Mine Water and the Environment (2023) 42:220–229
https://doi.org/10.1007/s10230-023-00932-8

TECHNICAL ARTICLE

A Practical Method for Floor Water Inrush Prediction Using a Hybrid Artificial Intelligence Model and GIS

Weitao Liu[1,2] · Mengke Han[1]

Received: 23 November 2022 / Accepted: 9 May 2023 / Published online: 21 June 2023

Abstract
Predicting floor water inrush has become increasingly challenging as coal is being mined at greater depths. We established a practical predictive method using a hybrid artificial intelligence (AI) model and geographic information system (GIS) techniques. The hybrid model is a classifier that mixes a back propagation neural network (BPNN) with an adaptive boosting algorithm (AdaBoost). To assess the effectiveness of the model, 33 borehole data points with known water inrush results in the Yangcheng coal mine were used as data samples to train and test the model. The outcomes demonstrated a predictive accuracy of 100%, far exceeding the accuracy and stability of the BPNN classifier alone using the same parameters. Then, GIS techniques were used to extend the approach throughout the mining region; the greatest risk was shown to be in the middle of the area. Given the limited data set, errors may exist in extending the risk predictions for the entire mining area, so more data needs to be collected to ensure the accuracy of subsequent predictions. Still, we believe that the methods and steps adopted in this study can be used to create practical predictive models in different mining regions.

Keywords Back propagation neural network · Adaptive boosting algorithm · Geographic information system · Mining

Introduction

China has been mining at greater depths in recent years, which has caused a number of engineering catastrophes, particularly due to the hazard posed by high-pressure Ordovician limestone water. A floor water inrush is caused by water in a confined aquifer suddenly passing through the water-resisting rock stratum between the mining coal seam and the confined aquifer and entering the mining space. The rock mass in deep mines is usually in complex geological environments, which increases the difficulty of predicting floor water inrush and results in frequent water inrush accidents. These pose a severe threat to the social economy and people's safety (Wu 2014), so accurately predicting floor water inrush events is of great concern.

Many experts and academics have applied various theories and methods to predict floor water inrushes. These include the water inrush coefficient method (Shi 2012), water inrush probability index method (Shi et al. 1995), fuzzy comprehensive evaluation method (Xu et al. 2018), vulnerability index method (Wu et al. 2007), improved vulnerability index method (Liu et al. 2021; Wu et al. 2009), set pair analysis method (Li et al. 2017), Bayesian network (Dong et al. 2012), the logistic regression analysis method (Liu et al. 2015), and many more. All of these studies have helped provide guidance for safer mine production and water disaster prevention and control. However, these methods typically require extensive geological field data and relatively complex calculation processes, and some are subjective and lack relevance.

Currently, artificial intelligence is being applied in many different research fields. Rafiei Sardoi et al. (2021) used the support vector machine (SVM), random forest (RF), and enhanced regression tree to assess urban flood risk. Zeng et al. (2022) used the SVM optimized by the genetic algorithm (GA) and RF algorithm. Lin et al. (2021) used the improved genetic algorithm and extreme learning machine

✉ Mengke Han
hmksdust@163.com

Weitao Liu
wtliu@sdust.edu.cn

1 College of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao 266590, China

2 College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266590, China

method to accurately distinguish the sources of water inrush. Moreover, artificial intelligence algorithms have also been used for water quality prediction (Egbueri 2022; Egbueri et al. 2022; Egbueri and Agbasi 2022a, b, c).

Because of the complexity of geological conditions, floor water inrush frequently involves many variables, and the relationship between these variables and prediction results is highly nonlinear. This makes it a potentially appropriate application for an artificial intelligence algorithm model. The advantages of this approach are that it is not necessary to determine the weight of variables, that the influence of human factors is reduced, and that it can more accurately describe the complex nonlinear relationship between variables and water inrush. In addition, the geographic information system (GIS) has been recognized as the primary tool for monitoring and analyzing groundwater disasters (Wu et al. 1996). Therefore, this paper proposes a two-step practical method based on a hybrid artificial intelligence model and GIS techniques to predict floor water inrushes.

Methods

The Hybrid Artificial Intelligence Model

The BPNN Algorithm

The back propagation neural network (BPNN) (Rumelhart et al. 1986) is a widely used machine-learning model, a multilayer feedforward neural network trained using the error back propagation algorithm. The network structure of BPNN includes the input layer, output layer, and multiple hidden layers (Xia et al. 1997), as shown in Fig. 1. The following are the main BPNN steps:

(1) Initializing the network structure. The number of nodes in each layer—output layer nodes m, input layer nodes n, and hidden layer nodes l—is determined by the input and output dimensions of the dataset. Initialize connection weight ω_{ij}, ω_{jk}, hidden layer threshold a, output layer threshold b, given neural excitation function f and learning rate η.

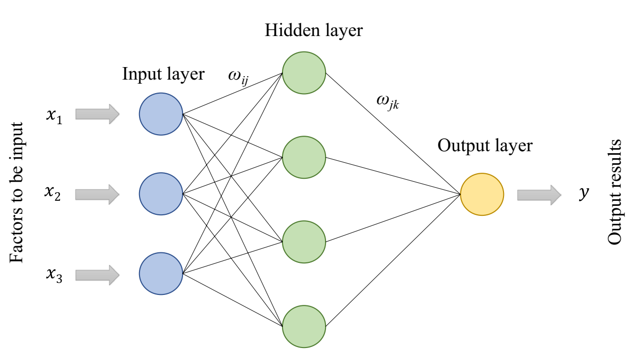

Fig. 1 BPNN structure diagram

(2) Output calculation of the hidden layer:

$$H_j = f\left(\sum_{i=1}^{n} \omega_{ij} x_i - a_j\right) \; i = 1, 2, \ldots, n \; j = 1, 2, \ldots, l \tag{1}$$

where x_i is the input variable, H_j is the output of hidden layer, f is Sigmoid function, and $f(x)$=1/(1+exp(−x)).

(3) Output calculation of the output layer:where O_k is the output of the output layer.

$$O_k = \sum_{j=1}^{l} H_j \omega_{jk} - b_k \; k = 1, 2, \ldots, m \tag{2}$$

(4) Output error calculation:where Y_k is the expected output and e_k is the output error.

$$e_k = Y_k - O_k \; k = 1, 2, \ldots, m \tag{3}$$

(5) Weight update:

$$\omega_{ij} = \omega_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^{m} \omega_{jk} e_k \quad i = 1, 2, \ldots, n \; j = 1, 2, \ldots, l \tag{4}$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \; j = 1, 2, \ldots, l \; k = 1, 2, \ldots, m \tag{5}$$

(6) Threshold update:

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^{m} \omega_{jk} e_k \; j = 1, 2, \ldots, l \tag{6}$$

$$b_k = b_k + \eta e_k \; k = 1, 2, \ldots, m \tag{7}$$

(7) Determine if the algorithm has completed its iteration. If not, go back to step 2.

BPNN-AdaBoost Algorithm

AdaBoost, an ensemble technique with adaptive boosting capabilities (Liu et al. 2019), can optimize weak classifiers by repeatedly adjusting the weightings. In essence, it is an online allocation algorithm that can integrate several low-accuracy models into a highly accurate model with a strong classification ability. Moreover, this high-accuracy model maintains good generalization capability and rarely over fits phenomena, even in the case of large training times, so it is highly applicable to small samples and high-dimensional data.

We used the AdaBoost algorithm with adaptive adjustment ability to optimize BPNN, allowing improved predictions. The BPNN-AdaBoost model takes several BPNNs as weak classifiers. With the enhancement advantage of AdaBoost, a strong classifier BPNN-AdaBoost was obtained after training BPNN repeatedly. The BPNN-AdaBoost algorithm's fundamental phases are as follows (Yan et al. 2019): given the sample space *(x, y)*, *h* groups of training samples are found, and the weight of each group is *1/h*. Then using the BPNN algorithm to run iteratively for *t* times, the weight distribution of the training sample is updated according to the classification results after each operation. Individuals that fail the classification will be given greater weight so that the next iteration pays more attention. The BPNN classifier gains a sequence of classification functions $p_1, p_2, \ldots, p_t$ through iteration and assigns a weight to each classifier function. The corresponding weight increases with the quality of the classification results. After *t* iterations, the BPNN-AdaBoost classification function *P* is obtained by weighting the BPNN classification function. The principle of the algorithm is shown in Fig. 2.

BPNN-AdaBoost Prediction Model

To enhance the performance of prediction, we combined the two algorithms to establish a hybrid artificial intelligence model—the BPNN-AdaBoost prediction model. These are the primary steps:

(1) Initialization: Set the weak classifier as BPNN, and the number is *T*. *h* groups of training samples are selected from the water inrush sample dataset. The distribution weights of the training samples are initialized to be $D_t(i) = 1/h$.

(2) Weak classifier prediction: Train the *t-th* weak classifier to obtain the prediction errors sum e_t of the weak prediction sequence f_t.

$$e_t = \sum_i D_t(i) i = 1, 2, \ldots, h \ (f_t \neq y) \tag{8}$$

where *y* is the expected output.

(3) Weight calculation of *t-th* BPNN weak classifier:where Z_t is the normalization factor, the goal is to set the sum of the distribution weights to 1 while the weight proportion remains unchanged.

$$q_t = \frac{1}{2} ln(1 - e^t/e^t) \tag{9}$$

(4) Weight adjustment of training samples: The weight of the subsequent training round should be adjusted using Eq. (10):where Z_t is the normalization factor, the goal is to set the sum of the distribution weights to 1 while the weight proportion remains unchanged.

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times exp[-q_t y_i f_t(x_i)] \ i = 1, 2, \ldots, h \tag{10}$$

(5) Strong classification function. After the *t*-cycle, *t* groups of BPNN classification functions $p\ (f_t, q_t)$ are obtained, and the *t* groups $p\ (f_t, q_t)$ are weightily combined to get the BPNN-AdaBoost classification function *P (x)*.

$$P(x) = sign[\sum_{t=1}^{T} q_t p(f_t, q_t)] \tag{11}$$

GIS Techniques

The BPNN-AdaBoost model uses borehole datapoints to fit and define the prediction at the point level; additional tools are needed to give the prediction spatial representation. GIS techniques were combined with the BPNN-AdaBoost model to predict the risk of water inrush throughout the mining region. As suggested by Wu et al. (2013), we used the strong spatial information processing function of GIS to quantify the main control factors and generate the thematic layer maps; we then extracted the relevant data of the main control factors from the map to provide the necessary input data for the BPNN-AdaBoost model. The trained model was used to process the input data to obtain the output results, which was then processed using the GIS interpolation analysis function; the results can then be shown graphically.

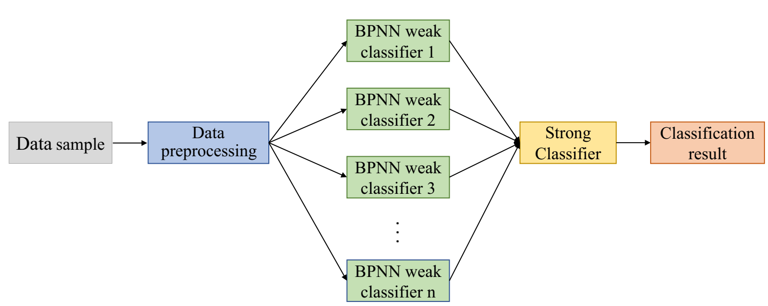

Fig. 2 BPNN-AdaBoost algorithm principle

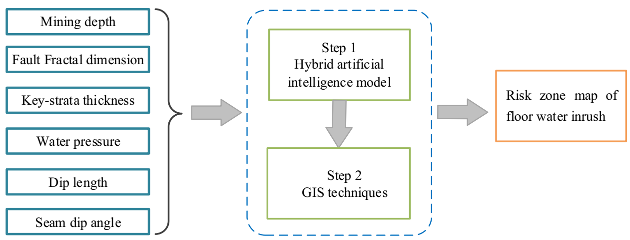

Fig. 3 Flow of the two-step practical method

The Two-step Practical Method Based on the Hybrid Artificial Intelligence Model and GIS Techniques

The practical method can be divided into two steps (Fig. 3). Specifically, in the first step, the hybrid artificial intelligence model is used to predict the risk of floor water inrush at the point level. The BPNN model is a weak classifier, so several BPNN classifiers are integrated with the AdaBoost algorithm to create a BPNN-AdaBoost strong classifier. The input variables of the predictive model are the six main control factors that influence floor water inrush: mining depth, fault fractal dimension, key-strata thickness, water pressure, dip length, and seam dip angle, and the corresponding output variables are the predicted water inrush results. Since the input variables contain the fault fractal dimension, the influence of the faults on the water inrush is considered. From the point of view of whether the floor contains faults or not, the model can be used to predict water inrush from a complete floor and from a floor with faults. The BPNN-AdaBoost prediction model was trained and tested using borehole data points with known water inrush results from the deep Yangcheng coal mine. An analysis of the testing results revealed that the BPNN-AdaBoost model accurately predicted the risk of floor water inrush. The second step was to expand the prediction using GIS techniques to expand the prediction created by the BPNN-AdaBoost.

Experiments and Results

The Study Area

The entire Yangcheng mining area was taken as the study area (Fig. 4). The boundaries of the study area in the east, northeast, south, northwest, and north directions are the F1, F2, F3, and F4 faults, and the Wensi branch faults, respectively. The western boundary is the coal seam outcrop. The distribution of faults is complicated due to a multi-stage structure's influence. There are 124 relatively well developed faults in the study area, of which 46 are large and medium-sized faults. Three principal aquifers impact mining. First, the lower Quaternary sand and gravel aquifer

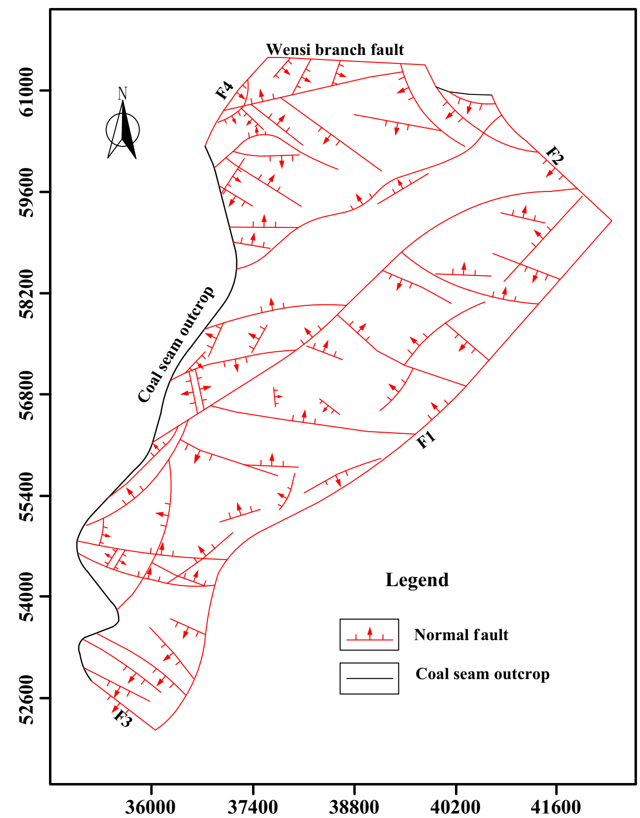

Fig. 4 Geological structures of the study area

is 197.3–337.1 m thick, which according to the lithological combination and water-bearing characteristics, can be divided into an upper aquifer and a lower aquifer. The upper aquifer has moderate water abundance, while the lower aquifer has weak water abundance. Second is the 66.05–110.00 m thick sandstone aquifer in the roof and floor of the no. 3 coal seam. The aquifer is weak in water abundance and has certain dynamic reserves, but its rechargeability is poor. Third is the 161.55–175.00 m thick Ordovician limestone aquifer, which is a regionally strong aquifer and threatens the safety of mining.

BPNN-AdaBoost Modeling for Floor Water Inrush Prediction

The four stages of the BPNN-AdaBoost modeling approach are as follows: (1) model assumptions; (1) variable selection; (2) data preprocessing; (3) model establishment; and (4) model validation and evaluation.

Model Assumptions

(1) It is assumed that faults are the only geological structures affecting floor water inrush.

(2) It is assumed that there is only "water inrush" or "no water inrush". An area with the GIS predictive result of "water inrush" is deemed to have a high risk of floor water inrush.
(3) It is assumed that the two classification results of "water inrush" and "no water inrush" are mutually exclusive.

Variable Selection

Nodes, which represent variables, are the basic element of the BPNN-AdaBoost predictive model. The number of input and output variables affects the size of the input and output nodes. In this research, the output variable is the predicted result of water inrush. The input variables should be the main control factors.

The risk of a floor water inrush is affected by many factors. The conceptual model of floor water inrush is shown in Fig. 5.Therefore, based on considering the geological, hydrogeological, and structural data of the Yangcheng coal mine and referring to the influencing factors from previous research results (Li et al. 2022; Tahershamsi et al. 2018; Yang et al. 2019; Yin et al. 2021), six main control factors were selected as the input variables for the BPNN-AdaBoost prediction model, the: mining depth (D1), fault fractal dimension (D2), key-strata thickness (D3), water pressure (D4), dip length (D5), and seam dip angle (D6). The supporting evidence is provided below.

(1) Mining depth (D1): As the mining depth increases, the accompanying mine pressure also increases, which means that the floor is more likely to be fractured, which creates water inrush channels. The probability of floor water inrush grows with deeper mining depth, making mining depth the critical factor in determining water inrush in deep mines.
(2) Fault fractal dimension (D2): The relevant factors of a fault mainly include fault distribution, fault intersection, endpoint, and so on. Because these factors are not easy to quantify mathematically, the relevant factors can be integrated by the fractal dimension method according to fractal theory and expressed as the value of the fault fractal dimension. The fault fractal dimension can be calculated using $d = lgN(s)/lgs$, where $N(s)$ is the number of grids containing fault traces in the study area and s is the length of the grid. The fault fractal dimension is a quantitative parameter reflecting fault complexity. The higher the fault fractal dimension, the more complex the fault structure, and the greater the risk of water inrush.
(3) Key-strata thickness (D3): The key strata are the one or more high-strength strata within the water-resistant strata. These strata are essential for mining because they can prevent water from an aquifer from entering the working face. The risk of water inrush increases with a decrease in the thickness of key strata.
(4) Water pressure (D4): When the groundwater pressure is great enough, water can easily break through the water-resistant strata and enter the mine, thus causing the floor water inrush.
(5) Dip length (D5): Dip length is the ratio of the working face length measured on the engineering plan to the cos a (a is the seam dip angle), which is a metric used to gage the mining space. The stress surrounding the mining face increases with the dip length, leading to more severe failure of the key strata and a greater probability of floor water inrush.
(6) Seam dip angle (D6): Seam dip angle can cause water pressure and stress differences in the working face, which can change the floor failure depth, thus affecting the floor water inrush. The larger the seam dip angle, the greater the floor failure depth, and the higher the possibility of a water inrush.

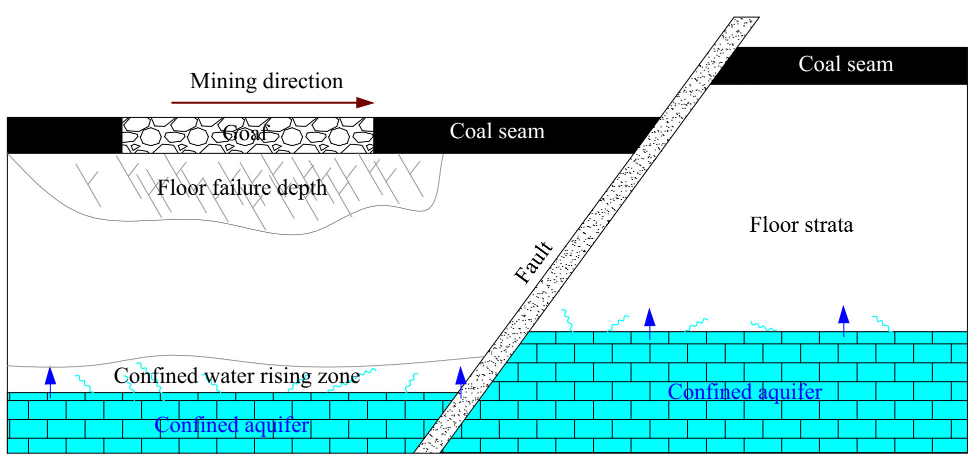

Fig. 5 Conceptual model of floor water inrush

Data Preprocessing

The 33 borehole data points (Li and Sui 2021) of the Yangcheng coal mine were used as the water inrush sample dataset for the simulation experiment, in which the value of 1 is set for "water inrush", and the value of -1 is set for "no water inrush". The sample dataset is shown in Table 1. The input variables in Table 1 (D1–D6) should be the main factors controlling the risk of floor water inrush. Because there are differences in the dimensions of sample data, direct use will affect the predictions. To eliminate these differences, the mapminmax function in MATLAB was used to normalize the input variable data in the sample dataset, producing a value range of [−1, 1]. Using the ratio of 8: 2 (Zhou 2016), the first 26 groups in Table 1 were used as training samples, and the last 7 groups were used as testing samples.

Table 1 should be placed near here during the final printing process.

Model Establishment

Based on the relevant BPNN theory, the number of input layer nodes is six, and there is one output layer node. According to the Wang et al. (2018), the number of hidden layer nodes of BPNN is optimized using the formula $l = \sqrt{m+n} + z$, where z is a constant between 1 and 10. If $z = 5$, the number of hidden layer nodes is 8. Therefore, a 6–8–1 BPNN structure was constructed. The specific parameters were established as follows: the maximum training

Table 1 Sample dataset

Groups	Input variables						Actual water inrush results
	D1 (m)	D2	D3 (m)	D4 (Mpa)	D5 (m)	D6 (°)	
1	706	0.706	56.33	0.3	175.51	10	− 1
2	680	0.733	56.05	0.3	250.24	10	− 1
3	929	0.756	30.29	0.5	200.25	12	− 1
4	915	0.879	33.12	0.5	200.65	12	− 1
5	243	0.793	65.00	0	98.58	12	− 1
6	726	0.808	37.29	0.5	182.15	10	− 1
7	750	0.796	34.77	0.5	192.84	10	− 1
8	740	0.875	40.00	0.5	192.68	10	− 1
9	570	0.884	32.84	0	143.56	15	− 1
10	887	0.927	45.56	1.2	197.55	15	− 1
11	374	0.752	33.01	0	170.26	20	− 1
12	670	0.884	38.56	2.3	198.02	20	− 1
13	925	0.927	45.66	3.6	159.53	10	1
14	956	0.740	35.50	2.3	209.38	10	1
15	547	0.915	10.50	3.4	275.34	20	1
16	262	0.656	20.00	0.3	150.02	9	− 1
17	760	0.714	28.60	2.1	183.92	20	1
18	872	0.712	32.00	0.6	180	20	− 1
19	392	0.756	10.00	3.3	200	10	− 1
20	871	0.743	58.89	1.1	108.94	15	− 1
21	390	1.062	12.81	2.5	120.5	20	− 1
22	529	0.971	12.37	3.7	200.1	20	− 1
23	901	0.680	33.42	0.3	91.47	12	− 1
24	267	0.695	38.80	0	189.82	9	− 1
25	252	0.695	25.00	0	189.82	9	− 1
26	409	0.717	27.00	1.3	190.05	11	− 1
27	769	0.979	15.00	4.2	275.42	20	1
28	519	0.718	18.63	5.2	200	25	1
29	431	0.791	34.57	1.3	300.63	11	− 1
30	699	0.614	36.60	1.2	175.47	13	− 1
31	350	0.911	35.00	0.2	347.55	10	− 1
32	671	0.831	48.80	0.5	220.05	10	− 1
33	746	0.870	64.00	1.2	250.13	10	− 1

time was 100, and the learning rate was 0.1. This BPNN-AdaBoost predictive model trained and tested the sample dataset using a strong classifier made up of 10 BPNN weak classifiers.

Model Validation and Evaluation

After the sample dataset in Table 1 were entered into the prediction model, the first 26 training samples were trained and studied (Fig. 6), and the last seven testing samples were used to predict the risk of water inrush. As seen in Fig. 7, among the seven test samples, the prediction results of the BPNN-AdaBoost model agreed well with the actual water inrush results. In this prediction, the average classification error rate of 10 BPNN weak classifiers was 12.86%. The prediction of a single BPNN weak classifier in the test was sorted; Table 2 demonstrates that the weak classifier was less accurate and less stable than the strong classifier. This shows that the strong classifier has better classification performance, accuracy, and stability than the weak classifier for the same basic parameters. It also shows that the BPNN-AdaBoost model has great statistical regularity and generalization ability under limited samples and can be used for floor water inrush prediction.

Extended Prediction Based on GIS

Because the BPNN-AdaBoost prediction model was only trained and tested based on the data of 33 boreholes, the prediction was only defined at the point level, and its spatial dimension was limited. However, GIS techniques can provide the spatial distribution results of floor water inrush

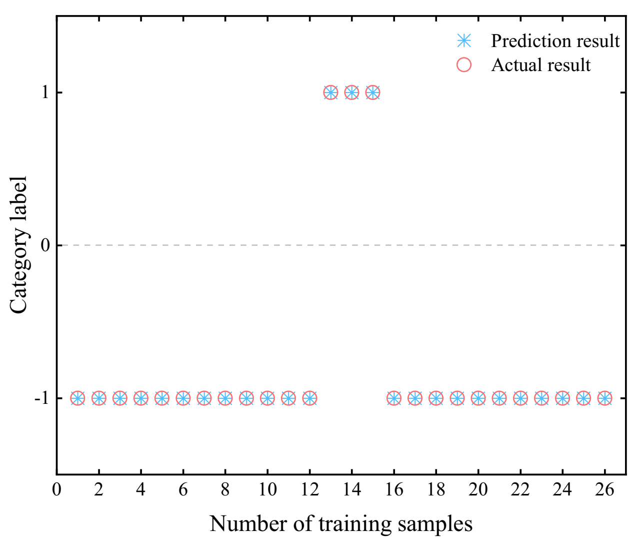

Fig. 6 Comparison between actual and prediction result of training samples

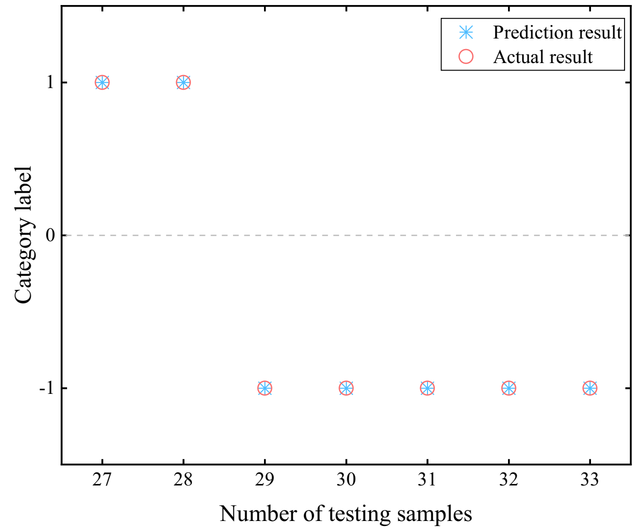

Fig. 7 Comparison between actual and prediction result of testing samples

prediction through visual maps. Using GIS techniques to expand predictions includes the following three stages: (1) establishment of a variable thematic layer map; (2) prediction by the BPNN-AdaBoost model; (3) display of the predicted results in GIS.

Establishment of Variable Thematic Layer Maps

The grid interpolation processing method in GIS was used to obtain a series of sampling points with known data, get the data of unknown points, and cover the entire mining area. The data of the six variables in Table 1 were entered into GIS and used to generate their respective thematic layer maps (supplemental Fig. S-1). This provided a data source for extracting the input data of the BPNN-AdaBoost

Table 2 Classification results of testing samples by BPNN weak classifiers

BPNN weak classifiers	Correct prediction	Wrong prediction
BPNN weak classifier 1	6	1
BPNN weak classifier 2	7	0
BPNN weak classifier 3	6	1
BPNN weak classifier 4	6	1
BPNN weak classifier 5	7	0
BPNN weak classifier 6	5	2
BPNN weak classifier 7	6	1
BPNN weak classifier 8	5	2
BPNN weak classifier 9	7	0
BPNN weak classifier 10	6	1

prediction model. At the same time, supplemental Fig. S-1 shows that D1 is larger in the east and northeast of the study area, D3 and D5 are larger in the south of the study area, and D2, D4, and D6 are larger in the middle of the study area. Generally speaking, except for D3, which is a variable to restrain water inrush, the larger the other variables, the greater the risk of a water inrush.

Prediction by the BPNN-AdaBoost Model

Through the interpolation processing described in the previous section, the data of all points in the mining area were obtained, and the data for these points were stored in six attribute tables in GIS, respectively. The lattice is uniformly distributed in the study area, with a total of 480 points (supplemental Fig. S-2). By using the overlay function of Arc Info, the thematic layer maps of six variables were compounded and superimposed, and the related data of the six variables at these 480 points was read from the attribute table of the compounded information storage layer. These data were entered into the trained BPNN-AdaBoost model and the output results at these points were calculated. An output of -1 indicates that no water inrush is anticipated at this point, while an output of 1 indicates that a water inrush is predicted to occur at this point.

Display of Prediction Results in GIS

The output results calculated by the BPNN-AdaBoost model were statistically analyzed and input into GIS to establish a relevant database. Then, the output results were processed by the grid interpolation method of GIS. The results are displayed in the form of graphics to divide the water inrush danger zone (Fig. 8). It can be seen that there is a large risk of a floor water inrush in the middle of the study area and this area needs more attention.

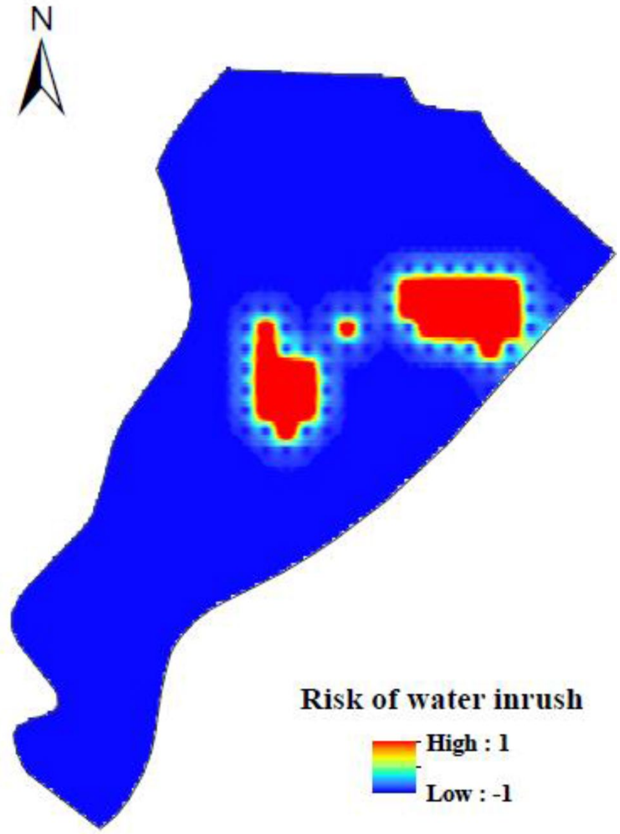

Fig. 8 Risk zone map of floor water inrush in the study area

Discussion

Misclassifying the risk of disaster events can have major economic and social impacts, so accurate model predictions are very important. In Table 1, the proportion of samples with a history of water inrush is very small, indicating that this was an unbalanced dataset. However, the existing classification studies based on BPNN ignore the imbalance of related data and can easily classify a few samples into a majority of samples, creating a high likelihood of misclassification. A successful model must have respectable overall accuracy. Our approach pays more attention to accurately identifying the water inrush cases that account for a small proportion of the inrush events. The categorization of unbalanced sample categories currently has two different types of solutions. The first focuses on data-level analysis, while the second aims to make the learning algorithm more adaptable to unbalanced samples. The major goal is to balance out the distribution of the data training sets (Guo et al. 2019); however, this procedure typically needs a lot of sample data. Alternatively, the boosting ensemble learning algorithm proposed by Schapire (1990) can transform a weak classifier into a strong classifier with a better classification level. AdaBoost is the most well-known representative of a boosting algorithm (Rutledge 2009). In this study, we adopted the second approach, using AdaBoost to improve BPNN to establish the BPNN-AdaBoost model, which contains 10 BPNN weak classifiers, in which the BPNN structure was 6–8–1. The BPNN-AdaBoost prediction model was trained and tested using the borehole data with the known water inrush results in the study area (Table 1) as the sample dataset. The experimental results show that the model has good classification t, good statistical regularity, and generalization ability despite the condition of limited samples, and is suitable for predicting the risk of floor water inrushes in the study area.

Because the BPNN-AdaBoost prediction model can only analyze and predict at the point level, 480 data points covering the study area were extracted by GIS. The extracted data points were classified and predicted using the trained model, and then the likelihood of water inrushes in the entire study area was predicted. As shown in supplemental Fig. S-1, the high risk zone is located in the middle of the study area. From the perspective of water inrush risk management, corresponding water inrush preventive measures should be taken in the red area. However, given the limited borehole data, the variable-related data for the 480 test points in the entire study area may be inaccurate in some areas, resulting in inaccurate water inrush predictions. Although the final predictive results basically reflect the distribution of water inrush risk in the study area, there is still much room for improvement in data collection Fig. 8, supplemental Fig. S-1.

Conclusions

To better predict the risk of floor water inrush, we established a practical method based on a hybrid artificial intelligence model, the BPNN-AdaBoost prediction model, and GIS techniques, and applied it to the Yangcheng coal mine.

(1) The risk zone map of the study area obtained by GIS showed that the risk of floor water inrush was high in the middle of the study area; corresponding water prevention measures should be taken.
(2) BPAN-AdaBoost overcomes the instability of traditional BPNN and makes it more accurate in limited unbalanced samples. Therefore, this model has certain application value in predicting the risk of a floor water inrush.
(3) The research results will contribute to the prediction, assessment, and management of the risk of floor water inrush in the study area. In addition, the methods and steps adopted in this paper can be applied to create prediction models in different mining areas, which offers a fresh idea for predicting the risk of floor water inrush.
(4) Due to the limited data used to build the model, errors may exist in the risk predictions for the entire mining area. Therefore, more data needs to be collected to enrich the training set to ensure the accuracy of subsequent predictions.
(5) Due to the limited geological data in the study area, only six main control factors were considered, but in the actual productive process, there may be more factors that affect the risk of a floor water inrush. In future research, we expect to consider as many factors as possible, such as water yield and mining thickness, to modify the model.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10230-023-00932-8.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant 51874192).

Data availability The data that support the findings of this study are available from the corresponding author, [Han MK], upon reasonable request.

References

Dong DL, Sun WJ, Xi S (2012) Water-inrush assessment using a GIS-based Bayesian network for the 12–2 coal seam of the Kailuan Donghuantuo coal mine in China. Mine Water Environ 31(2):138–146

Egbueri JC (2022) Incorporation of information entropy theory, artificial neural network, and soft computing models in the development of integrated industrial water quality index. Environ Monit Assess 194(10):693

Egbueri JC, Agbasi JC (2022a) Data-driven soft computing modeling of groundwater quality parameters in southeast Nigeria: comparing the performances of different algorithms. Environ Sci Pollut R 29(25):38346–38373

Egbueri JC, Agbasi JC (2022b) Combining data-intelligent algorithms for the assessment and predictive modeling of groundwater resources quality in parts of southeastern Nigeria. Environ Sci Pollut R 29(38):57147–57171

Egbueri JC, Agbasi JC (2022) Performances of MLR, RBF-NN, and MLP-NN in the evaluation and prediction of water resources quality for irrigation purposes under two modeling scenarios. Geocarto Int. https://doi.org/10.1080/10106049.2022.2087758

Egbueri JC, Unigwe CO, Agbasi JC, Nwazelibe VE (2022) Indexical and artificial neural network modeling of the quality, corrosiveness, and encrustation potential of groundwater in industrialized metropolises, southeast Nigeria. Environ Dev Sustain. https://doi.org/10.1007/s10668-022-02687-8

Guo SK, Liu YQ, Chen R, Sun X, Wang XX (2019) Improved SMOTE algorithm to deal with imbalanced activity classes in smart homes. Neural Process Lett 50(2):1503–1526

Li Q, Sui WH (2021) Risk evaluation of mine-water inrush based on principal component logistic regression analysis and an improved analytic hierarchy process. Hydrogeol J 29(3):1299–1311

Li B, Wu Q, Duan XQ, Chen MY (2017) Risk analysis model of water inrush through the seam floor based on set pair analysis. Mine Water Environ 37(4):1–7

Li A, Ding XS, Yu ZZ, Wang M, Mu Q, Dai ZX, Li HY, Zhang B, Han TR (2022) Prediction model of fracture depth and water inrush risk zoning in deep mining coal seam floor. Environ Earth Sci 81(11):315

Lin G, Jiang D, Dong DL, Fu JY, Li X (2021) A multilevel recognition model of water inrush sources: a case study of the Zhaogezhuang mining area. Mine Water Environ 40(3):773–782

Liu WT, Liao SH, Liu SL, Liu H (2015) Principal component logistic regression analysis in application of water outbursts from coal seam floor. J Liaoning Tech Univ Nat Sci Ed 34(8):905–909 (**in Chinese**)

Liu H, Zhang XC, Zhang XT (2019) PwAdaBoost: Possible world based AdaBoost algorithm for classifying uncertain data. Knowl-Based Syst 186:104930

Liu SQ, Wu Q, Li Z, Zeng YF, Yuan QD, Yu YL (2021) Vulnerability evaluation and application of floor water inrush in mining area

with multiple coal seams and single aquifer based on variable weight. J China Univ Min Technol 50(3):587–597 (**in Chinese**)
Rafiei-Sardooi E, Azareh A, Choubin B, Mosavi AH, Clague JJ (2021) Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. Int J Disast Risk Re 66:102614
Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536
Rutledge J (2009) The top ten algorithms in data mining. J Qual Technol 41(4):441–441
Schapire RE (1990) The strength of weak learnability. Mach Learn 5(2):197–227
Shi LQ (2012) Analysis of the origin of water inrush cofficient and its applicability. J Shandong Univ Sci Technol Nat Sci 31(6):6–9 (**in Chinese**)
Shi LQ, Han J, Song Y, Zhang BP, Zhang XP, Pu CS (1995) Forecast of water inrush from mining floor with probability indexes. J China Univ Min Technol 28(5):442–444 (**in Chinese**)
Tahershamsi A, Feizi A, Molaei S (2018) Modeling groundwater surface by MODFLOW math code and geostatistical method. Civ Eng J 4(4):812–827
Wang RB, Xu HY, Li B, Feng Y (2018) Research on method of determining hidden layer nodes in BP neural network. Comput Technol Dev 28(4):31–34 (**in Chinese**)
Wu Q (2014) Progress, problems and prospects of prevention and control technology of mine water and reutilization in China. J China Coal Soc 39(5):795–805 (**in Chinese**)
Wu Q, Cao HM, Zhang ZC, Li WP (1996) Chen YF (1996) Application of geographic information system in groundwater model. Hydrogel Eng Geol 1:43–44 (**in Chinese**)
Wu Q, Zhang ZL, Zhang SY, Ma JF (2007) New practical methodology of the coal floor water bursting evaluating II: The vulnerable index method. J China Coal Soc 32(11):1121–1126 (**in Chinese**)
Wu Q, Wang JH, Liu DH, Cui FP, Liu SQ (2009) A new practical methodology of the coal floor water bursting evaluating IV: the application of AHP vulnerable index method based on GIS. J China Coal Soc 34(2):233–238 (**in Chinese**)
Wu Q, Zhang B, Zhang WD, Liu SQ (2013) A new practical methodology of coal seam floor water burst evaluation: the comparison study among ANN, the weight of evidence and the logistic regression vulnerable index method based on GIS. J China Coal Soc 38(1):21–26 (**in Chinese**)
Xia YY, Xie YM, Zhu RG (1997) An engineering geology evaluation method based on an artificial neural network and its application. Eng Geol 47(1–2):149–156
Xu X, Guo BB, Tian KY, wang GZ, (2018) Fuzzy comprehensive evaluation of coal mine water disaster risk based on combination weighting. J Catastrophol 33(2):14–18
Yan H, Zhang JX, Zhou N, Li M (2019) Application of hybrid artificial intelligence model to predict coal strength alteration during CO2 geological sequestration in coal seams. Sci Total Environ 711:135029
Yang BB, Yuan JH, Duan LH, Liu Q (2019) Using GIS and fractal theory to evaluate degree of fault complexity and water yield. Mine Water Environ 38(2):261–267
Yin HY, Zhou XL, Lang N, Zhang LF, Wang ML, Wu T, Li X (2021) Prediction model of water inrush from coal floor based on GA-BP neural network optimized by SSA and its application. Coal Geol Explor 49(6):175–185 (**in Chinese**)
Zeng YF, Mei AS, Wu Q, Hua ZL, Zhao D, Du X, Wang L, Lv Y, Pan X (2022) Source discrimination of mine water inflow or inrush using a hydrochemical field machine learning analysis and hydrodynamic field reverse tracer simulation coupling technique. J China Coal Soc 47(12):1–14 (**in Chinese**)
Zhou ZH (2016) Machine Learning. Tsinghua University Press, Beijing (**in Chinese**)