Asian Journal of Water, Environment and Pollution. Vol. 23, No. 1 (2026), pp. 261-272.
doi: 10.36922/AJWEP025370287

ORIGINAL RESEARCH ARTICLE

Neural network-based prediction of drilling fluid leakage

Wenbing Wu[1,2], Tao Liu[2], Lianlu Huang[2], Jian Wang[1]*, Chenxin Wang[1]*, Hua Li[2], Justine Kiiza[1,3], Moussa Camara[4], Jie Zhong[1]*, and Jiafang Xu[1,5]*

[1]Department of Marine Oil, Gas and Hydrate, School of Petroleum Engineering, China University of Petroleum (Qingdao Campus), Qingdao, Shandong, China
[2]Drilling Company, China National Petroleum Corporation Offshore Engineering Company Limited, Tianjin, China
[3]College of Natural Sciences, Makerere University, Kampala, Uganda
[4]School of Chemistry, Petroleum and Energy, Institut National Polytechnique Felix Houphouët Boigny, Yamoussoukro, Côte d'Ivoire
[5]State Key Laboratory of Deep Oil and Gas, China University of Petroleum (Qingdao Campus), Qingdao, Shandong, China
***Corresponding authors**: Jian Wang (b20020065@s.upc.edu.cn); Chenxin Wang (s24020010@s.upc.edu.cn); Jie Zhong (dynamic.zhong@outlook.com); Jiafang Xu (xjiafang@upc.edu.cn)

Received: September 8, 2025; Revised: September 25, 2025; Accepted: October 30, 2025; Published online: November 19, 2025

Abstract: During the drilling process, reservoir fractures may lead to drilling-fluid loss, thereby slowing drilling progress and reducing well productivity. Therefore, it is necessary to choose the appropriate materials and formulations for plugging, and the leakage amount and rate are the most important indicators for selecting plugging agents. In this study, the amount of rigid mineral particles and plant fibers commonly used in drilling, as well as the width of formation fractures, were used as input variables, while leakage volume served as the output variable. By combining the multiple-population genetic algorithms (MPGA) and the backpropagation neural network (BPNN), an MPGA–BPNN prediction model was established to predict the leakage amount under different plugging formulations. The results showed that the correlation coefficient of the established prediction model reached 0.9741, indicating strong predictive accuracy for leakage volume and plugging performance under varying formulation conditions, providing useful reference and guidance for the optimization of plugging agents.

Keywords: Drilling fluid; Lost circulation control and plugging; Leakage amount; Neural networks; Multi-population genetic algorithm

1. Introduction

The huge loss of drilling fluid caused by the invasion into a formation is a common problem in drilling operations and occurs frequently during the drilling process. It not only prolongs drilling time but may also trigger complex downhole situations such as wellbore collapse, blowouts, and drill string sticking, resulting in severe economic losses.[1,2] Therefore, it is necessary to effectively prevent and control drilling-fluid loss to ensure the smooth progress of drilling operations. At present, the commonly used method is to add the relevant lost circulation control and plugging materials to the drilling fluid to suppress fluid loss. For example, lost circulation bridging materials include granular, fibrous, and flaky inert particles; lost circulation polymer cross-linking materials include acrylamide monomer polymers and similar substances; and rigid materials include calcite particles characterized by high strength and strong acid solubility. In addition,

many experts and scholars have conducted research on the development and optimization of lost circulation control and plugging agents. Ma *et al*.[3] proposed a dynamic fracture plugging experimental device and developed design criteria for lost circulation materials for induced fractures. Nazemi *et al*.[4] evaluated a novel smart fluid with high design flexibility in terms of shear rate, temperature, and shear history. Liu *et al*.[5] proposed a percolation fracture network that considers fracture connectivity evolution. Zeng *et al*.[6] addressed fracture-related fluid loss and reservoir protection in a South China Sea field by optimizing particle size grading criteria and the waterproofing agent, and by constructing a lost circulation control and plugging system for drilling fluids. At present, the selection of lost circulation control and plugging agents in drilling fluid mostly involves measuring the leakage amount of different components through orthogonal experiments to optimize the best formulations. This process usually requires a large number of laboratory experiments and consumes significant time and materials. Therefore, there is an urgent need for a low-cost and efficient method to optimize drilling fluid plugging agents.

With the development of computing technology, artificial intelligence technology has been widely applied in the petroleum industry, including petroleum exploration,[7-9] mining,[10,11] and transportation,[12-14] as shown in Table 1. Among these applications, artificial intelligence technology has also been used in the field of drilling fluids. Zhou *et al*.[15] used artificial intelligence to comprehensively assess wellbore stability and reduce the risk of lost circulation. Wu *et al*.[16] proposed a data-driven machine learning method to identify lost circulation risk. These studies have demonstrated that artificial intelligence technology can be effectively applied to lost circulation control and plugging, achieving good results. Among machine learning prediction methods, the backpropagation neural network (BPNN) has strong self-learning, self-adaptive, generalization, non-linear processing, and fault-tolerant capabilities. It demonstrates excellent predictive performance for regression problems and is thus widely applied in various fields. Meanwhile, genetic algorithms, especially the multiple population genetic algorithm (MPGA), possess a strong global search ability. They can solve the problems of local optima and overfitting that occur during BPNN training,[17] thereby improving prediction accuracy by finding the global optimal solution. Amish and Khodja[18] summarized several machine learning models for predicting fluid loss and discussed other possible applications. Arbi and Tamás[19] studied the potential of machine learning and deep learning algorithms for predicting the non-linear behavior of the rate of penetration using data-driven methods. Ismail *et al*.[20] applied various machine learning techniques to evaluate the gel strength and viscosity of drilling mud. Although relevant artificial intelligence models have been applied in the field of oil drilling, research specifically dedicated to drilling-fluid loss prediction remains limited. In this study, the MPGA is combined with the BPNN to develop an MPGA–BPNN drilling-fluid loss prediction model. This model can predict the leakage amount for different lost circulation control and plugging schemes, thereby providing valuable reference and guidance for the selection of plugging agents in drilling operations.

Table 1. Application of artificial intelligence in the petroleum industry

Field	Model name
Geophysics and exploration	Porosity prediction;[7] reservoir quality prediction;[8] and reservoir hydrocarbon channel prediction[21]
Drilling engineering	Drilling trajectory prediction;[22] wellbore temperature prediction;[23] and water production rate prediction[24]
Oil production engineering	Production prediction;[25] recovery factor prediction;[26] and reservoir saturation prediction[27]
Production, gathering, and transportation	Pipeline defect prediction;[28] flow pattern recognition;[29] and early warning systems[30]

2. MPGA–BPNN prediction model

2.1. BPNN

The mechanism of drilling fluid leakage involves the strong coupling of geological conditions, drilling parameters, fluid properties, and other factors, making it a highly non-linear and multivariable problem. As a powerful nonlinear mapping tool, BPNN can learn this complex relationship from historical data. Therefore, this study employs BPNN as the basic prediction model. As a multilayer feedforward network based on error backpropagation, the BPNN has been widely applied in numerous fields. Its network structure encompasses an input layer, a hidden layer, and an output layer.[31] After the input data enter the input layer, they are processed in the hidden layer, and the final result is generated by the output layer. The number of hidden layers in

the model has a certain impact on its performance. Research indicates that, theoretically, a greater number of hidden layers can improve accuracy;[32] however, it also increases training time and the risk of overfitting. If only one hidden layer is used, increasing the number of nodes in that layer can greatly simplify the BPNN structure and improve the convergence efficiency.[33-35] Therefore, in this study, only one hidden layer was adopted, and the model's performance was improved by adjusting the number of nodes in that layer, as shown in Figure 1.

Selecting the appropriate input parameters is a prerequisite for ensuring the accuracy of prediction results. The amount of slurry loss is related not only to the lost circulation control agents but also to the fracture width. In this regard, referring to the relevant literature,[36] this study collected and selected rigid mineral particle plugging agents for drilling (divided into GZD-A, GZD-B, GZD-C, and GZD-D, in descending order of particle size), composite plant fiber lost circulation control agents for drilling (divided into GDJ-I, GDJ-II, GDJ-III, and GDJ-IV, also in descending order of particle size), and the width of the open end (D_o) and the terminal width (D_e) of the simulated fracture as input parameters.

2.2. Genetic optimization algorithm

The loss mechanism of drilling fluid is highly complex and influenced by the nonlinear coupling of various factors. Although BPNN can fit this complex relationship, its learning method based on gradient descent is prone to being trapped in local optima when applied to limited field data. This can lead to unstable prediction performance and insufficient generalization ability of the model. To overcome this inherent limitation, this study introduces a genetic algorithm to optimize the BPNN.[37] As an optimization algorithm that simulates the biological evolution process,[38] the genetic algorithm adopts a global optimization strategy and can effectively optimize parameters like the weights in the neural network to find the global optimal solution. The GA used to optimize the BPNN adopts the concept of "survival of the fittest" in genetic evolution and achieves the global optimum through a competitive mechanism. The basic procedure involves using individuals to represent the initial weights and thresholds of the network. The norm of the test error of the BPNN for the prediction samples is taken as the output of the objective function, and the fitness value of each individual is then calculated. Through the operations of selection, crossover, and mutation, the algorithm continuously searches for the optimal individual—that is, the optimal initial weights and thresholds of the BPNN—and incorporates them into the BPNN model to achieve the optimal state.[39,40] The specific process is shown in Figure 2.

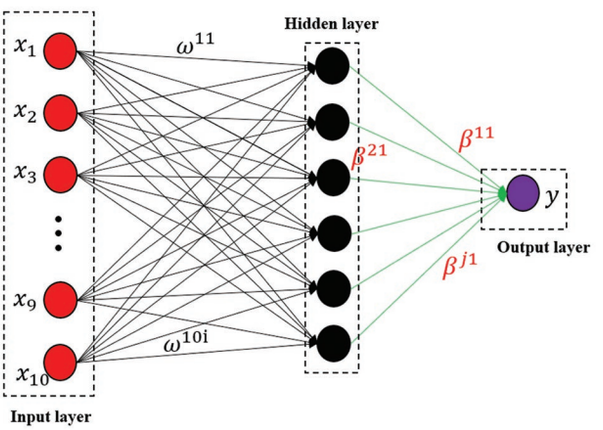

Figure 1. Neural network structure, where x_1, x_2,… x_n are the input parameters of the model, and y is the corresponding output parameter. ω^{ij} represents the weight between the i-th input layer and the j-th hidden layer, and β^{j1} represents the weight between the j-th hidden layer and the output layer

2.3. MPGA

Although the GA can enhance the prediction accuracy of the BPNN, it still suffers from the drawback of premature convergence. Specifically, in the presence of exceptional individuals within the population (i.e., individuals whose fitness is substantially higher than that of others), such individuals tend to be repeatedly selected during the selection process. Consequently, the new population becomes dominated by these individuals, reducing population diversity. To prevent this phenomenon, this study employed the MPGA to further improve the prediction performance of the model. The MPGA is an extended form of evolutionary computation, and its structure is shown in Figure 3.[41] This algorithm builds on the traditional genetic algorithm by introducing the concepts of population partitioning and co-evolution. In this approach, the weights and thresholds of the BPNN are treated as optimization variables, which are encoded as a series of numbers, represented as chromosomes. Crossover and mutation operations are then carried out using the genetic algorithm to continuously search for the optimal individuals. The entire search space is divided into multiple subgroups, and each subgroup independently performs genetic operations, including selection, crossover, and mutation, thereby realizing parallel exploration in different regions of the global

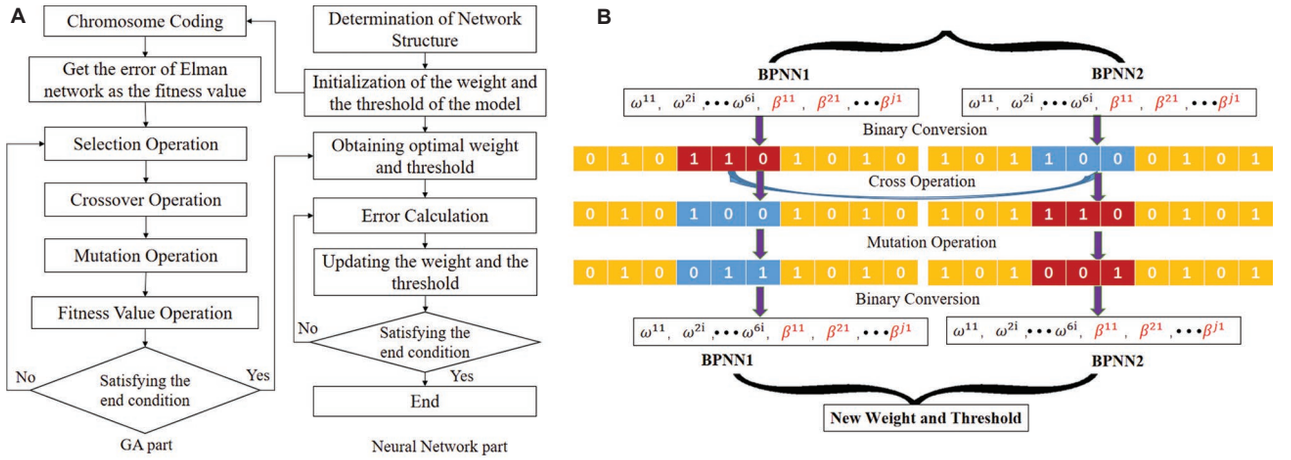

Figure 2. Process of using genetic algorithms (GA) to optimize the backpropagation neural network (BPNN). (A) Flowchart of GA–BPNN; and (B) example diagram of GA optimizing the weights and thresholds of BPNN.

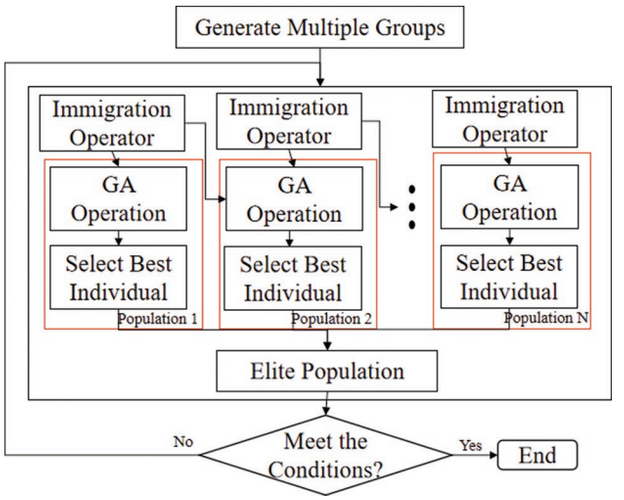

Figure 3. Multiple population genetic algorithm flowchart

Abbreviation: GA: Genetic algorithm.

search space. Finally, the optimal chromosomes are decoded into corresponding weights and thresholds and introduced into the BPNN model to reach the optimal state. This method promotes the exchange of high-quality genes while maintaining population diversity, preventing the algorithm from prematurely converging to the local optimum, and enhancing the global search capability of the algorithm. The corresponding specific parameters are listed in Table 2.

3. Selection of data analysis and evaluation methods

A total of 60 data samples (Table 3) were collected under different combinations of plugging agents and lost circulation control agents, and the data characteristics are summarized in Table 4. To further investigate the impact of these factors on drilling-fluid loss, a correlation analysis was conducted, and the results are presented in Figure 4. As shown in the figure, D_o, D_e, and GZD-A exhibit relatively strong correlations with the amount of drilling-fluid loss. Among them, D_o shows the strongest correlation, whereas other factors demonstrate weaker correlations with fracture width. From high to low correlation, the order is GDJ-IV, GZD-B, GZD-C, GDJ-II, GDJ-III, GDJ-I, and GZD-D. Overall, each parameter exerts a certain influence on drilling-fluid loss. Excluding any parameter may ignore its potential or synergistic effects, resulting in biased model results. Therefore, all parameters were retained as input variables in this study.

In this study, the data were divided into training and testing datasets. Among them, 55 samples were used for training and 5 for testing. Given the limited amount of data, the cross-validation approach was employed to construct the prediction model. The 55 training samples

Table 2. Specific parameters of the MPGA and BPNN algorithms

Parameter	MPGA	BPNN
Population size	10	–
Number of individuals	5	–
Number of genetic generations	20	–
Learning rate	–	0.01
Number of training cycles	–	8,000
Target error value	–	0.001

Abbreviations: BPNN: Backpropagation neural network; MPGA: Multiple population genetic algorithm.

Table 3. Experimental data

GZD-A/%	GZD-B/%	GZD-C/%	GZD-D/%	GDJ-I/%	GDJ-II/%	GDJ-III/%	GDJ-IV/%	D_o/mm	D_e/mm	SL/mL
0	0	0.5	0.5	0	0	0.5	0.5	1	0.5	121
0	1	1	0.5	0	2	0	1	1.5	1	50
0	2	1	0.5	0	2	2	2	1.5	1.5	12
2	1	0.5	0.5	0	2	0	1	2	1.5	216
3	1.5	1	0.5	0.5	1	1	1	3	3	312
0	0	0.5	1	0	0	0.5	1	1	0.5	81.5
0	2	1	0.5	0	1.5	0	1	1.5	1	33
0	2	1	0.5	0	1.5	2	2	1.5	1.5	28
2	1	0.5	1	0	1.5	0	1	2	1.5	237
3	2	1	0.5	1	0.5	1	1	3	3	74.5
0	0	0.5	0.5	0	0	0	1	1	0.5	49.5
0	2	1	1	0	1.5	0	1	1.5	1	67
0	2	1	0.5	0	1	2	2	1.5	1.5	19
2	0	1	0.5	0	1.5	0	1	2	1.5	174
3	2	1	0.5	1	1	0.5	1	3	3	104.5
0	0	1	0.5	0	0	0	1	1	0.5	36
0	2	1	1	0	1.5	0	1	1.5	1	61
0	2	1	0.5	0	1	2	2	1.5	1.5	26
3	0	1	0.5	0	1.5	0	1	2	1.5	58
3	2	1	0.5	1	1	1	0.5	3	3	128
0	0	0.5	0.5	0	0	0	1	1	0.5	70
0	2	1.5	1	0	1.5	0	1	1.5	1	68.5
0	2	1	0.5	0	1	1	2	1.5	1.5	47
1	0	1	0.5	0	1.5	0	1	2	1.5	312.5
3	2	1	0.5	1	1	0.5	0.5	3	3	126
0	0	0.5	1	0	0	0	1.5	1	0.5	79
0	2	1.5	1	0	1.5	0	1	1.5	1	139
0	2	1	0.5	0	1.5	1	2	1.5	1.5	20
3	0	1	0.5	0	1.5	0	1	2	1.5	18
4	2	1	0.5	1	1	0.5	0.5	3	3	144
0	0	0.5	0.5	0	0	0.5	0.5	1	0.5	121
0	2	1	0.5	0	1.5	0	1	1.5	1	33
0	2	1	0.5	9	1.5	2	2	1.5	1.5	28
2	0	1	0.5	0	2	2	2	2	1.5	174
4	2	1	0.5	1	1	0.5	0.5	3	3	126
0	0	0.5	0.5	0	0	0	1	1	0.5	70
0	2	1	1	0	1	0	1	1.5	1	61
0	2	1	0.5	0	1	2	2	1.5	1.5	47
3	0	1	0.5	0	2	1	1	2	1.5	58
5	3	2	1	1	1	0.5	0.5	3	3	47.5

(Cont'd...)

Table 3. ***(Continued)***

GZD-A/%	GZD-B/%	GZD-C/%	GZD-D/%	GDJ-I/%	GDJ-II/%	GDJ-III/%	GDJ-IV/%	D_o/mm	D_e/mm	SL/mL
0	0.5	0.5	0.5	0	0	0.5	1	1	0.5	93
0	2	1	0.5	0	1	0	0.5	1.5	1	12.5
0	2	1	0.5	0	2.5	0	1.5	1.5	1.5	73
3	0	1	0.5	0	1	0	0.5	2	1.5	5
4	2	1	0.5	1	1	0.5	0.5	3	3	144
0	0.5	0.5	1	0	0	1	0	1	0.5	91
0	2	1	0.5	0	1	0	0.5	1.5	1	25
0	2	1	0.5	0	1	1	2	1.5	1.5	15
3	0	0.5	0.5	0	1	0	0.5	2	1.5	32
5	3	1	0.5	2	2	0.5	0.5	3	3	59
0	0.5	1	0.5	0	0	1	0	1	0.5	87
0	2	1	1	0	1	0	0.5	1.5	1	36
0	2	1	0.5	0	2	0	1.5	1.5	1.5	14
3	0	1	0.5	0	1	0	0.5	2	1.5	27
5	3	1	0.5	2	1	2	1	3	3	79
0	0	0	1	0	0	1	0	1	0.5	135
0	2	1.5	1	0	1	0	0.5	1.5	1	69
0	2	1	0.5	0	3	0	1.5	1.5	1.5	8
3	0	0	1	0	1	0	0.5	2	1.5	16
5	3	2	1	2	2	1	1	3	3	47.5

Abbreviations: D_o: Width of the open end; D_e: The terminal width; SL: Slurry loss.

Table 4. Dataset characteristics of parameters

Parameter	GZD-A/%	GZD-B/%	GZD-C/%	GZD-D/%	GDJ-I/%	GDJ-II/%	GDJ-III/%	GDJ-IV/%	Do/mm	De/mm	SL/mL
Max	5	3	2	1	9	3	2	2	3	3	312.5
Min	0	0	0	0.5	0	0	0	0	1	0.5	5
Mean	1.2833	1.3000	0.9250	0.6250	0.3917	1.1000	0.5417	1.0083	1.8000	1.5000	79.0917
Median	0	2	1	0.5	0	1	0	1	1.5	1.5	61

Abbreviations: D_o: Width of the open end; D_e: The terminal width; Max: Maximum; Min: Minimum; SL: Slurry loss.

were partitioned, and the established model was trained and evaluated accordingly to improve the model's adaptability. To prevent differences in magnitude among parameters from affecting the results, and to enhance comparability, convergence speed, and model accuracy, the data were normalized. The numerical ranges of all parameters were scaled to the same interval[42,43] to avoid adverse effects caused by parameter magnitude differences. The specific normalization formula is as follows:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Where x_i represents the input data, x'_i represents the normalized data, and x_{min} and x_{max} are the minimum and maximum values of the corresponding data, respectively.

To evaluate and compare the prediction performance of the models, the root mean square error, correlation coefficient (R^2), mean absolute error, and mean absolute percentage error were selected as performance metrics. The calculation formulas[44,45] are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i')^2} \quad (2)$$

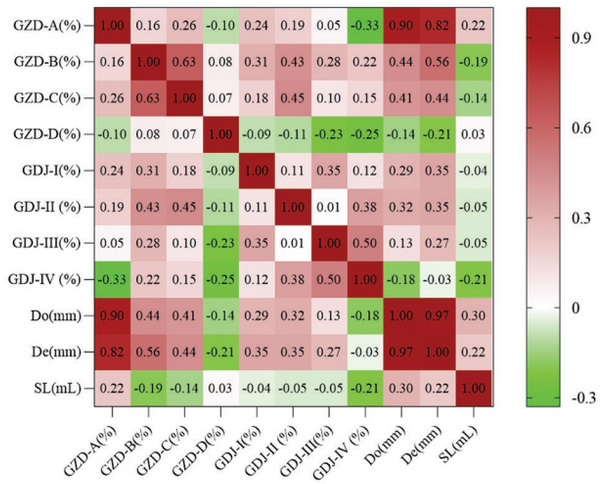

Figure 4. Correlation coefficient matrix of influencing factors

Abbreviations: D_o: Width of the open end; D_e: The terminal width; SL: Slurry loss

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i'}{y_i}\right| \times 100\% \tag{4}$$

$$MAE = \frac{1}{n}\sum_{n=1}^{n}|y_i - y_i'| \tag{5}$$

where y_i' is the predicted value, y_i is the corresponding actual value, and n is the number of samples.

4. Data structure selection and optimization

4.1. Model structure optimization

For the prediction model, the number of neurons in the hidden layer is a key parameter and has a significant impact on the model's predictive ability. If the number of neurons is insufficient, the model may not be able to capture the complex relationships between the input and output data, resulting in poor generalization ability. Conversely, if the number of neurons is excessive, the model may overfit the training data, thereby affecting prediction accuracy and increasing the time required for model training. To address this issue, this study adopted the empirical formula shown in Equation 6[46,47] to determine the number of nodes in the hidden layer of the model:

$$k = \sqrt{q + m} + a \tag{6}$$

Where a is an integer between 1 and 10; q and m are the number of nodes in the output layer and the input layer, respectively; and k is the number of nodes in the hidden layer.

In this study, the number of nodes in the input and the output layers was 10 and 1, respectively. Therefore, the number of nodes in the hidden layer ranged from 4 to 37. Models were established for each number of nodes within this range, and the prediction results are shown in Figure 5.

When the number of neurons in the hidden layer was five, the model achieved the minimum error and the highest correlation coefficient. Thus, the number of neurons in the hidden layer was determined to be five in this study.

4.2. Optimization effect analysis

To analyze the optimization effect of the MPGA, this study comparatively examined the prediction performance of the non-optimized BPNN, GA–BPNN, and MPGA–BPNN models. The results are shown in Figure 6 and Table 5.

As shown in Figure 6, there are relatively large differences among the data. The results indicate that larger leakage volumes generally occur in samples with wider fracture openings (>1.5 mm). At the same time, the particle size of the rigid mineral particles GZD used for drilling is relatively large, resulting in a higher leakage volume. In addition, Table 5 shows that the prediction performance of the model improved significantly after optimization by the genetic algorithm, with the correlation coefficient increasing to 0.9079, an improvement of 0.115. After optimization by the MPGA, the prediction effect was further enhanced, reaching 0.9741, an additional increase of 0.0662. The genetic algorithm significantly improved the prediction accuracy and overall model performance, while the MPGA further enhanced prediction performance on this basis.

5. Analysis of model results

5.1. Model prediction results

In this study, the dataset was used to train and validate the prediction model, and the obtained results are shown in Figure 7. As illustrated in the figure, during both the training and testing phases, the training and predicted values of the model were generally consistent with the

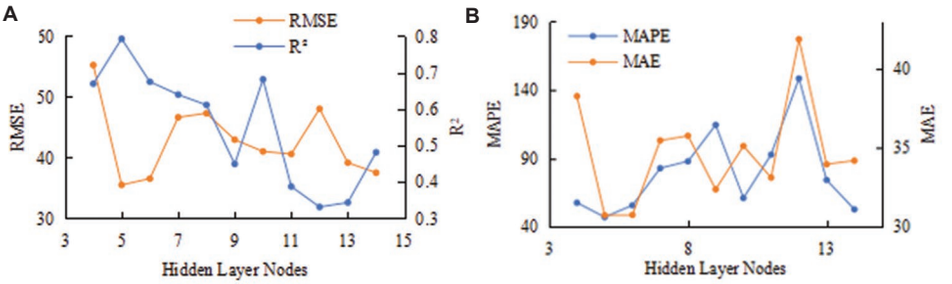

Figure 5. Error and R^2 of different hidden layer nodes. (A) RMSE and R^2; (B) MAE and MAPE.
Abbreviations: MAE: Mean absolute error; MAPE: Mean absolute percentage error; RMSE: Root mean square error; R^2: Correlation coefficient.

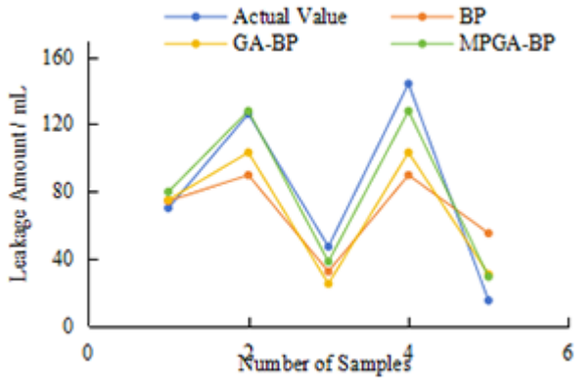

Figure 6. Prediction results of different optimization algorithms
Abbreviations: BP: Backpropagation; GA: Genetic algorithm; MPGA: Multiple population genetic algorithm.

actual observed values. Slight deviations were observed for a few data points, indicating that the optimized neural network model exhibits good generalization performance and adaptability.

5.2. Comparative analysis of different prediction models

To conduct a comparative analysis of the accuracy of the MPGA–BPNN prediction model, this study introduced three other prediction models: Random forest (RF), extreme learning machine (ELM), and polynomial fitting (PF). The same dataset was used to train these prediction models, and their prediction results were compared and analyzed. The prediction errors and correlation coefficients of each model are shown in Figure 8 and Table 6.

Under normal circumstances, if the correlation coefficient between the predicted values and the actual values of a model exceeds 0.8, it indicates that the model has a sufficiently high goodness of fit and the

Table 5. Error and R^2 values of different optimization algorithms

Index	BP	GA–BP	MPGA–BP
R^2	0.7929	0.9079	0.9741
RMSE	35.5205	25.708	12.1429
MAPE	47.1231	47.779	20.542
MAE	30.7267	22.0998	10.6277

Abbreviations: BP: Backpropagation; GA: Genetic algorithm; MAE: Mean absolute error; MAPE: Mean absolute percentage error; MPGA: Multiple population genetic algorithm; RMSE: Root mean square error; R^2: Correlation coefficient.

Table 6. Analysis of the prediction results of different models

Index	MPGA–BP	RF	ELM	PF
R^2	0.9741	0.7671	0.7062	0.9808
RMSE	12.1429	33.3303	39.2206	17.889
MAPE	20.542	51.4911	86.0494	24.4293
MAE	10.6277	29.3127	37.4922	15.1773

Abbreviations: BP: Backpropagation; ELM: Extreme learning machine; MAE: Mean absolute error; MAPE: Mean absolute percentage error; MPGA: Multiple population genetic algorithm; PF: Polynomial fitting; RMSE: Root mean square error; RF: Random forest; R^2: Correlation coefficient.

regression fitting effect is quite good. As shown in Figure 8 and Table 6, the prediction error of the ELM is relatively large. Although the RF and PF models had lower errors than the ELM, there was still a noticeable gap compared with the MPGA–BP model. It is believed, based on the analysis, that the ELM is prone to overfitting, leading to poor prediction performance, especially for small-sample datasets. In addition, its ability to handle nonlinear problems was limited, resulting in unsatisfactory prediction performance. Moreover, the RF was an ensemble algorithm composed of decision trees, and its performance was also suboptimal when

Table 7. Relationship between different parameters and relative error

GZD-A (%)	GZD-B (%)	GZD-C (%)	GZD-D (%)	D_o (mm)	D_e (mm)
0.081	−0.17	−0.065	−0.005	−0.02	−0.042
GDJ-I (%)	**GDJ-II (%)**	**GDJ-III (%)**	**GDJ-IV (%)**	**SL (mL)**	
−0.118	0.132	−0.164	−0.05	−0.327	

Abbreviations: D_o: Width of the open end; D_e: The terminal width; SL: Slurry loss.

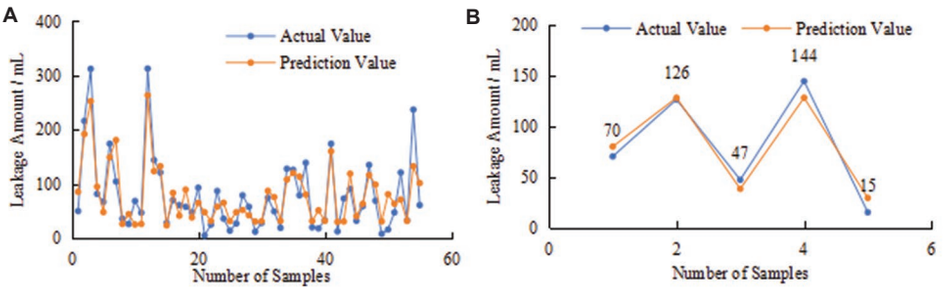

Figure 7. Prediction results of the multiple population genetic algorithm–backpropagation prediction model. (A) Training results; (B) Prediction results.

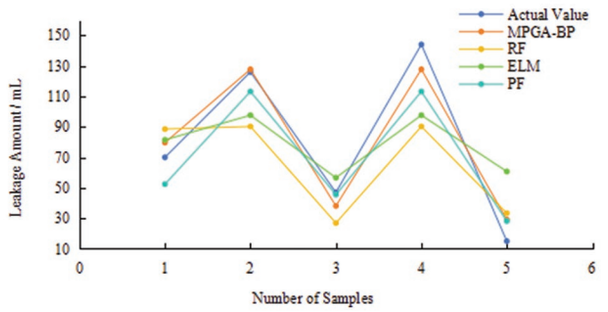

Figure 8. Prediction results of different prediction models

Abbreviations: BP: Backpropagation; ELM: Extreme learning machine; MPGA: Multiple population genetic algorithm; PF: Polynomial fitting; RF: Random forest.

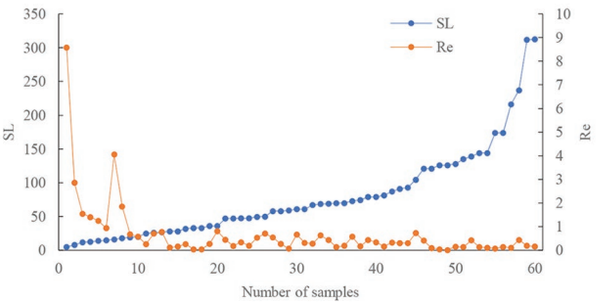

Figure 9. Relationship between slurry loss (SL) and relative error (Re)

dealing with datasets that have few features or small sample sizes. In contrast, the MPGA–BP model exhibited the smallest prediction error and produced results that are closer to the actual observed values, demonstrating high prediction accuracy. During model construction, the thresholds and weights of the BPNN were optimized using a genetic algorithm, which effectively improved the performance of the BPNN and further enhanced the accuracy and generalization capability of the prediction model.

5.3. Model adaptability analysis

Analyzing the predictive performance of a machine learning model under diverse conditions offers a deeper understanding of the model and helps assess its adaptability, thereby maximizing its value. Consequently, numerous researchers have analyzed the predictive performance of machine learning models under different parameters.[48] In this study, the computed relative error (Re) of the model's prediction outcomes was calculated, and the correlations between Re and multiple parameters were examined. The corresponding results are presented in Table 7.

As shown in Table 7, the parameters ranked by correlation strength with Re (from highest to lowest) are slurry loss, GZD-B, GDJ-III, GDJ-II, GDJ-I, GZD-A, GZD-C, GDJ-IV, D_e, D_o, and GZD-D. Notably, GZD-A and GDJ-II exhibit positive correlations with Re, whereas the other factors display negative correlations. The factor showing the strongest correlation with Re is slurry loss. To conduct a more detailed analysis of the relationship between Re and slurry loss, this study plotted the correlation curve between these two variables, as illustrated in Figure 9.

As shown in Figure 9, within the leakage range of 0–320 mL, Re generally exhibits a decreasing trend as the leakage increases. It can be concluded that the model developed in this study achieves high prediction accuracy for leakage cases but exhibits relatively lower prediction performance for small leakage cases.

6. Conclusion

In this study, relevant data were collected to analyze the factors influencing drilling fluid leakage, and a prediction model for drilling fluid leakage was established using a BPNN. The model structure was further optimized using the MPGA to achieve high predictive accuracy. The main conclusions are as follows:

(i) Using rigid mineral particle plugging agents, composite plant fibers, D_o, and D_e as input parameters, and the leakage amount as the output parameter, a new drilling fluid leakage prediction model was developed based on the BPNN. The model provides useful guidance for optimizing lost circulation control and plugging agents.

(ii) The number of hidden layer nodes in the model was optimized, and the model was further enhanced using the MPGA. The final prediction model achieved an R^2 of 0.9741, indicating high prediction accuracy and enabling reliable predictions of drilling fluid leakage under different formulations.

(iii) Due to the limited dataset, this study examined only two material systems (GZD and GDJ). With further research and an increase in available data, additional materials can be incorporated to expand the model's application range and improve its prediction accuracy.

Acknowledgments

The authors are grateful for the technical and logistical support of the Science and Technology Project of China National Petroleum Corporation Offshore Engineering Company Limited.

Funding

This research was supported by the Fundamental Research Funds for the Central Universities (grant number: 24CX02013A) and the Science and Technology Project of China National Petroleum Corporation Offshore Engineering Company Limited (grant number: 202303-0101).

Conflict of interest

Wenbing Wu, Tao Liu, Lianlu Huang, and Hua Li were employed by the Drilling Company, China National Petroleum Corporation (CNPC) Offshore Engineering Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The Drilling Company, CNPC Offshore Engineering Company Limited, had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Author contributions

Conceptualization: Wenbing Wu, Tao Liu
Data curation: Wenbing Wu
Formal analysis: Hua Li, Justine Kiiza, Moussa Camara
Investigation: Hua Li, Jie Zhong, Jiafang Xu
Methodology: Wenbing Wu
Resources: Chenxin Wang
Software: Wenbing Wu
Supervision: Lianlu Huang
Validation: Tao Liu, Lianlu Huang, Jian Wang
Visualization: Lianlu Huang
Writing–original draft: Wenbing Wu
Writing–review & editing: Tao Liu

Availability of data

The raw data supporting the conclusions of this article will be made available by the corresponding authors on reasonable request.

References

1. Bai Y, Liu C, Sun J, *et al*. High temperature resistant polymer gel as lost circulation material for fractured formation during drilling. *Colloids Surf A Physicochem Eng Asp.* 2022;637:128244. doi: 10.1016/j.colsurfa.2021.128244
2. Kang Y, Tan Q, You L, Zhang X, Xu C, Lin C. Experimental investigation on size degradation of bridging material in drilling fluids. *Powder Technol.* 2018;342:54-66. doi: 10.1016/j.powtec.2018.09.086
3. Ma C, Feng Y, Dou Y, *et al*. Experimental study on the design method of lost circulation materials for induced fractures. *Geoenergy Sci Eng.* 2024;240:213086. doi: 10.1016/j.geoen.2024.213086
4. Nazemi R, Moghadasi J, Ashoori S. Design and

experimental study on rheological behavior and sealing performance of shear sensitive fluids to control lost circulation during drilling. *Geoenergy Sci Eng.* 2024;237:212830.
doi: 10.1016/j.geoen.2024.212830

5. Liu J, Zhang Y, Zhang D, *et al*. Multi-sized granular suspension transport modeling for the control of lost circulation and formation damage in fractured oil and gas reservoirs. *Processes*. 2023;11(9):2545.
doi: 10.3390/PR11092545.
6. Zeng C, Wei A, Liu S, *et al*. Study on self-degradation drilling fluid for reservoir protection in a block of South China Sea. *Energy Chem Ind.* 2023;44(2):53-57.
doi: 10.3390/pr11061802
7. Mansoor IM, Tuama MH, Humaidi JA. Application of correlation-based recurrent neural network in porosity prediction for petroleum exploration. *Eng Res Express*. 2025;7(1):15241.
doi: 10.1088/2631-8695/ADA664
8. Elmaadawy GK, Hassan EAMM, Sallam MA. Rock typing and reservoir quality analysis of the abu madi reservoir: Distribution prediction using artificial neural networks in the West El Manzala Area, Onshore Nile Delta, Egypt. *Arab J Sci Eng*. 2024;49(1):913-944.
doi: 10.1007/S13369-023-08403-6
9. Han F, Zhang H, Rui J, *et al.* Multiple point geostatistical simulation with adaptive filter derived from neural network for sedimentary facies classification. *Mar Pet. Geol.* 2020;118:104406.
doi: 10.1016/j.marpetgeo.2020.104406
10. Luz LME, Spina ED, Pinto CJ. Application of transfer learning for modeling slugging in offshore oil production. *Geoenergy Sci Eng*. 2025;255:214069.
doi: 10.1016/j.geoen.2025.214069
11. Ewees AA, Qaness AAAM, Thanh VH, *et al*. Optimized neural networks for efficient modeling of crude oil production. *Knowl Inf Syst.* 2025;67:6171-6192.
doi: 10.1007/s10115-025-02415-4
12. Tadesse GC, Mamo BN. Artificial neural network and regression models for predicting intrusion of non-reacting gases into production pipelines. *Energies*. 2022;15(5):1725-1725.
doi: 10.3390/en15051725
13. Mayet MA, Alizadeh MS, Parayangat M, *et al*. ACO-based feature selection and neural network modeling for accurate gamma-radiation based pipeline monitoring in the oil industry. *Appl Radiat Isot*. 2025;215:111587-111587.
doi: 10.1016/j.apradiso.2024.111587
14. Salehuddin NF, Omar MB, Ibrahim R, Bingi K. A neural network-based model for predicting saybolt color of petroleum products. *Sensor*. 2022;22(7):2796.
doi: 10.3390/s22072796
15. Zhou D, Zhou C, Zhang Z, *et al*. Intelligent lost circulation monitoring method based on data augmentation and temporal models. *Processes*. 2024;12(10):2184.
doi: 10.3390/PR12102184
16. Wu S, Hu Y, Zhang L, Liu S, Xie R, Yin Z. Intelligent risk identification for drilling lost circulation incidents using data-driven machine learning. *Reliab Eng Syst Saf.* 2024;252:110407.
doi: 10.1016/J.RESS.2024.110407
17. Yao Q, Zheng X, Wang R, Liang W, Liu T, Chu W. Control of thermal uniformity in microwave heating process by BPNN and adaptive particle swarm optimization. *Heliyon*. 2024;10(21):e37971.
doi: 10.1016/j.heliyon.2024.e37971
18. Amish M, Khodja M. Review of detection, prediction and treatment of fluid loss events. *Arab J Geosci*. 2024;18(1):8.
doi: 10.1007/s12517-024-12142-9
19. Arbi MAB, Tamás M. Applying machine learning to predict the rate of penetration for geothermal drilling located in the Utah FORGE Site. *Energies*. 2022;15(12):4288.
doi: 10.3390/EN15124288
20. Ismail A, Rashid HMA, Gholami R, Raza A. Characterization based machine learning modeling for the prediction of the rheological properties of water-based drilling mud: An experimental study on grass as an environmental friendly additive. *J Pet Explor Prod Technol.* 2021;12(6):1677-1695.
doi: 10.1007/s13202-021-01425-6
21. Amir I, Farouk HE, Sahar N, *et al*. Gas channels and chimneys prediction using artificial neural networks and multi-seismic attributes, offshore West Nile Delta, Egypt. *J Pet Sci Eng.* 2022;208:109349.
doi: 10.1016/J.PETROL.2021.109349
22. Zhang N, Li F, Ren B, *et al*. Research on wellbore trajectory control of Rotary Steerable System using back-propagation neural network-fuzzy method. *Geoenergy Sci Eng.* 2025;257:214201.
doi: 10.1016/J.GEOEN.2025.214201
23. Xie J, Zhang J, Cheng Y, *et al*. Real-Time prediction of wellbore temperatures in deep shale gas drilling using a combination of PINN and heat transfer models. *Appl Thermal Eng.* 2025;279:127984.
doi: 10.1016/J.applthermaleng.2025.127984
24. Luo Y, Ma B, Wang X, Zhang J, Ma Y, Liu C. A new well loss prediction model was created by applying genetic algorithms to enhance the RBF neural network. *J Phys Conf Ser.* 2025;3048(1):012128.
doi: 10.1088/1742-6596/3048/1/012128
25. Zhang Z, Ertekin T, Ma X, Zhan J. An artificial-intelligence-based petrophysical property predictor for compositional volatile oil reservoir using three-phase production data. *Energy Explor Exploitation*. 2024;42(4):1284-1314.
doi: 10.1177/01445987231221593
26. Tabatabaei M, Attari N, Panahi SA, Asadian-Pakfar M,

Sedaee B. EOR screening using optimized artificial neural network by sparrow search algorithm. *Geoenergy Sci Eng*. 2023;229:212023.
doi: 10.1016/j.geoen.2023.212023

27. Chenji W, Ruijie H, Mingming D, *et al.* Characterization of saturation and pressure distribution based on deep learning for a typical carbonate reservoir in the Middle East. *J Pet Sci Eng.* 2022;213:110442.
doi: 10.1016/j.petrol.2022.110442
28. Lang X, Wang Z, Cao J, *et al*. A multidirection three-dimensional fusion neural network for irregular defect size estimation on magnetic flux leakage detection system. *Measurement*. 2025;257:118907.
doi: 10.1016/j.measurement.2025.118907
29. Wang F, Zhang Y, Xu Y, Zheng Q. Lightweight real-time network for multiphase flow patterns identification based on upward inclined pipeline pressure data. *Flow Meas Instrum*. 2025;102:102840.
doi: 10.1016/j.flowmeasinst.2025.102840
30. Fu J, Zou S, Sun J, Xu Q. Self-adaptive early warning of undesirable gas-liquid flow pattern in offshore oil and gas pipeline-riser system. *Process Saf Environ Prot*. 2023;182:254-278.
doi: 10.1016/j.psep.2023.11.055
31. Ren J, Shi X, Cao X. Fire recognition method based on PSO-BP neural network and ResNet50. *Int J Pattern Recognit Artif Intell.* 2024;38:2450018.
doi: 10.1142/S0218001424500228
32. Cui Y, Liu H, Wang Q, *et al*. Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network. *Combust Flame*. 2022;237:11852.
doi: 10.1016/j.combustflame.2021.111852
33. Guliyev NJ, Ismailov VE. A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function. *Neural Comput.* 2016;28(7):1289–1304.
doi: 10.1162/neco_a_00849S
34. Huang GB, Chen YQ, Babri H. Classification ability of single hidden layer feedforward neural networks. *IEEE Trans Neural Netw.* 2000;11(3):799-801.
doi: 10.1109/72.846750
35. Uzair M, Jamil N. Effects of Hidden Layers on the Efficiency of Neural Networks. In: *Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, 2020.
36. Wang L. *Research on Key Parameters of Drilling Leak Prevention and Sealing Based on BP Neural Network*. Master Thesis, Southwest Petroleum University. Chengdu, China; 2019.
37. Wang J, Zhao M, Wang B, *et al*. Prediction of coal seam permeability by hybrid neural network prediction model. *J Energy Eng.* 2024;150(4):040240211.
doi: 10.1061/jleed9.eyeng-5358
38. Ma J, Xu W. GA-BP-based nonlinear time series forecasting: Method and applications. *Acad J Comput Inf Sci.* 2024;7(8):15-20.
doi: 10.25236/AJCIS.2024.070803
39. Li Q, Li Z. Research on failure pressure prediction of water supply pipe based on GA-BP neural network. *Water*. 2024;16(18):2659.
doi: 10.3390/W16182659
40. Xu W, Ma J. Temperature prediction based on NEAT-optimized GA-BP neural network. *Autom Mach Learn.* 2024;5(2):25-32.
doi: 10.23977/autml.2024.050204
41. Tuo L. Application of multi population genetic algorithm in image processing. *Software*. 2023;44(8):143-146.
42. Zhang X, Hou L, Liu J, Zhong P, Lu H, Sigama A. Energy consumption prediction for crude oil pipelines based on integrating mechanism analysis and datamining. *Energy*. 2022;254:124382.
doi: 10.1016/j.energy.2022.124382
43. Zhu C, Guo B, Zhang Z, *et al*. Determining rock joint peak shear strength based on GA-BP neural network method. *Appl Sci.* 2024;14(20):9566.
doi: 10.3390/app14209566
44. Tang SZ, Li MJ, Wang FL, He YL, Tao WQ. Fouling potential prediction and multi-objective optimization of a flue gas heat exchanger using neural networks and genetic algorithms. *Int J Heat Mass Transf.* 2020;152:119488.
doi: 10.1016/j.ijheatmasstransfer.2020.119488
45. Wei L, Wu Y, Fu H, Yin Y. Modeling and simulation of gas emission based on recursive modified Elman neural network. *Math Probl Eng.* 2028;2018:9013839.
doi: 10.1155/2018/9013839
46. Yin H, Zhou X, Lang N, *et al*. Prediction model of water inrush from coal floor based on GA-BP neural network optimized by SSA and its application. *Coal Geol Explor.* 2021;49(6):175-185.
doi: 10.3969/j.issn.1001-1986.2021.06.021
47. Ye K, Wang J, Gao H, *et al*. Optimization of lapping process parameters of CP-Ti based on PSO with mutation and BPNN. *Int J Adv Manuf Technol.* 2021;117(9-10):2859-2866.
doi: 10.1007/S00170-021-07862-1
48. Hui G, Chen S, He Y, *et al.* Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors. J Nat Gas Sci Eng 2021;94:104045.
doi:10.1016/J.JNGSE.2021.104045.