Research on Prediction Model of Fracture Width in Loss Formation Based on Artificial Neural Network

Xuewen Yang[1], Hongtao Liu[1], Bo Zhou[1], Jinzhi Zhu[1], Shaojun Zhang[1], Tao Wang[1], Zhen Zhang[1], Haiying Lu[1], Erbiao Lou[1], Dan Bao[2,3,*]

[1] Petrochina Tarim Oilfield Company Korla 841000, China

[2] School of Chemistry and Chemical Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

[3] Research Center of Chongqing Oil & Gas Field Chemical Engineering Technology, Chongqing 401331, China

*Corresponding author: 18765920408@163.com

Abstract. Lost circulation in fractured formation has always been a worldwide technical problem in oil and gas drilling projects at home and abroad. The effect of bridging plugging depend on the matching of the fracture width and the particle size of lost circulation material. The plugging formula can be optimized according to the bridge rules to improve the success rate of plugging. In this paper, a prediction model of fracture width based on artificial neural network was established. Input layer, hidden layer and output layer were included. 17 input parameters such as well depth, formation, density of drilling fluid and displacement were determined. The fracture width was the output parameter. By optimizing the number of hidden layers and their nodes, the optimized training set had 2 hidden layers, and the number of nodes in the first hidden layer was 4, and the number of nodes in the second hidden layer was 10, the prediction accuracy of the model was the best, with the RMSE of 1.24 and the maximum R2 of 0.94. The model runs well in the test set with RMSE of 1.28 and R2 of 0.92. The fracture width prediction results of 5 Wells showed that the prediction results of this model match well with the field measurement results, and the average relative error percentage is only 2.94%. The established model had high accuracy, which can provide a basis for early and accurate prediction of well loss risk on site and for guiding optimization of engineering parameters to prevent well loss.

Keywords: Lost circulation, plugging, big data, artificial neural network, fracture width.

1. Introduction

Lost circulation in fractured formation is one of the most common types of drilling fluid loss during drilling [1-2]. Bridging plugging technique is the most commonly used method to deal with lost circulation in fractured formation, and its effect of plugging depends on the particle size of lost circulation material and the matching of the fracture, which has gradually formed the 1/3 rule, 2/3 rule, D50 rule, D90 rule, 3/10 and 6/5 rule and other basic theory of bridge plugging, guide the optimization of the particle size distribution of plugging materials [3-4]. However, in most cases, field engineers cannot obtain the fracture width when a lost circulation occurs, and the success rate of plugging is low when the particle size distribution of plugging material is not reasonable. If the fracture width of the downhole loss layer is known, the plugging formula can be optimized according to the bridging rules to form a dense pressure-bearing plugging layer to prevent drilling fluid loss [5]. Therefore, it is of great significance to obtain the fracture width of the leaky layer to deal with the Lost circulation efficiently. Traditionally, the fracture width is mainly predicted by field experience, or identified and predicted based on imaging logging, seismic and offset well data, but it lacks universality and accuracy, so it is urgent to establish a fracture width prediction model.

With the development of data science and technology, it is of great significance to collect, integrate and optimize all kinds of data from drilling using emerging technologies such as machine learning, big data and cloud computing, and to explore the potential value of massive drilling data to improve drilling efficiency and scientifically guide drilling decisions [6]. Lost circulation parameters, drilling parameters, drilling fluid parameters are intrinsically related to fracture width, and it is difficult to find a reasonable analytical solution to predict fracture width due to the high complexity and nonlinearity of these parameters. Machine learning can simulate this complex physical process and is able to build complex high precision prediction models by typical learning processes to develop correlations, transformations or mappings between data. Among them, artificial neural network (ANN) is the most powerful and effective technique to establish complex relational models based on previous experience at reasonable cost and time [7]. Therefore, it is of great significance to establish modeling between fracture width and drilling engineering parameters based on artificial neural network algorithm to predict fracture width for improve plugging efficiency.

He [8] proposed to establish a fracture width prediction model by using the optimized BP neural network method, which can solve the problem of difficult fracture width prediction and provide a better auxiliary decision in plugging engineering operations. Based on the rock mechanics mechanism of formation fracture generation, Chen [9] was based on the mechanism of formation of cracks in rock mechanics, identified six key mechanical and engineering factors affecting the fracture width, and established an analytical model of downhole formation fracture width using the nonlinear and large data characteristics of neural network calculation, to solve the problems of large error in judging the fracture by experience alone at the drilling site and the high cost of relying on imaging logging. Abbas [10] established a loss rate prediction model based on artificial neural network, analyzed the relationship between loss rate and drilling parameters, and controlled drilling parameters can effectively prevent drilling fluid loss. Al-Hameedi [11] conducted in-depth statistical analysis of more than 500 wells in the Rumaila oilfield, proposed a model for predicting loss in the Dammam formation, and proposed a model for optimizing drilling operations. Sabah [12] collected a large amount of data from 61 wells recently drilled in Iran's Marun oilfield, and established artificial neural network, decision tree, adaptive neural-fuzzy inference system, and genetic algorithm-multilayer perception model to quantitatively predict lost circulation. Hou [13] established lost circulation prediction system model by artificial neural network. He [14] applied big data processing ideas, used data processing techniques such as data correlation analysis, normalization processing, outlier processing, and imbalance processing, and optimized the deep neural network model to establish the lost circulation prediction model method.

In this paper, we have developed specification for lost circulation data acquisition, used big data analysis technology, compiled and analyzed on-site drilling data and lost circulation data in Block X of China's Tarim Oilfield, to simulation training with artificial neural network algorithms, established the prediction model between geological parameters, drilling parameters, drilling fluid parameters and fracture width, predicted loss layer fracture width during drilling, and provided technical support for scientific and efficient plugging in oilfield sites..

2. Artificial Neural Networks

Establishing the relationship between multiple input parameters, such as geological parameters, drilling parameters, and drilling fluid parameters, and the fracture width of the loss layer, involving complexity, uncertainty, and nonlinearity, the traditional mathematical methods have limitations in solving these problems, while artificial neural networks can be used to simulate the complex relationships between multiple input and output parameters.

Artificial neural network is an information processing system which approximate and simplified simulation the biological learning processes and has similar characteristics to biological neural network. It is composed of a large number of highly interconnected neurons that work in concert to solve a specific problem. Artificial neural networks consist of three parts, namely, input layer, one or more hidden layers and output layer. The optimal number of hidden layers and hidden layer neurons is determined by iterative process of repeated experiments until the error reaches the preset value. The forward flow of data is determined by the weight value and transfer function, and the adjustment in the artificial neural network modeling process is accomplished though trial-and-error techniques.

When the data training fails to reach the preset error value, it is propagated backwards, it is passed from the output layer to the input layer, arrives at the hidden layer, the data are multiplied by the adjusted weight values (w1j..wij), and then flows forward through the transfer function to generate the output of neurons until the error reaches a preset value and the training model is established [15].

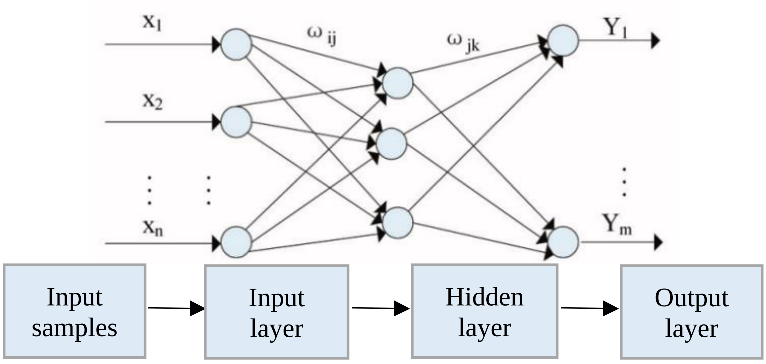

Figure 1. Artificial neural network flow chart

3. Establishment of fracture width prediction model and results

3.1. Input/output parameter acquisition

Table 1. Table of input/output parameters

Parameter Type		Parameter Name	Unit	Minimum value	Maximum value
Input Parameters	Geological parameters	Depth	m	5430	7950
		Formation	-	1	2
		Borehole size	mm	168.3	241.3
		Lithology	-	1	3
	Lost circulation Parameters	Working condition	-	1	4
		Loss speed	m/h^3	0	200
	Drilling parameters	Displacement	L/s	7	50
		Drilling pressure	kN	0	120
		Rotational speed	r/min	0	110
		Pump pressure	MPa	5	25
	Drilling fluid parameters	Density	g/cm^3	1.73	2.49
		Solid phase content	%	25	47
		Filtration loss	mL	2.0	9.0
		Dynamic shear	Pa	2.0	21.5
		Plastic viscosity	mPa-s	11.0	80.0
		Initial gel strength	Pa	1.0	6.0
		Final gel strength	Pa	2.0	18.0
Output parameter		Fracture width	mm	0	12.2

The preparation of database is a key step in the process of artificial neural network modeling, and the adequacy and accuracy of data are crucial for the accuracy of the model. The loss layer fracture width as an output parameter depends on input parameters such as geological parameters, drilling parameters, drilling fluid parameters, etc. Increasing the number of input parameters will reduce the running efficiency of the model, so input parameters are selected reasonably to maximize the running efficiency of the model. The field data such as drilling logs, drilling well history, drilling fluid history, and comprehensive logging records in Block X of Tarim Oilfield were collected, and the influencing factors of fracture width were comprehensively analyzed, established the specification of lost circulation data acquisition. The input parameters and output parameters are shown in Table 1, among which 17 parameters are used as input parameters: well depth, layer, borehole size, lithology, loss working condition, loss rate, displacement, drilling pressure, rotational speed, pump pressure, drilling fluid density, solid phase content, filtration loss, dynamic shear, plastic viscosity, initial cut, and final cut. The fracture width of the loss layer was used as the output parameter, and its value was obtained by inversion of the D90 particle size characteristic value of a successful plugging formulation in the field. A total of 1286 valid data were collected, including 876 loss data and 410 no loss data.

3.2. Data preprocessing

(1) Abnormal data processing: The abnormal values in the collected data are caused by human and equipment errors. Delete all records with abnormal values and eliminate data with large abrupt changes and unreasonable values.

(2) Digital processing: The artificial neural network modeling process only recognizes numerical type data, so the numbered category method is used to convert textual or symbolic data such as layer, lithology and loss working condition into numerical data, as shown in Table 2.

Table 2. Digitization conversion codes

Parameter Type	Name	Code
Formation	E	1
	K	2
Lithology	Fine Sandstone	1
	Mudstone	2
	Paste-bearing salt rock	3
Working condition	Drilling	1
	Circulation	2
	Down drilling	3
	Well curing	4

(3) Normalization process: In order to avoid the influence of the dimension and value range of different parameters on the accuracy of the model, the data normalization processing can eliminate the influence of the difference in the dimension and value range between indicators, and it is necessary to standardize the processing. The unified algorithm is used to convert the data uniformly, and the data are scaled according to the proportion to make it fall into a specific area for comprehensive analysis. The standardization of the input data depends on the transfer function used to construct artificial neural network [16]. For feedforward back propagation algorithms with logsig transfer functions, the parameters of the input and target data are normalized to the range 0 to +1 (Eq. 1). When using the tansig transfer function, the parameters of the input and target data are normalized to a range of -1 to +1 (Eq. 2).

$$\overline{x_i} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

$$\overline{x_i} = \frac{2(x_i - x_{\min})}{x_{\max} - x_{\min}} -1 \tag{2}$$

In the formula, x_i $\overline{x_i}$, $x_{\min}$, and $x_{\max}$ respectively represent the initial sample data in the sample information of each group, normalized data, and the minimum and maximum values in the initial sample data.

3.3. Model control parameter setting

The activation function ensures nonlinear variation of the input and output parameters. Each implicit layer is paired with an activation function that provides nonlinear variation. The most commonly used transfer functions from the input layer to the implicit layer are logsig and tansig. logsig is a logical transfer function as shown in Eq. (3). Tansig is a hyperbolic tangent function as shown in Eq. (4). The purelin transfer function is chosen for the implied layer to the output layer as a linear function, as shown in Eq. (5). Since there is a logical transfer relationship between the fracture width of the lost circulation layer and the engineering parameters, the logsig transfer function is selected for the implied layer and the Purelin transfer function is selected for the output layer in this paper.

$$\text{logsig}(n)=\frac{1}{1+\mathrm{e}^{-n}} \tag{3}$$

$$\text{tansig}(n)=\frac{2}{1+\mathrm{e}^{n}}-1 \tag{4}$$

$$\text{purelin}(n)=n \tag{5}$$

Through pre-modeling optimization, the number of hidden layers is set to 1~2 in this paper. If the number of neurons in the hidden layer is too less, it may not be able to train a mature network and recognize the unknown samples, which lacks the abilities of necessary learning and dealing information; on the contrary, the network tends to fall into partial optimization and causes a complex network structure, which makes the learning speed slow down. Therefore, the selection of the number of neurons in the hidden layer plays a crucial role in neural network modeling [17]. The optimal number of implied layers and implied layer neurons is determined by an iterative process of repeated trials. An empirical formula is used to calculate the number of hidden layer neurons as shown in Eq. (6).

$$l = a + \sqrt{n+m} \tag{6}$$

Where: l is the number of neurons in the hidden layer; n is the number of neurons in the input layer; m is the number of neurons in the output layer; α takes the value of 1~10. The range of the number of nodes of neurons in the hidden layer can be calculated according to the empirical formula.

To calculate the number of implied layer neurons, the output consequences are obtained by testing different numbers of implied layers and implied layer neurons, and the root means square error (RMSE) is calculated by comparing the generated output value f(xi) with the desired output value yi by equation (7) until the minimum mean square error is reached, which is the optimized number of implied layers and implied layer neurons.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(x_i)-y_i)^2} \tag{7}$$

3.4. Model Performance Evaluation

The most common way to assess performance evaluation is to split the dataset into a training dataset and a test dataset. The training dataset is used to modify and optimize the control parameters of the model, and the test dataset is used to test the performance of the developed fracture width prediction system model. The number of training datasets account for 70% and the number of test datasets account for 30% of the 1286 valid data collected. Compared the output values from the two modes with the actual measurement to assess the accuracy of the fracture width prediction system model. The root means square error (RMSE) and correlation coefficient (R2) is usually used to evaluate the accuracy of the model [18], as shown in Eq. 7and8. The smaller the RMSE value is and the closer to 1 the R 2value, the better the accuracy of the fracture width prediction system model is. The best prediction system model is established by adjusting the model control parameters (number of implied layers, number of implied layer neurons, etc.).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(f(x_i) - y_i)^2}{\sum_{i=1}^{n} f(x_i)^2 - \sum_{i=1}^{n} y_i^2 / n} \tag{8}$$

In the above equation, y is the actual fracture width (obtained by inversion of the D90 grain size characteristic value of a successful plugging formulation in the field), mm; f(x) is the predicted fracture width, mm; n is the total number of data used for model estimation.

3.5. Model prediction results

Based on the determined input parameters and output parameters, the optimal selection of the hidden layers and their neurons is carried out for the training set to determine the artificial neural network model with the best simulation results, and the results are shown in Table 3. When there is 2 hidden layer, the number of neurons in the first hidden layer is 4, and the number of neurons in the second hidden layer is 10, the simulation results are optimal, the accuracy of lost circulation rate prediction is the highest, the RMSE is the smallest as 1.24, and the R2 is the largest as 0.93.

Table 3. Preferred results of hidden layers and their neuron numbers for the training set (partial)

Number of hidden layers	Number of neurons in the first hidden layer	Number of neurons in the second hidden layer	Error Analysis	
			RMSE	R^2
1	7	0	5.13	0.51
1	8	0	4.28	0.62
1	9	0	3.46	0.76
1	10	0	1.95	0.81
1	12	0	2.04	0.84
1	14	0	1.92	0.89
2	2	9	2.17	0.84
2	2	12	1.85	0.87
2	3	8	4.02	0.74
2	3	10	2.17	0.86
2	3	9	3.04	0.80
2	3	11	2.76	0.78
2	4	6	1.75	0.91
2*	4	10	1.24	0.94
2	5	4	1.94	0.79
2	6	8	2.74	0.81
2	7	4	2.01	0.86
2	8	6	1.98	0.87

Note: * is the optimal solution

Based on the prediction of the artificial neural network fracture width established by the training set, using 30% testing set data to test analysis and verify the accuracy of the training set model, and the results are shown in Table 4. The model runs well in the test set with RMSE of 1.28 and R2 of 0.92, which proves that the established fracture width prediction system model has high simulation accuracy and can be used to predict the fracture width of lost circulation in the field.

Table 4. Error analyses of the training and test sets

Data set	Error Analyses	
	RMSE	R^2
Training set	1.24	0.94
Test set	1.28	0.92

4. Application Examples

The fracture width predicted model based on an artificial neural network was used to predict the fracture width of the newly drilled 5 wells in the field, as shown in Table 5. In the table, the fracture width is the D90 inversion value of a successful plugging formulation. Compared prediction fracture width with actual fracture width, the result shows that the error between the predicted fracture width and the actual fracture width by the artificial neural network model is smaller, with an average relative error percentage of only 2.94%, indicating that the fracture width prediction system model based on artificial neural network has high accuracy and provides a basis for reasonable optimization of the particle size distribution of plugging materials.

Table 5. Application examples of fracture width prediction system models based on artificial neural networks

Parameter Name	Unit	Well A	Well B	Well C	Well D	E well
Depth	m	6310	6535	5942	6183	5750
Formation	-	2	2	1	2	1
Borehole size	mm	168.3	168.3	241.3	168.3	241.3
Lithology	-	2	1	3	1	3
Working condition	-	1	1	1	1	1
Loss speed	m/h^3	30	60	12	6.5	28
Displacement	L/s	26	18	12	21	12
Drilling pressure	kN	40	60	60	50	80
Rotational speed	r/min	90	60	70	60	60
Pump pressure	MPa	20	22	18	23	10
Density	g/cm^3	1.73	1.85	1.76	1.78	2.13
Solid-phase content	%	31	36	36	40	44
Filtration loss	mL	4.2	5.0	4.0	3.8	3.2
Dynamic shear	Pa	4.0	4.0	4.0	5.5	7.0
Plastic viscosity	mPa-s	27.0	27.0	29.0	32.0	64.0
Initial gel strength	Pa	2.5	3.0	3.0	4.0	3.0
Final gel strength	Pa	4.0	5.5	6.5	6.0	7.5
Actual fracture width	mm	3.85	7.23	2.42	0.86	4.24
Predicted fracture width by artificial neural network	mm	3.74	7.05	2.38	0.90	4.11
Relative error percentage	%	2.85	2.48	1.65	4.65	3.06

5. Conclusions

(1) Determine those 17 input parameters and the output parameter is fracture width. The fracture width prediction system model was established with an artificial neural network to accurately predict the fracture width and provide a basis for the optimization of the bridge plugging formula.

(2) The optimized training set model has two hidden layers, and the prediction accuracy is optimal when the number of neurons in the first hidden layer is 4 and the second hidden layer is 10, with a minimum RMSE of 1.24and R2 of 0.94. The model runs well in the test set with RMSE of 1.28 and R2 of 0.92.

(3) Using the established artificial neural network model to predict the fracture width of 5 new wells, the predicted result of the model highly matched the actual result with an average relative error of only 2.94%, the established artificial neural network fracture width prediction system model has high accuracy to guide the optimization of the bridge plugging formula.

References

[1] Feng Y., Gray K.E. Review of fundamental studies on lost circulation and wellbore strengthening [J]. Journal of Petroleum Science and Engineering, 2017, 152: 511-522.

[2] Xu Tongtai, Liu Yujie, Shen Wei. Technology of lost circulation resistance and control during drilling engineering [M]. Beijing: Petroleum Industry Press, 1997.

[3] Qiu Zhengsong, Bao Dan, Liu Junyi, Chen Jiaxu, Zhong Hanyi, et al. Microcosmic mechanism of fracture-plugging instability and experimental study on pressure bearing and tight plugging [J]. Acta Petrolei Sinica, 2018, 39(05): 587-596.

[4] Sandeep D. Kulkarni and Dale E. Jamison, Kushabhau D. Teke. Managing suspension characteristics of lost-Circulation materials in a drilling fluid [C]. SPE-170271-MS. 2014.

[5] Cook J., Growcock F., Guo Q., et al. Stabilizing the wellbore to prevent lost circulation [J]. Oilfield Review, 2011, 23(4): 26-35.

[6] Weixiao Li. Research on drilling optimization control based on oil drilling big data Technology [D]. Xi'an Shiyou University, 2018.

[7] Junwen Cai. Research on well loss diagnosis model based on neural network fusion technology [D]. Southwest Petroleum University, 2014.

[8] Tao He, XianTao Xie, Jun Wang, et al, an optimized BP neural network was used to establish a crack width prediction model [J]. Drilling fluid and completion fluid, 2021, 38(02): 201-206.

[9] Zengwei Chen. Downhole fracture diagnosis and plugging technology based on neural network algorithm [J]. Drilling fluid and completion fluid, 2019, 36(1): 20-24.

[10] Abbas A K, Bashikh A A, H Abbas, et al. Intelligent decisions to stop or mitigate lost circulation based on machine learning [J]. Energy, 2019, 183: 1104-1113.

[11] Al-Hameedi A, Alkinani H H, DunnNorman S, et al. Using Machine Learning to Predict Lost Circulation in the Rumaila Field, Iraq [C]. SPE-191933-MS. 2018.

[12] Sabah M, Talebkeikhah M, Agin F, et al. Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: A case study from Marun oil field [J]. Journal of Petroleum Science & Engineering, 2019. 124: 182-193.

[13] X Hou, Yang J, Yin Q, et al. Automatic Gas Influxes Detection in Offshore Drilling Based on Machine Learning Technology [C]. SPE-182463. 2019.

[14] Pengfei He, Xiaobin Liu, Zhen Chen, et al. Research on well loss prediction based on deep neural network model [J]. Tianjin science and technology, 2019, 46(S1): 21-23.

[15] Leiwen Wang. Research on key parameters of drilling leak prevention and plugging based on BP neural network [D]. Southwest Petroleum University, 2019.

[16] Abbas A K, Al-Haideri N A, Bashikh A A. Implementing artificial neural networks and support vector machines to predict lost circulation [J]. Egyptian Journal of Petroleum, 2019. 175: 1025-1037.

[17] Alkinani H, Al-Hameedi A, Dunn-Norman S. Minimizing Lost Circulation Non-Productive Time Using Expected Monetary Value and Decision Tree Analysis [C]. SPE-192263. 2021.

[18] Xiyu Tu. Application and analysis of industrial data mining in well loss accident [D]. Shanghai Jiaotong University, 2019.