Artificial Intelligence In Source Discrimination of Mine Water: A Deep Learning Algorithm For Water Source Discrimination

Zhenwei Yang
  Henan Polytechnic University
Junchao Yue
  Henan Polytechnic University
Hang Lü
  Henan Polytechnic University
Xinyi Wang (✉ 359836588@qq.com)
  Henan Polytechnic University

Abstract

With increasing coal mining depth, the source of mine water inrush becomes increasingly complex. The problem of distinguishing the source of mine water in mines and tunnels has been addressed by studying the hydrochemical components of the Pingdingshan Coalfield and applying the artificial intelligence (AI) method to discriminate the source of the mine water. 496 data of mine water have been collected. Six ions of mine water are used as the input data set: $Na^+ + K^+$, Ca^{2+}, Mg^{2+}, Cl^-, $SO2-4$, and $HCO-3$. The type of mine water in the Pingdingshan coalfield is classified into surface water, Quaternary pore water, Carboniderous limestone karst water, Permian sandstone water, and Cambrian limestone karst water. Each type of water is encoded with the number 0 to 4. The one-hot code method is used to encode the numbers, which is the output set. On the basis of hydrochemical data processing, a deep learning model was designed to train the hydrochemical data. Ten new samples of mine water were tested to determine the precision of the model. Nine samples of mine water were predicted correctly. The deep learning model presented here provides significant guidance for the discrimination of mine water.

1. Background

With increasing coal mining depth, the source of mine water inrush becomes increasingly complex. Water inrush in mines can lead to serious disasters anywhere in the world due to the complicated hydrogeological conditions found in parts of China, which are uncommon elsewhere in the world[1]. Therefore, rapid and accurate discrimination of the source of water inrush is very important and necessary for both resuming production and rescuing miners[2].

The upcoming technological revolution has been termed Industry 4.0. Examples of the use of artificial intelligence (AI) in parameter identification of groundwater systems, management of groundwater and mine hydrogeology. Both today and in the foreseeable future, it is important to take advantage of new technological developments and innovations in source discrimination of mine water inrush. The development that has taken the world by storm in the past few years is artificial intelligence, which has been widely adopted in many fields, such as computer vision, intelligent robots, natural language processing and data mining [3]. As an important method of AI, deep learning is a hot topic in various fields because of its strong ability to automatically extract high-level representations from complex data, which has been applied widely in the fields of natural science, social science and engineering[4]. The value of source discrimination of mine water in the prevention and cure of mine water has been well established over the past several decades[5].

Hydrochemistry and mathematical methods are widely used to identify water sources in hydrogeology. The ion proportions of different aquifers in mines differ greatly, such as Na^+, K^+, Ca^{2+}, Mg^{2+}. However, even in the same aquifer, the content of hydrochemical ions has a great difference[6]. Therefore, some mathematical goelogy methods, such as Bayes and principal component analysis, are used to source identification of mine water. Characteristic ion contrast and ion proportional coefficients were applied to aquifers with distinct chemical characteristics to establish a characteristic index discrimination system[7].

Because of the artificial neural network structure, deep learning excels at identifying patterns in unstructured data such as images, sound, video, and text. As a result, deep learning is rapidly transforming many industries, including healthcare, energy, finance, and transportation. These industries are rethinking traditional business processes[8]. Therefore, the study of source discrimination of mine water with artificial intelligence is of great importance. Artificial intelligence in this paper elaborates deep learning algorithms to process the main ionic composition of groundwater to better discriminate the source of mine water[9].

The organization of the paper is as follows. Section 2 presents the geological and hydrogeological conditions of the study area. The source discrimination of mine water problems in the framework of the DNN model is introduced in detail in Section 3. The results of deep learning for the source discrimination of mine water are demonstrated in Section 5. This paper closes with some conclusions and final remarks.

2. Geological And Hydrogeological Conditions

2.1 Outline of the coalfield

The Pingdingshan coalfield (113°00-114°E, 33°30'-34°00'N), located in the central and western parts of Henan Province, northern China (Fig. 1), is the third largest coal producer in China. The coalfield is approximately 40 km long E-W and 20 km wide N-S. The Pingdingshan Coalfeld is located in the low hilly area, which is divided into eastern and western areas by the Guodishan fault. Structurally, it is a large syncline with symmetrically gently dipping limbs. The coal-bearing sediments are mostly Permian in age, comprised of sandstone,

siltstone and carbonaceous shale, which are overlain by Neogene, Paleogene and Quaternary deposits. The entire sequence is underlain by Cambrian karstic limestone (Fig. 1).

2.2 Major structures

The main coal-bearing measures are dominated by strike-parallel compressional structures. Of these, most folds and faults are concentrated within a narrow zone, known locally as a compressional zone or disturbance zone, which together with the Likou syncline has been interpreted to occur during Indosinian late Triassic orogenic compression.

2.3 Stratum

The exposed strata from old to new in the Pingdingshan coalfield are archaean metamorphic rock series, Upper Proterozoic Sinian, Lower Paleozoic Cambrian-Ordovician, Upper Paleozoic Carboniferous to Permian, Mesozoic Triassic and Cenozoic Neogene, and Paleogene to Quaternary, as shown in Fig. 2. The main coal-bearing strata in the study area are Carboniferous-Permian.

2.4 Hydrogeological background

The research area is situated in a transitional zone from a warm temperate zone to a subtropical zone, with a long-term average precipitation of 747.4 mm/year, mainly concentrated between July and September. The geomorphology in the east and south is an alluvial plain with a layer of 200 m~500 m thickness. The ground elevation is +75~80 m. With a surface elevation varying from 900 m to 1040 m, the topography is low in the southeast and high in the northwest. Influenced by the topographic features, the surface water is mainly distributed in the south and north of the mining area, that is, the Shahe River, Ruhe River, Zhanhe River and Baiguishan Reservoir. The Ruhe River and the Shahe River are perennial rivers that lie on the northern and southern margins of the study area. There are some seasonal rivers and man-made ditches, such as Zhanhe, Beigan Canal and Xigan Canal. The riverbed inserts into Cambrian limestone or Neogene marl, which has a certain replenishment effect on the groundwater of limestone in the Qikuang mine in the southwest of the Pingdingshan coalfield.

The main aquifer is a limestone aquifer of the Taiyuan Formation. On the basis of the borehole pumping test data, the water inflow per unit of karst aquifer of the Taiyuan Formation is 0.00018~0.3569 L/s m, and the permeability coefficient is 0.0076~3.047 m/day.

3. Data And Deep Neutral Network

3.1 Data

In the Pingdingshan coal mine, 496 mine water data points were collected. Due to the large amount of data, some of the data are shown in Table 2. Table 2 clearly shows that there is a variance of five orders of magnitude. Data are most valuable when you have something to compare it to, but these comparisons aren't helpful if the data is bad or irrelevant. Data standardization is about ensuring that data are internally consistent, that is, each data type has the same content and format. Standardized values are useful for tracking data that isn't easy to compare otherwise. The raw data are normalized individually according to Eq (1)

$Z_{ij} = (x_{ij}\text{-mean}(x_j))/\text{std}(x_j)$ (1)

where the subscript i means the row of the data matrix, the subscript j means the column of the data matrix, Z_{ij} represents the data after standardization, x_{ij} represents the source data, and the symbol std represents the standard deviation of related data[10].

Table 2
Hydrochemical compositions and discriminant results of the water filling aquifer (unit: mg/L. In the last column, which is groundwater type (label column), 0 represents the surface water, 1 represents pore water of Quaternary limestone, 2 represents karst water of Carboniderous limestone, 3 represents sandstone water of Permian limestone, and 4 karst water of Cambrian limestone.)

Na$^+$+K$^+$	Ca^{2+}	Mg^{2+}	Cl$^-$	SO2- 4	HCO- 3	Groundwater type
284.16	13.03	7.05	31.56	4.94	768.84	3
68.31	57.72	23.35	18.08	94.14	329.5	3
27.37	143.08	19.68	36.16	179.15	314.25	2
29.21	179.36	25.64	58.49	208.93	386.26	2
40.1	80.59	11.08	31.75	80.05	257.92	2
18.03	82.56	10.21	17.33	40.35	268.48	2
12.65	75.59	11.55	18.53	32.51	248.96	2
15.85	86.11	8.52	16.97	39.31	253.099	2
18.4	78.6	12	12	21	295.9	2
5.52	80.7	10.6	8.5	14.3	271.2	2
2.69	98.4	6.6	28.4	18.1	268.49	2
5.28	79.65	14.99	9.46	8.82	82.02	2
0.14	93.29	15.95	14.3	44.45	225.75	2
7.43	112.44	16.68	25.75	80.67	275.37	2
37.03	91.98	49.82	59.56	107.11	389.92	2
77.05	94.59	31.71	36.51	244.47	278.25	2
83.72	169.14	36.21	63.46	303.03	423.48	2
38.87	82.16	18.1	30.13	101.34	263.61	4
133.57	44.09	19.46	86.55	47.86	379.18	4
246.07	41.21	30.3	63.48	218.37	430.97	4
246.07	41.21	30.3	63.48	218.37	430.97	4
234.01	48.22	30.92	64.38	277.47	433.91	4
31.17	89.04	10.98	12.94	31.38	345.67	0
20.17	148.8	24.46	70.08	35.78	454.57	0
32.34	50	3.76	10.78	41.46	184.67	0
35.19	68.25	7.46	15.28	51.35	246.23	0
10.28	72.95	10.21	10.99	19.21	241.63	1
16.74	94.37	11.1	8.55	11.73	348.41	1
452.1	12.4	5.96	116.42	86.16	922	2
303.51	8.02	2.43	51.06	5.76	629.7	2
119.6	87.17	22.72	60.26	179.15	365.51	0
110.86	119.64	27.22	114.86	352.54	149.5	0
47.59	75.55	14.34	28.61	90.96	257	0
11.57	30.32	6.93	6.2	21.4	118.98	0

Na$^+$+K$^+$	Ca^{2+}	Mg^{2+}	Cl$^-$	SO2- 4	HCO- 3	Groundwater type
52.9	201.8	24.18	62.39	298.75	389.31	1
100.97	114.63	16.65	64.52	310.75	194.65	1
29	87.06	8.04	9.8	72.32	266.98	1
809.6	4.8	15.07	151.02	143.12	1796.42	3
1109.73	15.63	8.88	94.66	24.98	2498.77	3
1036.85	10.42	1.22	64.17	4.8	2288.25	3
284.16	13.03	7.05	31.56	4.94	768.84	3
321.34	3.8	1.38	81.52	21.13	599.42	3
68.31	57.72	23.35	18.08	94.14	329.5	3
158.24	167.13	173.4	30.49	1335.23	52.48	3
31.27	76.35	11.07	21.27	46.5	277.03	3
14.02	76.18	12.64	19.01	40.33	248.96	3
14.25	73.21	13.98	18.53	39.09	249.57	3
27.37	143.08	19.68	36.16	179.15	314.25	2
29.21	179.36	25.64	58.49	208.93	386.26	2

In the datasets, the label column is categorical data (string values). These labels have no specific order of preference, and since the data are string labels, the deep learning model cannot work on such data directly[11]. One approach to solve this problem can be label encoding, where we assign a numerical value to these labels, for example, the surface water and pore water of the Quaternary mapped to 0 and 1. However, this can add bias in our model, as it will start giving higher preference to the pore water of the Quaternary parameter as 1>0, and ideally, both labels are equally important in the datasets. To address this issue, we will use the one hot encoding technique, which will create a binary vector of length 5. Here, the label 'the surface water', which is encoded as '0', has a binary vector of [0,0,0,0,1]. As is shown in Table 3.

<div align="center">

Table 3
One-hot encoding

Natural number	One-hot encoding
0	0,0,0,0,1
1	0,0,0,1,0
2	0,0,1,0,0
3	0,1,0,0,0
4	1,0,0,0,0

</div>

3.2 Deep learning basics

A machine learning algorithm is an algorithm that is able to learn from data. As a special machine learning algorithm, most modern deep learning models are based on artificial neural networks (ANNs), which form the basis of most deep learning methods and are a class of supervised learning techniques that mimic biological neural networks (Fig. 3). ANN is built from one or more layers containing a series of neurons[12]. The weights and biases between different neurons adjust as learning proceeds with the aim of minimizing the loss between the predicted output and actual output. The training processes of the ANN are the adjustment processes of weights and biases, which are carried out by a back propagation procedure. In the procedure, the gradient descent algorithm is used to update the weights and biases of neurons by estimating the gradient of the loss function. In the process of training, weights and biases accept an adjustment proportional to the partial derivative of the loss function relative to the current weights and biases. With the increasing number of layers, the problem of vanishing gradients, however, makes ANNs hard to train[13].

Typically, when training an ANN model, we have access to a training set, we can compute some error measure on the training set, called the training error, and we reduce this training error. Thus far, what we have described is simply an optimization problem. The training and test data are generated by a probability distribution over datasets[14].

3.3 Deep learning architectures

Deep learning is a subset of machine learning where the artificial neural network comes in relation. It solves all the complex problems with the help of algorithms and its process. This idea is that the additional level of abstraction improves the capability of the network to generalize to unseen data and hence outperforms traditional ANN on data outside of the network training set. The learning process is deep because the structure of artificial neural networks consists of multiple input, output, and hidden layers. Each layer contains units that transform the input data into information that the next layer can use for a certain predictive task[15].

While indisputably powerful tools, traditional artificial neural networks (ANNs) and more classical machine learning techniques rely on developers identifying the typical features that describe the problem. In this work, a deep learning approach is applied to the problems of source discrimination of mine water inrush. Deep learning further exploits the power of ANNs by relying on the network itself to identify, extract, and combine the inputs into abstract features that contain much more pertinent information to solve the problem, that is, predicting the output, as illustrated in Fig. 4. $Na^{+}+K^{+}$, Ca^{2+}, Mg^{2+}, Cl^{-}, SO2- 4, HCO- 3. Every neuron accepts inputs from neurons on the previous layer based on linear or nonlinear activation functions (e.g., ReLU). The contents of six elements are delivered from the input layer to the output layer, where the output layer corresponds to the expectation to be predicted, which are the surface water, pore water of Quaternary limestone, sandstone water of Permian limestone, karst water of Carboniderous limestone and karst water of Cambrian limestone.

An ANN with three hidden layers and one output layer is shown in Fig. 5. Every layer constitutes a module through which one can back-propagate gradients. At every layer, we compute the total input i to every unit first, which is a weighted sum of the outputs of the units in the layer below. Then, a nonlinear function f is applied to i to obtain the output of the unit. For the sake of simplicity, the bias terms are omitted. The nonlinear functions in the hidden layer using the ANN include the rectified linear unit (ReLU) $f(z)= max(0,z)$. At the output layer, softmax is used to calculate the probability of the water source, which is commonly used in recent years[16].

At every hidden layer, we calculate the error derivative with respect to the output of every unit, which is a weighted sum of the error derivative with respect to the total inputs to the units in the layer above. Then, we convert the error derivative with respect to the output into the error derivative with respect to the input by multiplying it by the gradient of f. At the output layer, the error derivative with respect to the output of a unit is calculated by differentiating the cost function. This gives $y_l - t_l$ if the cost function for unit l is $1/2*(y_l - t_l)^2$, where t_l is the target value. Once $\partial E/\partial z_k$ is known, the error derivative for the weight w_{jk} on the connection from unit j in the layer below is just $y_j \partial E/\partial z_k$.

The Python deep learning library Keras, with a TensorFlow backend and GPU acceleration, is used to train the ANN. TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that allows researchers to push the state of the art in DL, and developers easily build and deploy DL-powered applications. The model parameters of the intelligent evaluation of the DNN model are shown in Table 4.

Table 4
Model parameters of the intelligent evaluation of the DNN model

Number	Parameter	value
1	Type of model	Sequential model
2	the number of neurons in the input layer	6
3	the number of hidden layer and neurons	3,5
4	the number of neurons in the output layer	5
5	activation function of hidden layer	ReLU
7	activation function of output layer	Softmax
8	Epoch	200
9	Learning rate	0.01
10	optimizer function	Adam
11	batch_size	10
12	dropout rate	0.5
13	error limitation	$1*10^{-4}$
14	momentum coefficient	$\eta = 0.8$

4. Results And Discussion

All of training process and result can be shown from Tensorboard, which is a browser based application that helps to visualize your training parameters (weights, biases & metrics). This shows the distribution of tensors in histograms and is used to show the distribution of weights and biases in every epoch regardless of whether they change as expected. We plot the histogram distribution of the weight for the first fully connected layer every 20 iterations. It takes an arbitrarily sized and shaped tensor and compresses it into a histogram data structure consisting of many bins with widths and counts.

The data source of distributions is the same as the histogram, which is shown in different former (Fig. 6). The distribution of weights and bias of the first layer are shown in Fig. 6((a) and (c)). The abscissa represents the training times, and the ordinate represents the range of weights. It shows the range of weight values in the training process as a whole, which is constrained to learn to the layer by optimizing its weights or the layer truly 'eats up' many errors. Almost the same number of weights have values of -0.8 to 0.8 and everything in between. There are some weights having slightly smaller or higher values, but it might not be using its full potential.

In comparison, this simply looks like the weights have been initialized using a uniform distribution with zero mean and value range -0.8 to 0.8(Fig. 6 (c) and (d)). The histogram of the layer forms a bell curve-like shape. The values are centered around a specific value, but they may also be greater or smaller than that. Each slice in the histogram visualizer displays a single histogram. The slices are organized by step; older slices are further 'back' and darker, while newer slices are close to the foreground and lighter in color. The y-axis on the right shows the step number. Most values appear close around the mean of 0, but values do range from -1.3 to 1.2. With increasing training times, the color of the curves gradually becomes lighter from back to front. There are many slices in Fig. 6(b,d), and each slice represents the frequency of the weight in the distribution of weights.

Accuracy and loss are unitless numbers that indicate how closely the classifier fits the validation training data. A loss value of 0 represents a perfect fit. The further the accuracy is from 0, the more accurate the fit. Separate loss plots are provided for the batches completed within each epoch and for the completed epochs. Ideally, the loss values reported in the epoch_loss and epoch_val_loss plots should decrease rapidly during the first few epochs and then converge toward 0 as the number of epochs increases. The epoch_val_loss plot corresponds to the 'loss' value displayed in the training model dialog as a model is being trained.

The accuracy and loss of training metric of deep neural network and BP neural network have been compared, which are drawn by Matplotlib(Fig. 7). In Fig. 7, the abscissa axis represents the times of forward calculation and back propagation, and the vertical axis represents the accuracy or loss. The blue curve represents the accuracy and loss of training metric of deep neural network, and the red curve represents the accuracy and loss of training metric of BP neural network. It can be seen that the accuracy of training metric of deep

neural network is higher than BP neural network, and the loss of training metric of deep neural network is lower than BP neural network. It means that the deep neural network do a better job than BP neural network in the source discrimination of mine water.

The probability is based on the fraction of correctly predicted values to the total number of values predicted to be in a class, which is calculated by the softmax function (Fig. 8). Ten mine water samples were inputted into the trained DNN model to test its accuracy of prediction. The data is also be inputted into BP model. The prediction result is shown in Table 5. From the table, we can see that nine water samples have been predicted correctly with the DNN model, and one water sample has been predicted incorrectly. With the BP model, four water samples were predicted correctly, and six water samples were predicted incorrectly. The prediction result can also be seen in the 3-D histogram, as shown in Fig. 9.

Table 5
Probability of sources of mine water by the BP and DNN methods

Sample	Probability of surface water(□)	Probability of pore water of Quaternary(□)	Probability of sandstone water of Permian(□)	Probability of karst water of Carboniderous limestone(□)	Probability of karst water of Cambrian limestone(□)	Predicted by DNN	Predicted by BP	Real source of mine water
NO.1	0%(0%)	0%(0%)	0%(0%)	1%(99%)	99%(1%)	□	□	□
NO.2	0%(0%)	0.5%(0.05%)	99%(99.8%)	0.5%(0.05%)	0%(0.1%)	□	□	□
NO.3	0%(0%)	0%(0%)	0.5%(63%)	99%(36%)	0.5%(1%)	□	□	□
NO.4	0%(0.6%)	0%(1.09%)	99%(57.7%)	0.7%(0%)	0.3%(40.61%)	□	□	□
NO.5	0%(1.53%)	0%(3.3%)	99%(89.1%)	0.7%(0.03%)	0.3%(6.04%)	□	□	□
NO.6	0%(0.185%)	0.8%(4.45%)	99%(46.12%)	0.2%(0%)	0%(49.245%)	□	□	□
NO.7	0%(4.22%)	0%(1.38%)	0%(62.3%)	95.8%(0.23%)	4.2%(19.45%)	□	□	□
NO.8	84.7%(0%)	5.8%(0%)	2.34%(44.76%)	1.15%(54.68%)	5.96%(0.56%)	□	□	□
NO.9	0%(0%)	0.5%(0%)	99%(34.12%)	0%(65.81%)	0.5%(0.07%)	□	□	□
NO.10	0.03%(0.14%)	52.43%(3.16%)	41.39%(84.58%)	5.89%(0%)	0.26%(1.21%)	□	□	□

5. Conclusions And Outlooks

In the research reported here, we apply deep learning methods to discriminate the source of mine water.

(1) Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. The method has dramatically improved the state of the art in the source discrimination of mine water. Deep learning discovers intricate structures in large data sets by using the back propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.

(2) On the basis of hydrochemical data processing, a deep learning model was designed to train the hydrochemical data. Ten new samples of mine water were tested to determine the precision of the model. Nine samples of mine water were predicted correctly. The deep learning model presented here provides significant guidance for the discrimination of mine water.

(3) This high predictive accuracy, combined with very low computational costs-execution of the full framework takes place on the order of milliseconds-makes the developed networks very well suited for discriminating source of mine water.

(4) However, for many practical applications, the combination of execution speed, accuracy, and ease of use are the main concerns. Due to the flexibility and predictive power of DL ANNs, several other aspects of the ANN could be approached in a similar way in the future.

Declarations

Author Contributions

Zhenwei Yang and Hang Lü designed the model. Junchao Yue collected the mine water samples. Junchao Yue and Xinyi Wang performed model training and data analysis. Zhenwei Yang and Xinyi Wang wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

References

1. Chunlu Jiang, Y., An, L., Zheng & Wangwang Huang. Water Source Discrimination in a Multiaquifer Mine Using a Comprehensive Stepwise Discriminant Method. *Mine Water and the Environment*, **40**, 442–455 (2021).
2. Renguang Zuo, Y., Xiong, J. & Wang, Emmanuel John, M. C. Deep learning and its application in geochemical mapping. *Earth Science Review*, **192**, 1–14 (2019).
3. Qiang Wu, W. *et al.* Source discrimination of mine water inrush using multiple methods: a case study from the Beiyangzhuang Mine, Northern China. *Bulletin of Engineering Geology and the Environment*, **78**, 469–482 (2019).
4. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature,2015,521,436-444.
5. Zhou, Z. Q. *et al.* An optimal classification method for risk assessment of water inrush in karst tunnels based on gray system theory. *Geomechanics and Engineering*, **8** (5), 631–647 (2015).
6. Yang, Y. Y. *et al.* Mining-induced geo-hazards with environmental protection measures in Yunnan, China: an overview.Bull Eng Geol Environ,2015,74(1):141–150
7. Yin, S. X., Zhang, J. C. & Liu, D. M. A study of mine water inrushes by measurements of in situ stress and rock failures. *Nat Hazards*, **79** (3), 1961–1979 (2016).
8. Qian, J. Z. *et al.* Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine,Anhui, China. Environ Earth Sci,2016,75(9):1–10
9. Ma, D. *et al.* Effect of mining on shear sidewall groundwater inrush hazard caused by seepage instability of the penetrated karst collapse pillar. *Nat Hazards*, **82** (1), 73–93 (2016).
10. Gu, H. Y. *et al.* Assessment of water sources and mixing of groundwater in a coastal mine:the Sanshandao gold mine, China. *Mine Water Environ*, **37** (2), 351–365 (2018).
11. Blake, A. *et al.* A deep learning framework for neuroscience. *Nature neuroscience*, **22**, 1761–1770 (2019).
12. Grob, L. *et al.* Filter-less separation technique for micronized anthropogenic polymers from artificial seawater.Environmental Science Water Research & Technology. 2021,online
13. Dauphin, Y. *et al.* Identifying and attacking the saddle point problem in highdimensional nonconvex optimization. In Proc. Advances in Neural Information Processing Systems, 2014, 27: 2933-2941
14. Tompson, J., Goroshin, R. R., Jain, A., LeCun, Y. Y. & Bregler, C. C. Efficient object localization using convolutional networks. In Proc. Conference on Computer Vision and Pattern Recognition, 2014,1411:4280
15. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting.J. Machine Learning Res., 201415,1929–1958
16. Xiong, H. Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease., **347**, 6218 (2015).
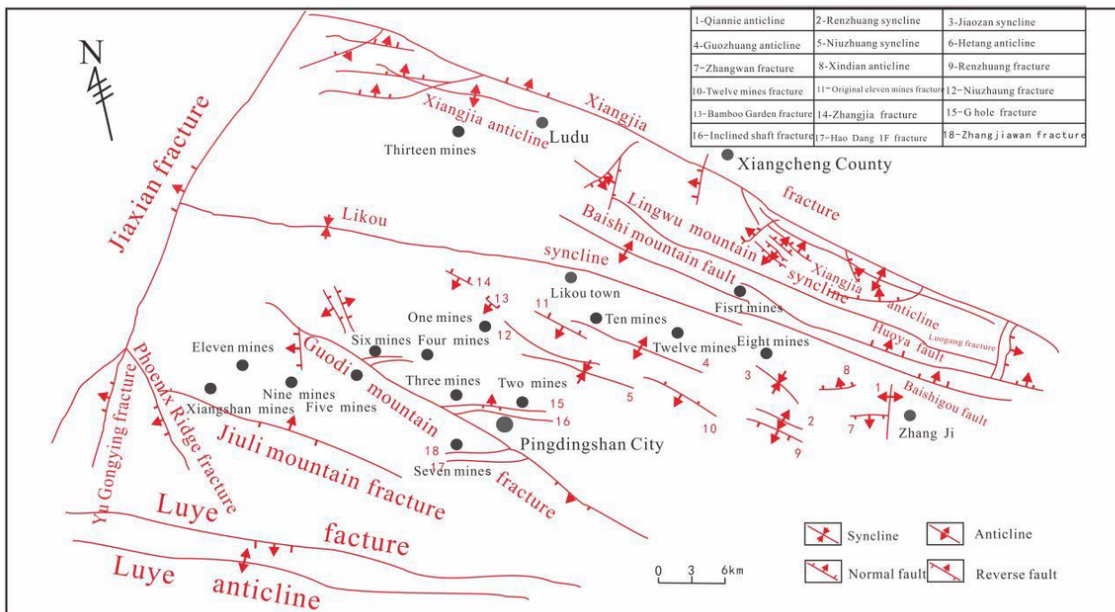
Figures

Figure 1

Structural chart of the Pingdingshan coalfield and its location in China
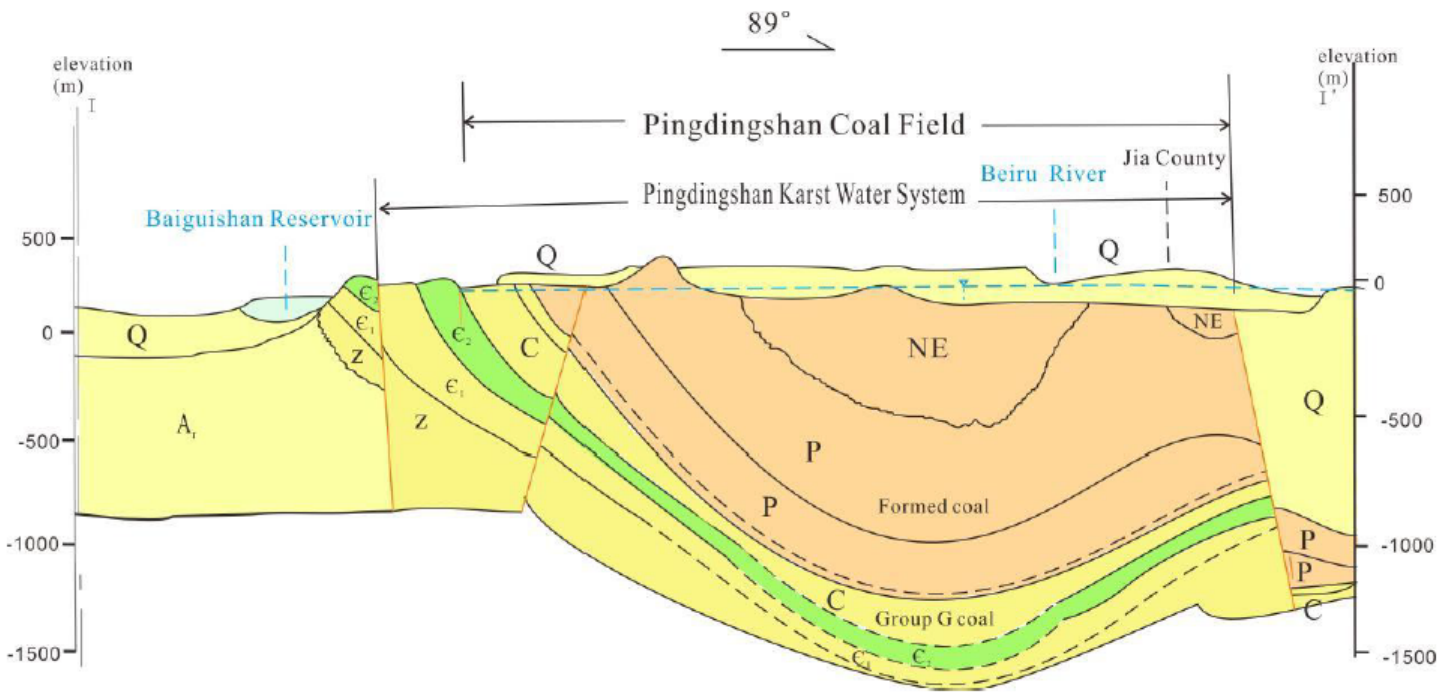
Permian.

Figure 2

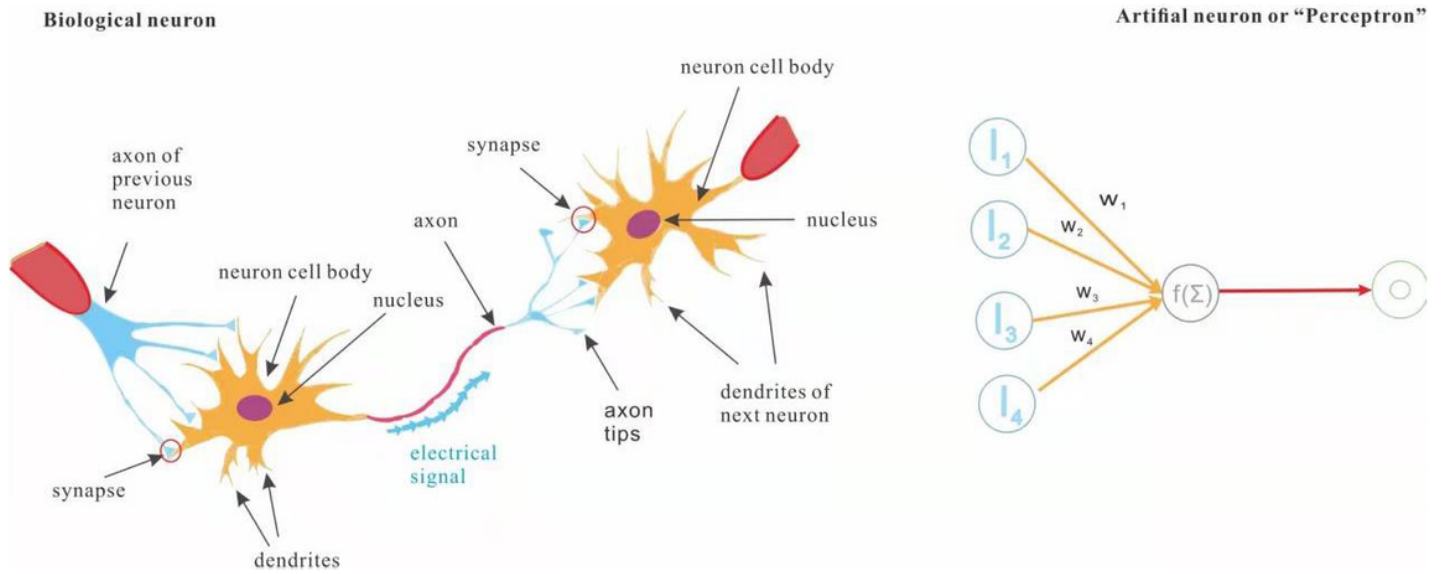The geological section of Pingdingshan coalfield

Figure 3

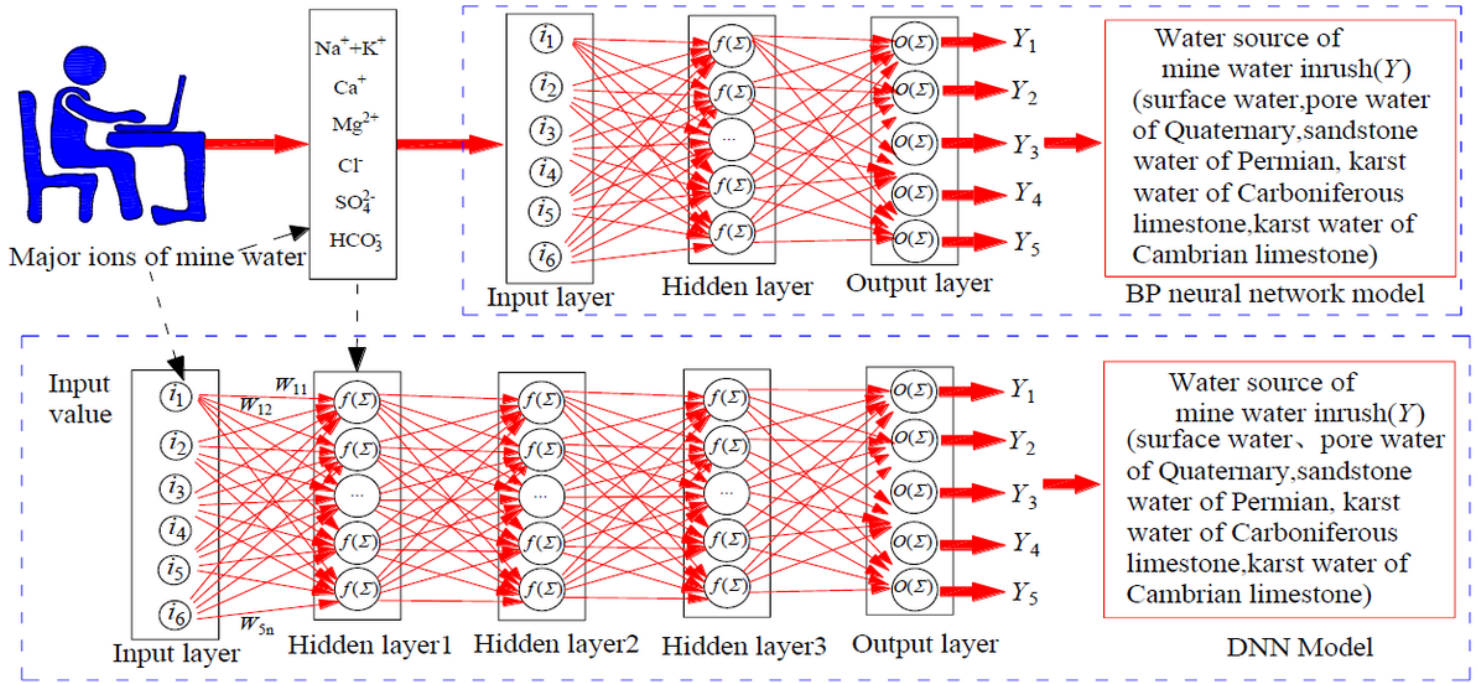Analogy of a biological neuron (left) and an artificial neural network (right)

Figure 4
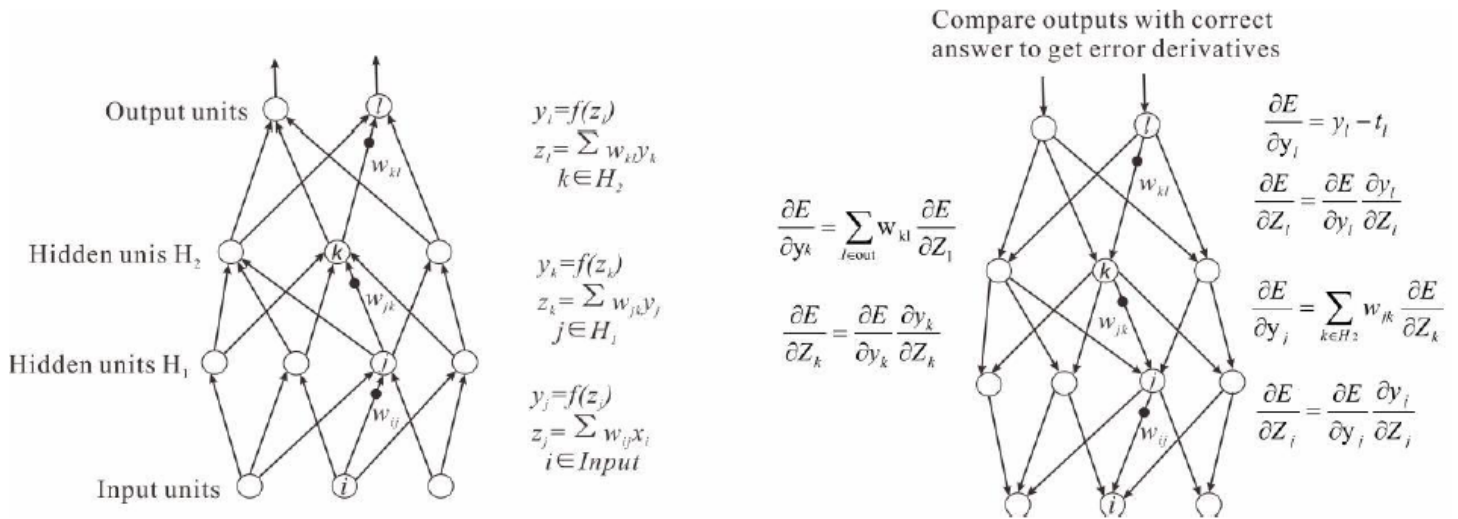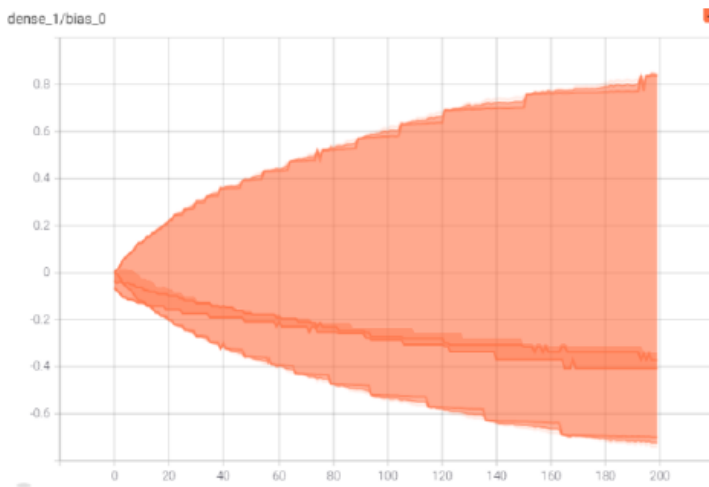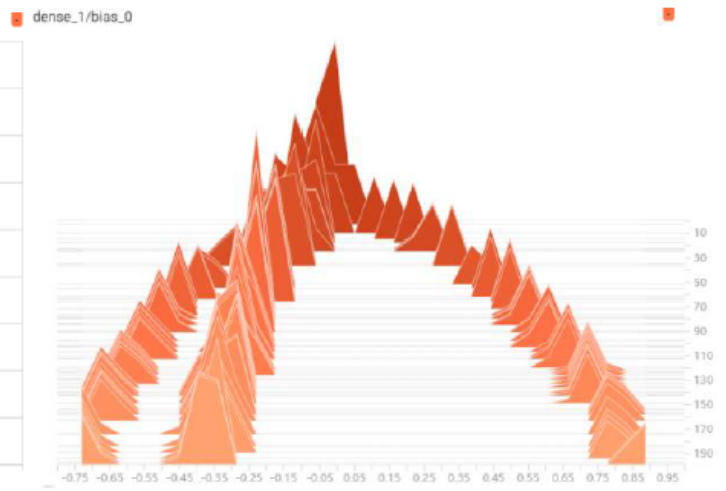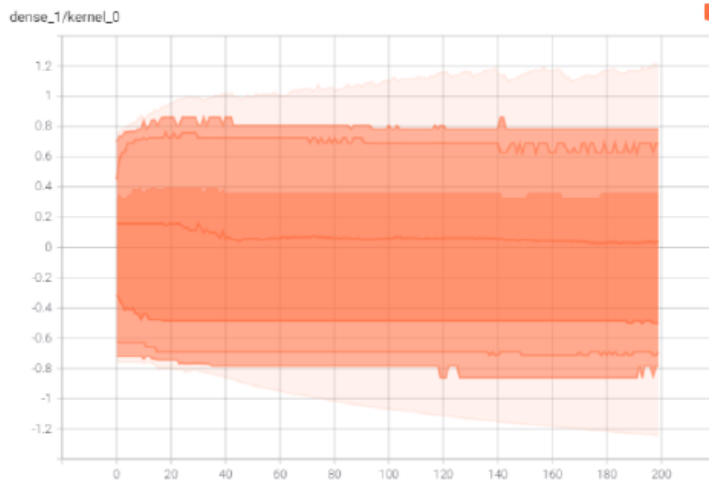
BP neural network compared with a deep learning ANN

Figure 5

Multilayer neural networks and backpropagation (Left: The equations used for computing the forward pass. Right: The equations used for calculating the backward pass). Compare outputs with correct answers to obtain error derivatives.

(a) The distribution of the bias

(b) The histogram of the bias

(c) The distribution of the weights
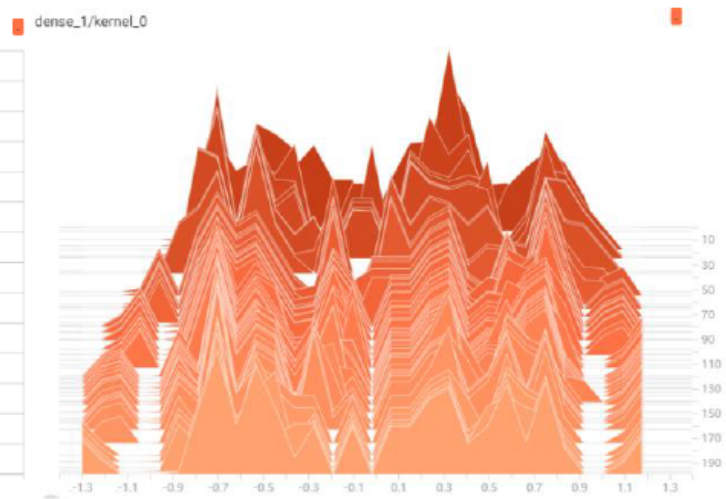
(d) The histogram of the weights

Figure 6

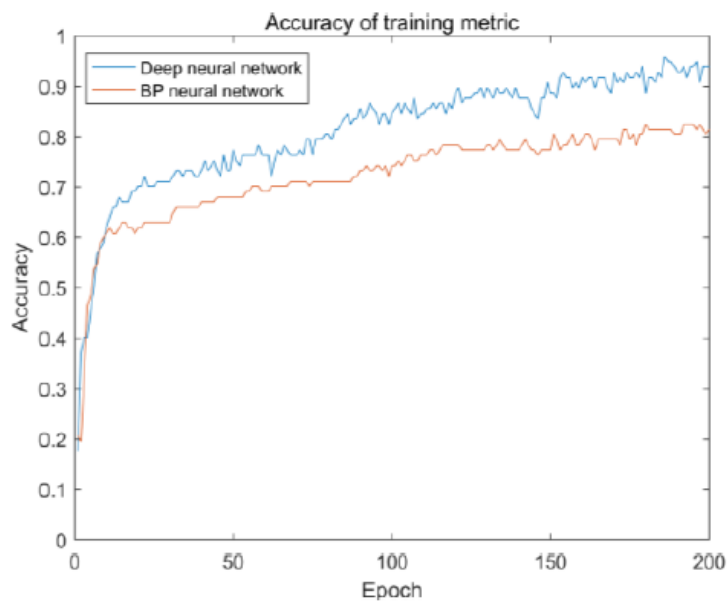The distribution and histogram of the weights and bias
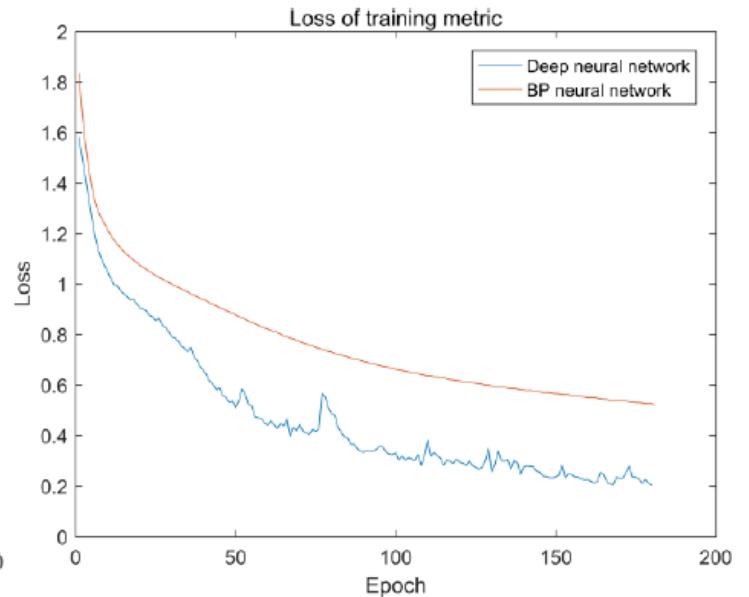
Figure 7
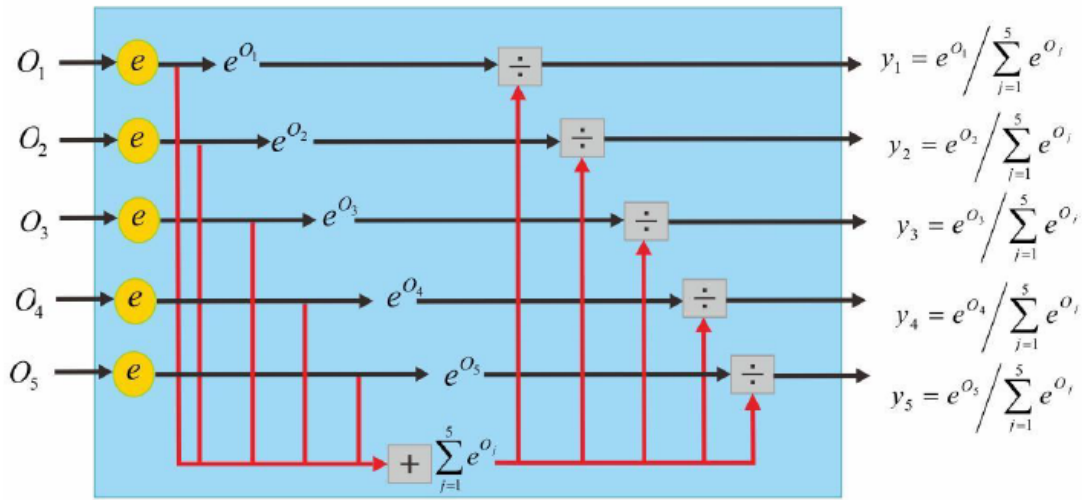
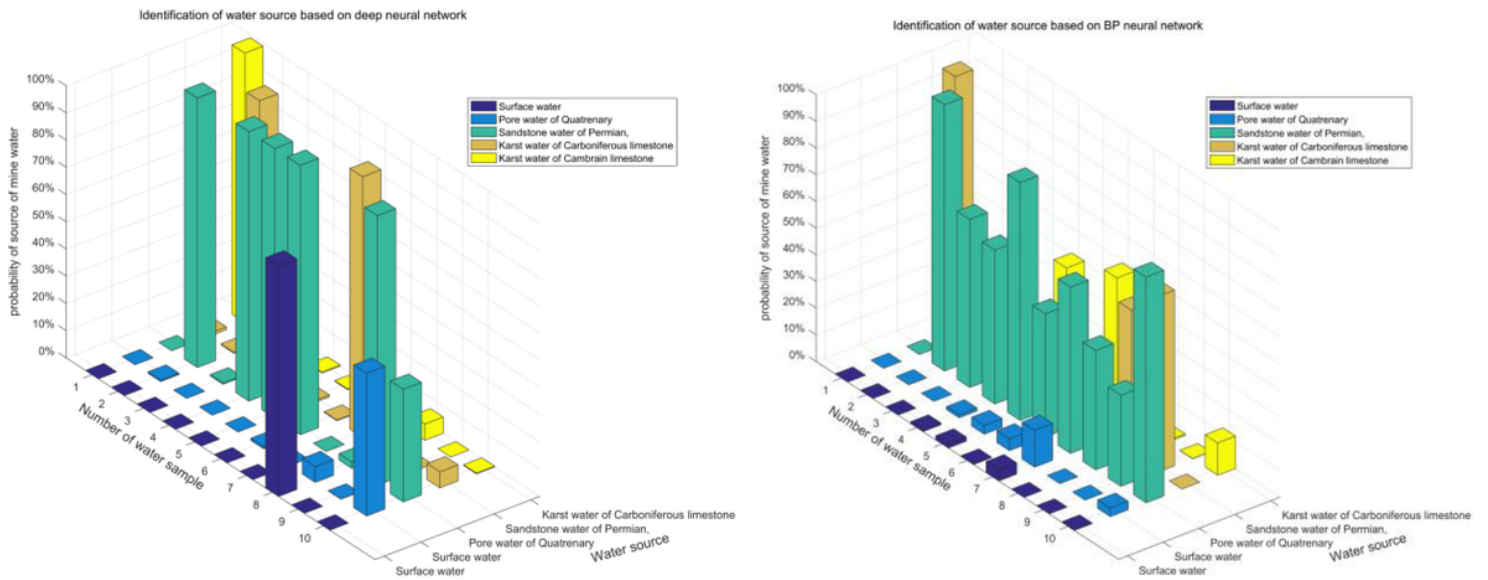Accuracy and loss of training metric

Figure 8

Softmax layer as the output layer

Figure 9

3D-Histogram of probability of source of mine water