure, neurons located in the same layer have no correlation with each other. This structure is developed on the basis of a Boltzmann machine (BM), which solves the shortcoming of the unacceptably slow training speed of traditional RM and improves the training speed of the network (Liu Jianwei 2014).*



**Fig 1** RBM structure diagram

RBM is composed of two layers of neurons as shown in Fig1. There is undirected full connection between different neurons, and there is no connection between neurons in the same layer. Data are input by the visual layer and output by the hidden layer after training by neurons and weight matrix.

With given a cell node  $(v, h)$ , the energy function of RBM is

$$\begin{aligned}
 E(v, h) &= -a^T h - b^T v - v^T w_{n \times m} h \\
 &= -\sum_i b_i v_i - \sum_j a_j h_j - \sum_{i,j} w_{ij} v_i h_j
 \end{aligned} \tag{1}$$

Based on the energy function, the following probability distribution under the condition  $\theta = (w_{n \times m}, a, b)$  can be obtained:

$$p_{\theta}(v, h) = \frac{e^{E_{\theta}(v, h)}}{\sum_{v, h} e^{E_{\theta}(v, h)}} \tag{2}$$

$$p_{\Theta}(v) = \sum_h p_{\Theta}(v, h) = \frac{1}{Z} \sum_h e^{E_{\Theta}(v, h)} \quad (3)$$

$$p_{\Theta}(h) = \sum_v p_{\Theta}(v, h) = \frac{1}{Z} \sum_v e^{E_{\Theta}(v, h)} \quad (4)$$

$Z$  is the normalized coefficient.

The activation probabilities of  $h$  and  $v$  are obtained after the activation function sigmoid:

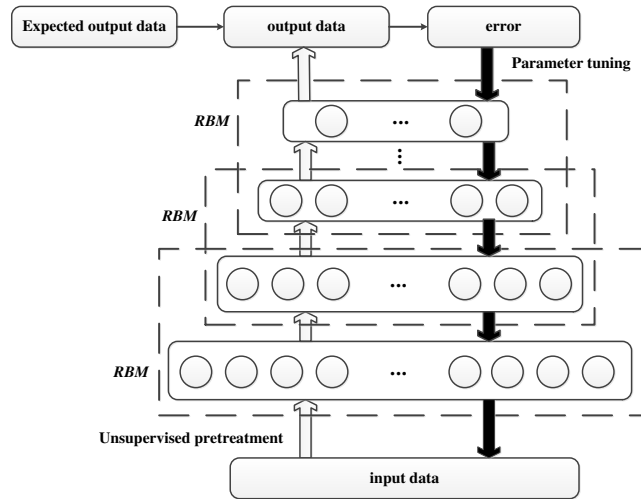
$$p(h_j = 1|v) = \text{sigmoid}(a_j + \sum_i v_i w_{ij}) \quad (5)$$

$$p(v_i = 1|h) = \text{sigmoid}(b_i + \sum_j h_j w_{ij}) \quad (6)$$

The core formula of the RBM algorithm is the activation formula of  $h$  and  $v$ . Data are input from the visual layer, and the characteristic index is mapped from the visual layer to the neurons of the hidden layer through Equation (5). Then, the output value obtained is reconstructed to the visual layer  $v$  through Equation (6), and the error between the reconstructed data in the original data domain is calculated. The weight parameters between the visible and hidden layers are adjusted by the error minimization rule so that the reconstructed data can represent the original input data to the maximum and achieve the goal of feature extraction. In fact, the goal of the training process of the RBM algorithm is to solve the Markov maximum likelihood estimation problem; that is, under the condition of fixed data input, the  $P_{\Theta}(v)$  value is maximized by adjusting the internal parameters of the RBM.

### **DBN network structure**

A DBN is composed of multiple stacked RBMs, which construct a typical DBN network model. Compared with the shallow neural network, this kind of stacked DBN structure has a deeper network level and better model generalization ability. Traditional neural networks rely on the selection of data features, while DBN can extract hidden features from input data by setting multiple hidden layers (Lecun Y et al 2015).



**Fig 2** DBN model

The DBN is composed of a cascading RBM and a back propagation algorithm adopted in the top layer as shown in Fig 2. The algorithm training process is divided into two parts: pretraining and parameter fine-tuning. Pretraining means that the input data are trained layer by layer unsupervised by the bottom RBM, and the output of the previous layer will be used as the input data of the upper-layer RBM. This structure can effectively screen out the feature information. The parameter fine-tuning process involves overall tuning and supervised training. The error between the expected data in the output data domain is propagated back layer by layer to fine-tune the parameters of the entire network (Gao Qiang and Ma Yanmei 2016).

### **PCA-DBN prediction model**

The data of water inrush accidents in coal mines are nonlinear and high-dimensional, and there are complex interrelations among the related factors of water inrush accidents. Most of the current prediction and evaluation methods cannot effectively extract a large number of hidden features from the data, resulting in the establishment of a water inrush accident model that is more one-sided, which affects the prediction accuracy and is unable to provide effective support for coal mine safety. Therefore, the design idea of the model in this paper mainly is: converting the high-dimensional influencing factors into low-dimensional data that is easy to train and more complete extraction of the characteristic quantity in the data (Wu Kai et al 2020).

### **PCA data dimension reduction**

The *PCA* algorithm is used to reduce the dimension of the main control factors of water inrush in coal mines and to standardize the data of the actual coal mine measurements. Linear transformation of data is carried out by adopting a deviation standardized formula, and the data are mapped to the interval of [0,1] to smooth the data optimization

process. According to the normalized data matrix, a covariance matrix is obtained, and the eigenvalue, principal component contribution rate (*VCP*) and cumulative variance contribution rate (*CVCP*) of the original data are calculated. Using *SPSS* software and on the basis of the testing data on the corresponding *PCA*, the main component of the selection criteria is that the cumulative variance contribution rate must be more than 80% because the value of the first to sixth principal component *CVCP* is approximately 83%. These six components already contain most of the information needed for water inrush prediction, so the first six components are used for floor water inrush evaluation. The *CP* and *CVCP* of principal components are shown in table 2.

**Table 2** *VCP* and *CVCP* of principal components

Component	<i>vcp</i> (%)	<i>cvcp</i> (%)
1	32.38	32.38
2	15.67	48.05
3	13.52	61.57
4	10.01	71.58
5	6.02	77.6
6	5.82	83.4

***PCA-DBN model training***

The data after *PCA* dimension reduction are input into the *DBN* for pretraining. The pretraining initializes the weight matrix between each layer and traverses the input vector and the neuron nodes of the hidden layer. After the training of the first *RBM*, the output neuron parameters are taken as the input vector of the second *RBM* and are then transmitted layer by layer to the highest level. According to the results of the pretrained output layer and the error between the output layer and the expected output, the parameters of each layer are updated by backward propagation from each output layer to the hidden layer (Chen Kai et al. 2017).

The advantage of the model is that the output data features are abstracted by the *DBN*, and the neural network is used as the top unit of the *DBN* to predict water inrush after new features are extracted. The prediction process of the model is as follows:

A. Establish a 4-layer *RBM* network. The number of nodes in the input layer is determined by the data dimension, and the number of nodes in the hidden layer is obtained by the "trial and error" method.



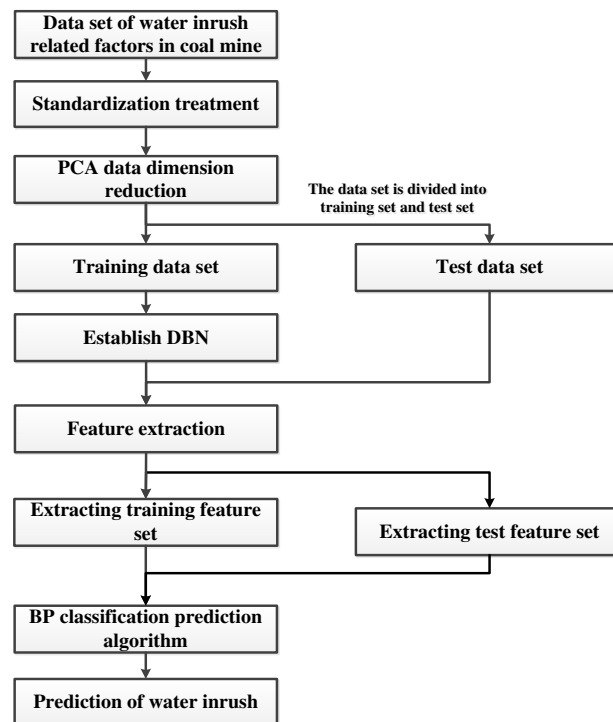
B. Input the training data without labels after dimension reduction by PCA into the DBN network, pretrain the RBM parameters layer by layer, and locally optimize the RBM parameters.

C. Input the training data with labels, back-propagate the errors layer by layer, and update the DBN network weights using the gradient descent method until convergence.

D. Input all the data into the network for feature learning and extract the output reconstructed feature data.

E. The data are divided into test and training data. The training data with labels are input into BP for training, and the trained network is used to predict the test set, compare the predicted water inburst situation with the actual situation, and evaluate the predicted results.

The algorithm flow chart is shown in Fig 3.



**Fig 3** algorithm flow chart

### Algorithm verification

The proposed method is used to test the water inrush data of typical mining areas in North China, and the PCA-DBN and BP algorithms are used to predict and compare the water inrush situations. The number of samples used in these three modelling methods is the same, with 100 groups, among which 80 groups are used as training sets and 20

groups are used as test sets. After processing the original data by PCA, the dimension of the eigenvalues is reduced to 6.

**Table 3** Partial data after PCA dimensionality reduction

Number	s1	s2	s3	s4	s5	s6	result
1	0.117647	0.435	0	0.23	0	0.4307	0
2	0.647059	0.675	0	0	0	0.779434	1
3	0.235294	0.495	0.590164	0.17	0	0.38152	0
4	0.352941	0.48	0	0	0	0.353204	0
5	0.117647	0.555	0.491803	0	0	0.390462	0
6	0.705882	0.53	0.655738	0	0	0.351714	0
7	0.176471	0.435	0.491803	0.18	0	0.324888	0
8	0.294118	0.615	0.606557	0.4	0.091017	0.508197	0
9	0.117647	0.45	0.491803	0	0	0.350224	0
10	0.470588	0.545	0.622951	0.39	0.134752	0.415797	1
11	0.176471	0.56	0.606557	0.3	0	0.470939	1
12	0.058824	0.665	0.836066	0.28	0.165485	0.488823	1
...	...	...	...	...	...	...	...
77	0.411765	0.905	0.688525	0.21	0.22695	0.535022	1
78	0.352941	0.97	0.639344	0	0	0.350224	1
79	0	0.755	0.737705	0.46	0	0.627422	0
80	0.529412	0.725	0.655738	0.46	0.153664	0.564829	1

The data from Table 3 are input into the PCA-DBN model, and the results are shown in Figure 4 and Table 4. As shown in the table, there are two false predictions (serial number 86,97); therefore, the correct rate is 90%. The reason for the incorrect prediction of samples may be related to the insufficient number of samples. Increasing the number of training samples in the training process can improve the accuracy of the model. At the same time, the accuracy rate of the BP neural network trained by the original features is 70%, and the accuracy rate of the water inburst coefficient method is 60%. The proposed model is superior to the two methods.

Water inrush case No	Prediction results	Actual results	Conformance to reality
81	0	0	Yes
82	0	0	Yes
83	1	1	Yes
84	0	0	Yes
85	0	0	Yes
86	0	1	No
87	1	1	Yes
88	0	0	Yes
89	0	0	Yes
90	1	1	Yes
91	0	0	Yes
92	0	0	Yes
93	1	1	Yes
94	0	0	Yes
95	0	0	Yes
96	0	0	Yes
97	0	1	No
98	0	0	Yes
99	1	1	Yes
100	0	0	Yes

**Table 4** Comparison of model results and actual results

bp using RBM features:				
	precision	recall	f1-score	support
0	0.87	1.00	0.93	13
1	1.00	0.71	0.83	7
accuracy			0.90	20
macro avg	0.93	0.86	0.88	20
weighted avg	0.91	0.90	0.90	20
bp using raw pixel features:				
	precision	recall	f1-score	support
0	0.71	0.92	0.80	13
1	0.67	0.29	0.40	7
accuracy			0.70	20
macro avg	0.69	0.60	0.60	20
weighted avg	0.69	0.70	0.66	20

**Fig 4** Prediction results of PCA-DBN and PCA-BP

The proposed model can be directly applied to water inrush forecasts in the North China coalfields, and the predicted results show that the DBN can effectively extract features. This kind of nonlinear factor affecting water inrush reflects interrelated data that have a very good pre-treatments effect and can markedly improve the BP neural network prediction effect. In conclusion, the PCA-DBN prediction model has a good prediction effect on water inrush data and can provide a more accurate water inrush risk assessment to better ensure coal mine safety.

## Conclusion

There are many risk factors affecting coal floor water inbursts, and some data are redundant. Principal component analysis reduces the data dimension without damaging the integrity of the data and saves the cost of the training algorithm. By training relative to the original features of PCA and BP, the PCA-DBN model is more effective for extracting the characteristics of water inrush that influence the original data, improving the training accuracy and generalizing the performance of the model. As a result, the PCA-DBN model can eliminate the defects of traditional algorithms for feature selection, extract implicit characteristics in complex hydrogeological information, and effectively filter the missing and noise data to establish a more reliable evaluation model for water inrush accidents. The case analysis shows that the predicted value of the model is consistent with the actual situation of water inrush in coal mines, and the following conclusions are drawn:

(1) The multidimensional redundant input data will complicate the structure of the DBN. PCA is used to reduce the dimensionality of the data, extract the nonlinear features of the high-dimensional data, and input them into the deep confidence network, which can simplify the network structure and improve the accuracy of the model.

(2) Compared with the traditional BP network, the PCA-BP network model and the water inburst coefficient method, the PCA-DBN model proposed in this paper has the highest prediction accuracy. In subsequent research, the network model can be optimized from the structure of the DBN network itself, and other algorithms can be integrated to further improve the model's accuracy.

### **Acknowledgements**

This work was supported by the National Key R&D Program of China (2016YFC0801800). The authors gratefully acknowledge the anonymous reviewers who substantially improved the manuscript.

### **References**

- Cao Qingkui, Zhao Fei (2011) Prediction of water inrush from coal floor based on genetic-support vector regression. *Journal of Coal* 36(12):2097-2101
- Chen Kai, Zhou Xincong, Fang Junqiang, et al (2017) Fault feature extraction and diagnosis of gearbox based on EEMD and deep briefs network. *International Journal of Rotating Machinery* 40(1):119
- Du Chunlei, Zhang Xueying, LI Fenglian (2014) Application of improved CART algorithm in prediction of water inrush from coal seam floor. *Industry and Mine Automation* 40(12):52-56
- Duan Hongfei (2012) Study on Method of Water Inrush Risk Assessment of Mining above Confined Aquifer. China University of Mining and Technology, Xuzhou (in Chinese)
- Gao Qiang, Ma Yanmei (2016) Research and Application of the Level of the Deep Belief Network. *Science Technology and Engineering* 16(No.23):234-262
- Huang Hao, Wang Jingming (2015) Research on water inrush from the blind fault of coal floor by physical experiment. *Journal of North China Institute of Science and Technology (Natural Science)* 12(1):11-16
- Lecun Y, Bengio Y (2015) Hinton G. Deep learning. *Nature* 521(14539):436 .
- Liu Jianwei, Liu Yuan, and Luo Xionglin (2014) Research and Development on Boltzmann Machine 50(1):1-16
- Liu Weitao, Liao Shanghui, Liu Shiliang, et al (2015) Principal component logistic regression analysis in application of water outbursts from coal seam floor. *Journal of Liaoning Technical University* 34(8):905-909
- Liu Zaibin, Jin Dewu, Liu Qisheng (2009) Prediction of water inrush from seam floor based on binomial logistic regression model and CART tree. *Coal Geology & Exploration* 37(1);56-61
- Liu Zaibin, Jin Dewu, Liu Qisheng (2011) Prediction of water inrush through coal floors based on data mining classification technique. *Procedia Earth & Planetary Science* 3:166-174
- Li Fengjun, Zheng Jidong (2010) The prediction and forecast of coal floor water-inrush based on GIS: A case study on the I-1 mining district in the 5# coal mine in Pingdingshan area. *International Conference on Environmental Science and Information Application Technology. IEEE 2010:663-666.*
- Li Pei (2014) The Application of Forecasting Model Based on PCA - ELM in Coal Mine Water Burst Prediction. China University of Mining and Technology, Xuzhou (in Chinese)

- Ma, D, Duan, H, Cai, X, Li, Z, Li, Q, Zhang, Q. A (2018) Global Optimization-Based Method for the Prediction of Water Inrush Hazard from Mining Floor. *Water*, 10(11):1618
- Qian Minggao, Miao Xiexing, Li Liangjie (1995) Mechanism of mail of the fracture behaviours floor in longwall mining. *Chinese Journal of Geotechnical Engineering*, 17(6): 56- 61
- Qiao Yufeng (2010) Application Research of Genetic Algorithm and Artificial Neural Networks in the Prediction of Mine Water Gushing-out. Xi'an University of Architecture and Technology, Xi'an(in Chinese)
- Yan Zhigang, Bai Haibo, Zhang Hairong (2008) A novel SVM model for the analysis and prediction of water inrush from coal mine. *Journal of Chinese Security Science*, 18(7):166–170.
- Shi LQ, Han J (2004) Floor water inrush mechanism and prediction. China University of Mining and Technology Press, Xuzhou,
- Shi L, Gao W, Han J, et al (2017) A Nonlinear Risk Evaluation Method for Water Inrush Through the Seam Floor. *Mine Water and the Environment* 34(4):1-9
- Shi Longqing, Han Jin (2005) Theory and practice of dividing coal mining area floor into four-zone. *Journal of China University of Mining & Technology* 42(1):16-23
- Shi Longqing (2012) Analysis of water inrush coefficient and its applicability. *Journal of Shandong University of Science and Technology* 31(6):6–9.
- Shi Longqing, Tan Xipeng, Wang Juan, et al (2015) Risk assessment of water inrush based on PCA-Fuzzy-PSO-SVC. *Journal of China coal society* 40(1):167-171
- Wu Kai, Liu Xinzhong, Zhang Xiaoxiong, Miao Yu (2020) Feature extraction of hot strip rolling data based on PCA-DBN[J]. *Metallurgical Industry Automation* 44(3):21-25
- Zhao Zuopeng (2014) HU Mengke. Multi-level forecasting model of coal mine water inrush based on self-adaptive evolutionary extreme learning machine. *Applied Mathematics & Information Sciences Letters* 2(3):103–110
- Zhao Z, LI P, XU X (2013) Forecasting model of coal mine water inrush based on extreme learning machine. *Applied Mathematics & Information Sciences* 7(3):1243–1250
- Zhang Jincai, Liu Tianquan (1990) On depth of fissured zone in seam floor resulted from coal extraction and its distribution characteristics. *Journal of China Coal Society* 2:35-38