REGULAR ARTICLE

OPEN ACCESS

Prediction of drilling leakage locations based on optimized neural networks and the standard random forest method

Junlin Su[1], Yang Zhao[*], Tao He[2], and Pingya Luo[1]

[1] Institute of Petroleum and Natural Gas Engineering, Southwest Petroleum University, Chengdu 610000, China
[2] China National Petroleum Corporation Chuanqing Drilling Engineering Co., Ltd., Drilling Fluid Technology Service Company, Chengdu 610000, China

Abstract. Circulation loss is one of the most serious and complex hindrances for normal and safe drilling operations. Detecting the layer at which the circulation loss has occurred is important for formulating technical measures related to leakage prevention and plugging and reducing the wastage because of circulation loss as much as possible. Unfortunately, because of the lack of a general method for predicting the potential location of circulation loss during drilling, most current procedures depend on the plugging test. Therefore, the aim of this study was to use an Artificial Intelligence (AI)-based method to screen and process the historical data of 240 wells and 1029 original well loss cases in a localized area of southwestern China and to perform data mining. Using comparative analysis involving the Genetic Algorithm-Back Propagation (GA-BP) neural network and random forest optimization algorithms, we proposed an efficient real-time model for predicting leakage layer locations. For this purpose, data processing and correlation analysis were first performed using existing data to improve the effects of data mining. The well history data was then divided into training and testing sets in a 3:1 ratio. The parameter values of the BP were then corrected as per the network training error, resulting in the final output of a prediction value with a globally optimal solution. The standard random forest model is a particularly capable model that can deal with high-dimensional data without feature selection. To evaluate and confirm the generated model, the model is applied to eight oil wells in a well site in southwestern China. Empirical results demonstrate that the proposed method can satisfy the requirements of actual application to drilling and plugging operations and is able to accurately predict the locations of leakage layers.

1 Introduction

Circulation loss is a common but complex occurrence during the drilling process. Downhole leakage considerable increases drilling cost and downtime [1–3], which often leads to serious accidents because leakage management is a tedious process [4]. Furthermore, the reduction in the height of the liquid column in the wellbore is reduced, resulting in a decrease in effective static liquid column pressure at the well's bottom; this may easily lead to an imbalance in the formation of pore pressure, causing overflow and even blow-out [5]. Oil companies and universities have spent considerable amount of money on research related to the problem of drilling fluid leakage and plugging and explored various avenues including hole reinforcement [6], plugging materials [7], mud flow law [8–10], drilling fluid diffusion pressure [11] and drilling fluid formulas [12, 13].

Identifying the location of the circulation loss layer is a key element of dealing with the problem of circulation loss,

and accurate judgment in that regard can greatly aid decision-making in the field. Currently, various instruments are used to measure the location of the leakage layer, including acoustic testers, eddy current testers, radioactive tracers, and well temperature testers [14, 15]. It is not only needs professional team, but also has high cost and difficult to popularize. Furthermore, it is extremely difficult to quickly obtain leakage layer prediction results with measurements from such instruments during the occurrence of a well leakage. Therefore, to determine the location of the leakage layer, the engineering practice of plugging typically relies on experience or test plugging, which has both poor accuracy and is wasteful in terms of human and material resources. Therefore, for high leakage loss in wells around the world, the lack of a method for predicting the location of leakage layers is one of the primary reasons.

Although Artificial Intelligence (AI) emerged in the 1950s, it is still a relatively new area of science that studies and develops the theory, method, technology, and application of systems that are used to simulate, extend, and expand human intelligence. To date, it has been successfully

* Corresponding author: 1275728683@qq.com

applied to various industries, including economics, computer science, financial trade, medical diagnosis, industry, transportation, and telematics [16–18]. Currently, AI technology has proven able to solve certain difficult problems in the oil industry [19–23].

Because traditional methods cannot completely or even effectively address the issue, many researchers have previously experimented with different approaches for using AI methods to reduce the cost of drilling and plugging and to improve the success rates of plugging, including predicting the direction of lost circulation in a coordinate system [24], formation type and lithology [25], and drilling fluid density [26].

Currently, AI has been involved in many classic cases in the field of drilling and plugging; however, it is far from being able to form a complete system, and there is the conspicuous lack of an AI model for predicting leakage layer locations. Therefore, it is necessary to construct an AI model to predict leakage layer locations.

This study aims to use real-time well history data and drilling parameters to predict the locations of circulation loss layers and to propose a new and effective prediction method for use in drilling and plugging operations. Through a literature review and practice comparison, we concluded that the Genetic Algorithm-Back Propagation (GA-BP) neural network and the standard random forest model were best suited to mining potentially useful information from the drilling data such as different lithologies and the magnitude of the impact made by different amounts of drilling pressure on the location of the leakage layer. Using AI to precisely analyze the data related to these factors, a potential law for predicting leakage layer location may be derived.

2. Data source

In this study, the original data set used was of 1.4 million well history records selected from 240 wells in southwestern China, including 1029 cases of circulation loss.

Since these 240 wells are all vertical wells or near vertical wells (the deviation angle is within 15°), the parameters such as well deviation angle, azimuth angle, deviation change rate and closure distance are not considered in this study. Therefore, there were 20 unfiltered parameters including well depth, vertical pressure, outlet density, inlet density, equivalent density, horizon, hook load, lithology, bit type, bit size, total pool volume, rotation speed, torque, outlet flow, drilling speed, inlet flow, Weight On Bit (WOB), drilling fluid density, funnel viscosity, and three-turn reading, all of which may be related to predicting the location of circulation loss in the drilling data set. Thus, correlation analysis was required to determine whether they were input parameters.

The well history data focused on the official drilling and logging data, which reflects the well's formation information, drilling time records (whole meter data), full performance of drilling fluid, bit usage, drilling fluid records for each shift, and other data information closely related to the entire drilling process (Tab. 1). The parameters of the

drilling machinery, most of which were based on instrument readings, were obtained from real-time sensors. The primary difficulty with collecting data in a well site is the quality of data. Over the course of drilling, the levels of data measurement uncertainty and inaccuracy can be rather high, mostly because of both operator and equipment error. Because of the abovementioned problems, it was necessary in this study to preprocess the measured well site data so as to improve the accuracy of prediction results.

In this process, the mud signal is converted to the front-end of the drilling fluid, and converted to the downhole signal by the mwd-mud technology The results of mud MWD deep well adaptability test show that the system can work normally at well depth of 7000 m and temperature below 150 °C, so as to ensure the reliability of data source.

3 Data preprocessing

Before data mining using AI, it was necessary to preprocess the data, the purpose of which was to eliminate and modify certain factors that would affect data mining of the well history data and drilling parameters. The preprocessing procedure primarily included data parameter selection, data coding, processing of abnormal and missing values, and data specification.

3.1 Parameter selection

In this study, the ANOVA method proposed by Fisher was used to select parameters for analyzing drilling leakage data. This method involves the division of the total variance of measurement data, depending on the source of variations, according to processing (inter group) effects and error (intra group) effects and making a quantitative estimation to determine multiple parameters that influence the research results (*i.e.*, the location of the leakage layer).

SPSS can then be used along with the ANOVA method to measure the relationships between variables when determining the attribute parameters of leakage layer location. By inputting a large number of data parameters, the strength of the relationships between variables can be accurately measured, and the most influential parameter variables related to leakage layer location can be determined. A disadvantage of this method is that it cannot use these relationships to predict data; moreover, it does not refine or solidify the relationships between variables to form a model. Therefore, it needs to be used along with a GA-BP neural network or other models.

3.2 Data coding

There is a large amount of textual information recorded in Chinese characters and English in the data mining database regarding lithology, horizon, and bit type. However, this type of textual information cannot be directly utilized in data mining; therefore, it needs to be further processed and digitized. Because of the disorder between categories, we were unable to use natural ordinal coding. Instead, we used the unique technology of hot coding.

Table 1. Data sources.

	Classification	Data
Logging information	Depth	Drilling time record
	Formation information	Cuttings logging data
		Stratigraphic stratification
Drilling information	Bit usage	Bit usage
	Drilling log	Drilling performance
		Daily BHA
		Daily bit
	Drilling fluid information	Actual measurement of drilling fluid instruments
		Usage record of drilling fluid

Table 2. Text data of lithology used in this study.

Lithology	Code	Lithology	Code
Limestone	000001	Carbonaceous mudstone	001101
Mudstone	000010	Coal seam	001110
Argillaceous limestone	000011	Nodular limestone	001111
Shale	000100	Medium sandstone	010000
Bauxite mudstone	000101	Fine sandstone	010001
Pelitic siltstone	000110	Siltstone	010010
Silty mudstone	000111	Sandy shale	010011
Shell limestone	001000	Gypsum layer	010100
Medium shale	001001	Dolomitic limestone	010101
Tuff	001010	Gypsum dolomite	010110
Pebbly fine sandstone	001011	Siliceous mudstone	010111
Pebbly siltstone	001100	Siliceous siltstone	011000

Single hot coding, also known as one-bit effective coding, uses n-bit status registers to code n states; each state has its own independent register bits, and only one of them is valid at any one time. The advantage of this method is that it can ignore the code–size relationship induced by direct coding, and thus avoid the prediction error introduced by the size relationship of parameter codes such as lithology, bit type, and horizon (Tab. 2).

3.3 Abnormal data value and missing value processing

The box chart method is used to detect and process outliers; it can intuitively display and eliminate the outliers in a large amount of data. The next step after removing outliers is to fill in the fields with a missing rate of <30% (a few pieces of data with a missing rate of >30% can be considered untrusted and subsequently deleted). According to the principle that the parameters represented by the fields can be collected and obtained at the drilling site, as much data information can be retained in the cleaning data as possible for the next processing, so the data can be divided into different areas. The Newton interpolation method was used to fill in the gaps in different regions.

ID index primary key fields and flag fields were added to all data tables in the data warehouse to record whether a piece of data has been pushed to data specification. For pushed records, the value was recorded as 1. For newly entered data records that were not pushed, the value was recorded as 0.

3.4 Data specification and normalization

To avoid the phenomenon of data inundation because of very few pieces of leakage point data, well depth without leakage was selected from the original data, and data reduction was made in units of 10 m. Most non-leakage data was thus abandoned, and the surrounding data was able to be used to represent these non-leakage data points, thus improving the model's prediction accuracy.

Furthermore, to prevent parameters from being influenced by dimension and data value ranges at the beginning of training, the Pandas module in Python was used to arrange all sample data in [0, 1] sets.

After the above data preprocessing process, the original 96 data tables of >125 000 well history records, including 1029 well history records, were consolidated into one data table comprising 57 576 well history records, including 661 well history records (Fig. 1). On this basis of the aforementioned, the final model of data mining achieved a high level of precision.
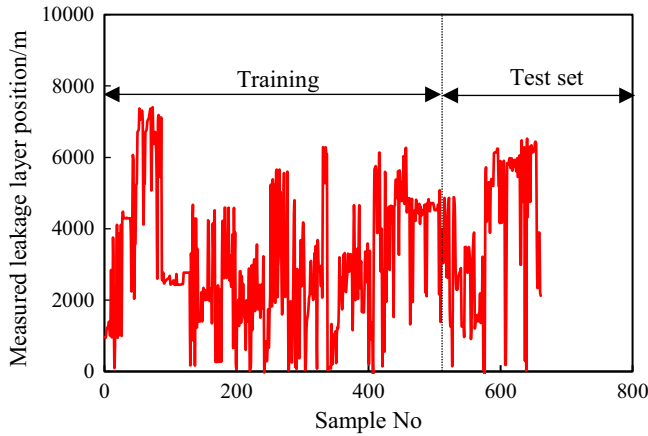
Fig. 1. Overall distribution results of circulation loss data sets after pretreatment.

4 Model introduction

4.1 GA-BP neural network

The BP neural network is the most extensively used error backpropagation training algorithm. It presents the distinct advantage of mapping well to any non-linear relationship. Figure 2 shows the structure of a general three-layer BP network model. It is a feedforward neural network [27, 28], which describes multiple linear and nonlinear uncertain mapping relationships in which each neuron can learn and store a large number of mapping relationships without understanding the exact mathematical equation for the unknown mapping relationship.

Using the parameters provided in Table 1, a knowledge base with 18 tables and 30 056 well history records was developed; neural network model training was then performed using this knowledge base. For conducting this study, after many experiments and error analyses, the number of hidden layers did not have a great impact on the accuracy of the model's predictions. Therefore, this study used only three layers of the BP neural network (as shown in Fig. 2), namely, an input layer, an output layer, and a hidden layer. The input layer comprised 17 parameters, the dimension was 17; the output layer comprised 1 parameter; and the hidden layer neuron comprised 20 parameters.

GA can optimize the initial values of a BP network's parameters, which improves the correct recognition rate of the optimized output; the effect of this is significantly improved compared with general BP neural network. GA, as a global optimization algorithm, corrects the parameter values of BP according to errors in network training, and finally outputs a prediction value with a globally optimal solution, as shown in Figure 3.

The specific optimization process is as follows:

(1) Initialize the population and code:
Determine the total number of nodes in the BP structure and code the real number of population individuals; each individual includes all parameters of the BP neural network.
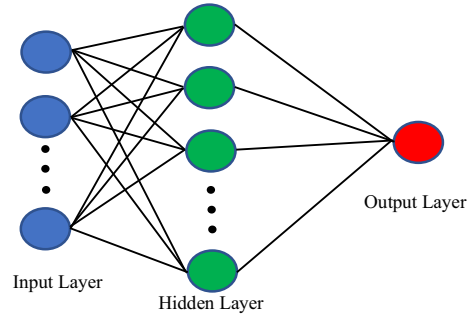

Fig. 2. Structure of the three-layer neural network.

(2) Individual fitness:
The predicted individual fitness is related to the absolute value and the error; the individual fitness value is then recorded as F:

$$F = k\left(\sum\nolimits_{i=1}^{n} \mathrm{abs}(y_i - o_i)\right), \qquad (1)$$

where k is the correction coefficient, n is the total number of predicted outputs, y_i is the true value of the ith output of the network, and o_i is the predicted value of the ith output.

(3) Selection:
The GA selected is related to the fitness value, and the selection of individual I is as follows:

$$\begin{cases} f_i = k/F_i \\ p_i = \dfrac{f_i}{\sum_{j=1}^{N} f_i} \end{cases}, \qquad (2)$$

where F_i is the fitness value of individual I, k is the correction coefficient, and N is the population size.

(4) Cross:
Crossover generates a new individual by exchanging parts of two original individuals. The crossover formula is as follows:

$$\begin{cases} a_{kj} = a_{kj}(1 - b) + a_{lj}b, \\ a_{lj} = a_{lj}(1 - b) + a_{kj}b. \end{cases} \qquad (3)$$

(5) Variation:
In GA, new individuals are generated by introducing mutations to improve population adaptability. The process involves the mutation of gene a_{ij}, the formula for which is as follows:

$$a_{ij} = \begin{cases} a_{ij} + \left(a_{ij} - a_{\max}\right)f(g), r > 0.5 \\ a_{ij} + \left(a_{\min} - a_{ij}\right)f(g), r \leq 0.5 \end{cases}, \qquad (4)$$

$$f(g) = t(1 - g/G_{\max})^2. \qquad (5)$$

Among them, $a_{\max}$ and $a_{\min}$ are the upper and lower bounds of a_{ij}, g is the number of evolutions, $G_{\max}$ is the maximum number of evolutions, and the variation probability is generally between 0.001 and 0.1.
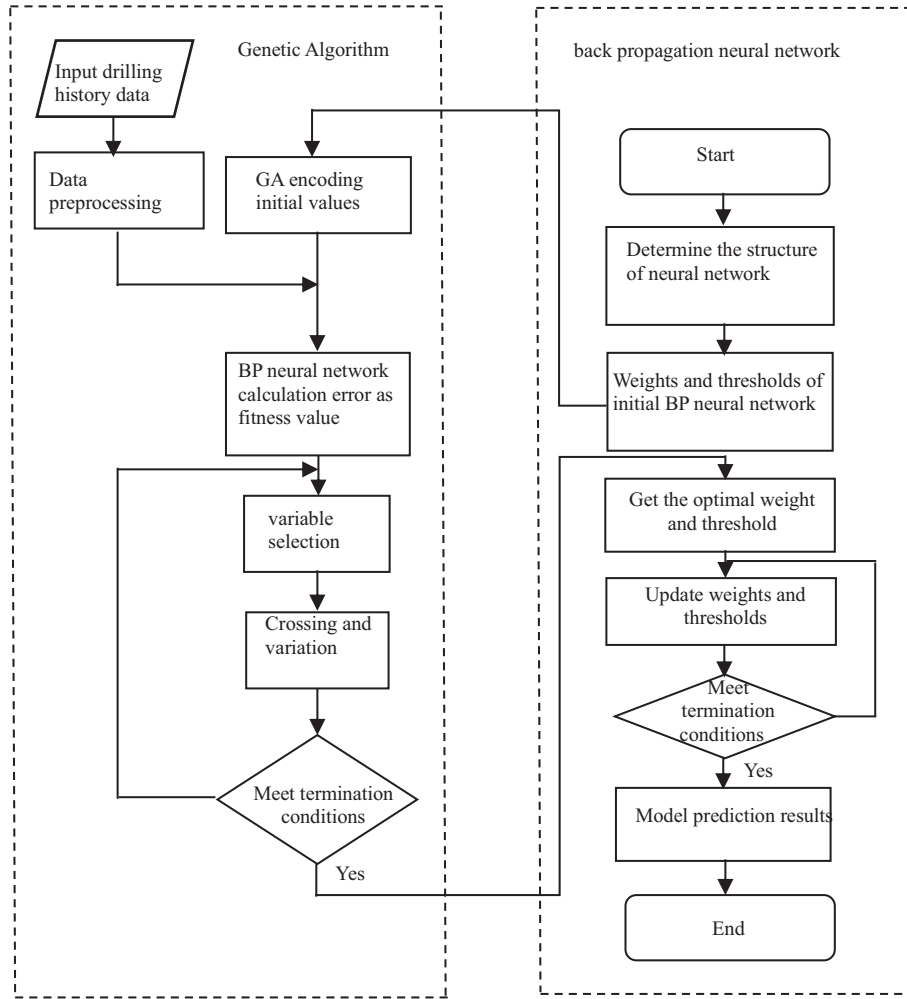
Fig. 3. Flow chart of a GA-optimized BP neural network.

4.2 Standard random forest

The Standard Random Forest (SRF) uses the idea of decision tree integration. In the forest, each tree is independent, and 99.9% of the unrelated trees make prediction results covering all situations. While these prediction results will offset each other, a few excellent trees will have good prediction results.

A typical problem with regression is the prediction of the location of drilling leakage layer based on well history data. To realize random forest regression, every decision tree in the random forest must be a regression tree. The process uses a recursive partition to divide the data into different homogeneous regions, subsequently averaging the results of all the regression trees. Each tree then independently grows to the maximum size (~70%) based on the guidance samples in the training data set without any pruning (*i.e.*, the selection of input variables will not be stopped at each node). For each tree, the SRF randomly selects a subset of variables (mtry) to determine the split at each node. The calculation formula at node v of Gini(v) with the Gini coefficient is as follows:

$$\text{Gini}(v) = \sum\nolimits_{j=1}^{p} \hat{p}_c^v (1 - \hat{p}_c^v), \qquad (6)$$

where $\hat{p}_c^v$ is the observation value of the jth variable at node v. The "Gini" coefficient gain (X_i, v) of X_i at the split node v is the impurity difference v between the nodes and sub-node v of nodes, which is updated as follows:

$$\text{Gain}(X_i, v) = \text{Gini}(X_i, v) - \omega_L \text{Gini}(X_i, v^L) \\ - \omega_R \text{Gini}(X_i, v^R), \qquad (7)$$

where v^L and v^R are the left and right child nodes of node v, respectively, and w_L and w_R are the ratios of the characteristic variables to the left and right child nodes, respectively. On each node, mtry (mtry $\approx \sqrt{p}$) randomly selects variables from among P variables and finally obtains the characteristic variable with the maximum information gain for the split of node v. The formula for calculating the importance X_i of the variable is as follows:

$$\text{imp}_i = \frac{1}{\text{ntree}} \sum_{v \in S_{X_i}} \text{Gain}(X_i, v), \qquad (8)$$
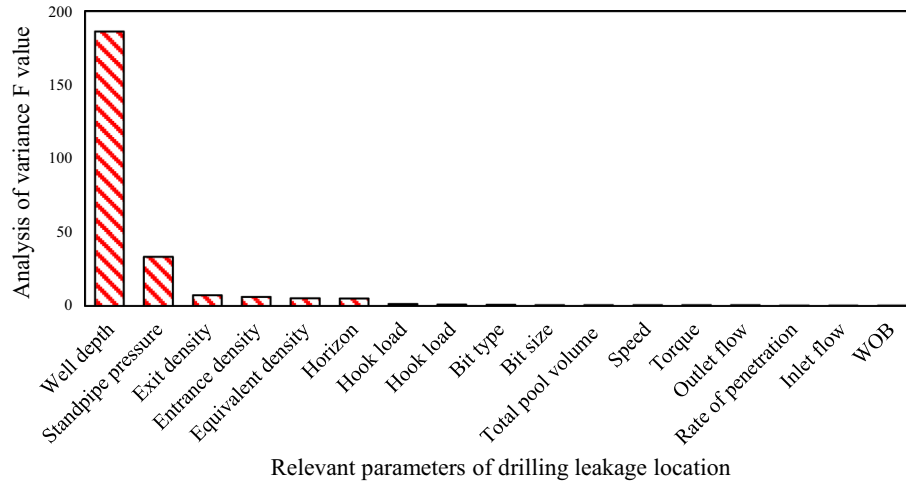
Fig. 4. Analysis of variance results (F value) of the relevant parameters of leakage layer position.

where SX_i is the set of nodes divided into X_i in the random forest of the nTree tree. Importance scores are used to evaluate the contribution of characteristic variables to the prediction.

The random forest process only comprises four steps: 1) bootstrap sampling of the original samples; 2) randomly selecting mtry features to establish the decision tree; 3) repeating the previous two steps nTree times (*i.e.*, forming an nTree decision tree of random forests); and 4) predicting the average value of all decision trees for new data. The principle of node splitting is then used to minimize prediction error.

5 Results and discussion

5.1 Correlation analysis of input parameters

As per the parameters of the collected well history data, the independent variables of the variance analysis were defined, and the parameters were determined as the main sources of intra-group error. The actual difference in locations of the leakage layer in the same block is the main performance of the intra-group error, and the location of the leakage layer was the dependent variable. Figure 4 shows the results obtained using SPSS.

Generally, α in the p value of a multifactor analysis is either 0.05 or 0.1. To consider as many factors as possible, α was considered as 0.1. Therefore, as shown in Figure 5, the 17 parameters that had the greatest influence on fracture width during variance analysis were selected as the input parameters of the neural network; these included well depth, vertical pressure, outlet density, inlet density, equivalent density, horizon, hook load, lithology, bit type, bit size, total volume of pool, rotation speed, torque, outlet flow, drilling speed, inlet flow, and WOB.

5.2 GA-BP neural network

To understand the prediction performance of the model after selecting the GA-BP neural network, 70% of the data

set was used as the training set for training the neural network model; the remaining 30% was used as the test set for model verification. Figures 6a and 6b show the results of the comprehensive statistic and a graphic error analyses of the model's performance. This figure shows the prediction scatter diagram for the GA-BP neural network model's training and test sets, respectively.

To confirm the reliability of the Artificial Neural Network (ANN) model's leakage layer position prediction results, the decision coefficient (R^2), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) of the two models (*i.e.*, the BP neural network and the GA-BP neural network) were calculated as follows:

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2},$$
(9)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2},$$
(10)

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|,$$
(11)

where y is the number of actual missing solutions, $\hat{y}$ is the number of correctly simulated missing solutions using machine learning methods, and n is the total number of data types used for model evaluation. As shown in Table 3, the model with the highest R^2, the lowest RMSE, and the lowest MAPE can be considered as the optimal model.

Table 3 shows that the GA-BP neural network that was optimized via the genetic algorithm was much more accurate and stable than the BP neural network that was not optimized.

5.3 Random forest

Note that 70% of the data set was used as the training set for decision tree model's training, and the remaining

Table 3. Prediction performance of the artificial neural network model in training and test sets.

Data sets	R^2	RMSE	MAPE	Number of data
Training set	0.94	0.03	0.09	40 303
Test set	0.86	0.05	0.33	17 273

Table 4. Prediction performance of the proposed random forest model in training and test sets.

Data sets	R^2	RMSE	MAPE	Number of data
Training set	0.92	0.09	0.14	40 303
Test set	0.89	0.16	0.17	17 273
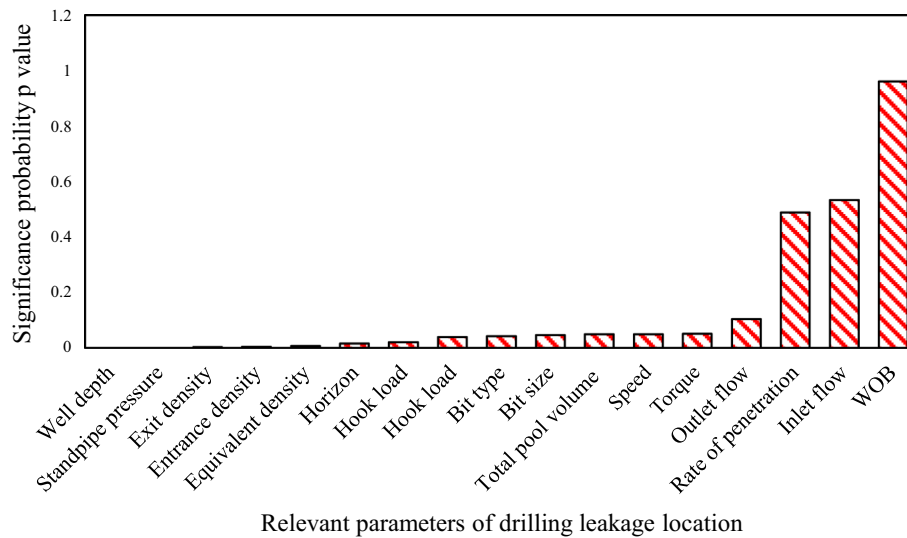
Fig. 5. ANOVA results of parameters related to leakage layer position (significance p value).

30% was used as the test set for model verification. The results are shown in Figures 7a and 7b, in which the scatter diagrams of the decision tree model and the random forest model are presented.

Similarly, the decision coefficient R^2, RMSE, and MAPE of the two models were calculated; the results are presented in Table 4.

5.4 Result analysis

Although the above scatter diagram and model parameters may reflect certain advantages and disadvantages of several of the prediction models, the results are still not intuitive. The data of for the training and test sets were thus input into the two models for analysis, and the prediction results and trend charts for the GA-BP neural network and random forest algorithm were obtained.

Based on the mining of well history data and drilling parameters by the GA-BP neural network and random forest algorithm, the GA-BP neural network and random forest methods played an extremely unique role in predicting

leakage layer position. As shown in Figure 8, the evaluation standards of each model, and the trend charts, not only the coefficient of determination (R^2) is very high, but also the prediction results are in good agreement with the actual data. The results thus measured agreed well with the actual data.

6 Model application example

The Permian layer in Block C of an oil field in southwestern China often encounters large-scale circulation losses. The conventional methods of bridge plugging, high filtration, cement plugging, and composite plugging have low success rates and take an inordinate amount of time. By summarizing geological and engineering data analyses and early plugging work, the primary difficulties with plugging in this leakage layer were characterized as follows: there are multiple natural horizontal and vertical fractures, and the location of the leakage layer is unclear, resulting in inaccurate mud injection. Whether the location of the leakage layer
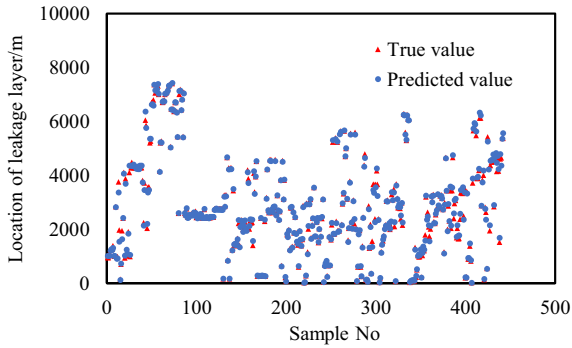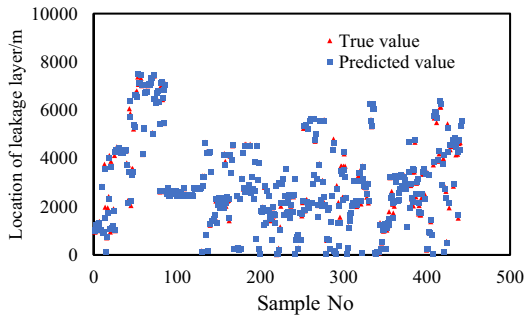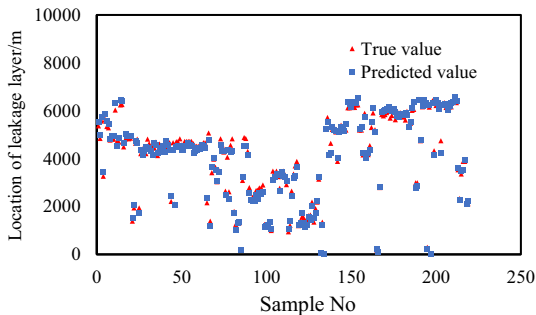
a. Training set prediction results
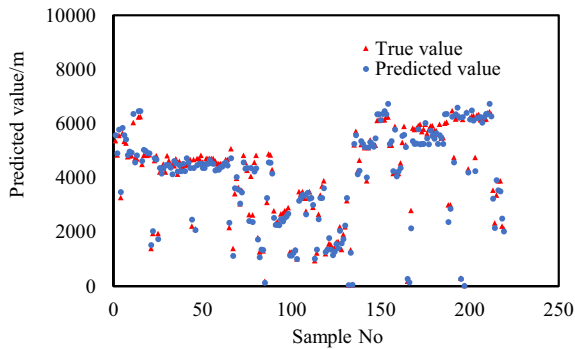
b. Test set prediction results

Fig. 6. Prediction scatter diagram of the training and test sets in the GA-BP neural network model.
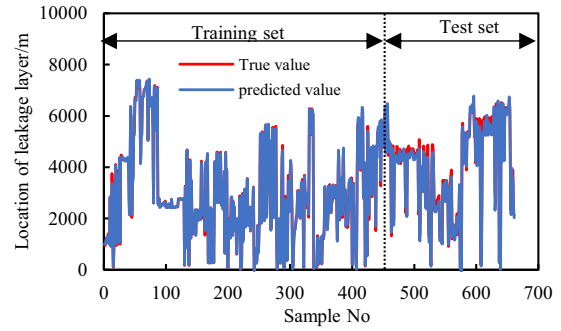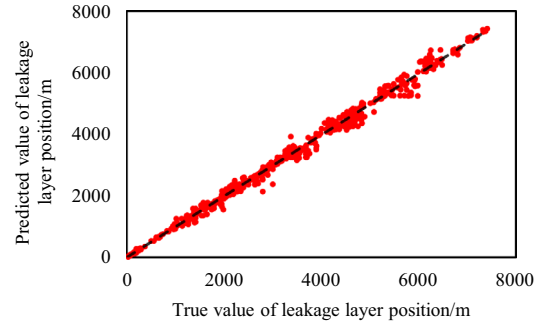

a. Training set prediction results
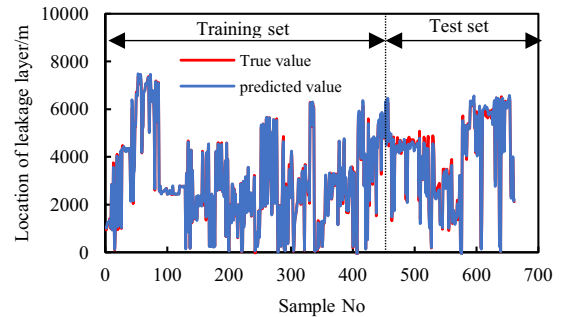

b. Test set prediction results

Fig. 7. Prediction scatter diagram of the training and test sets for the standard random forest model.
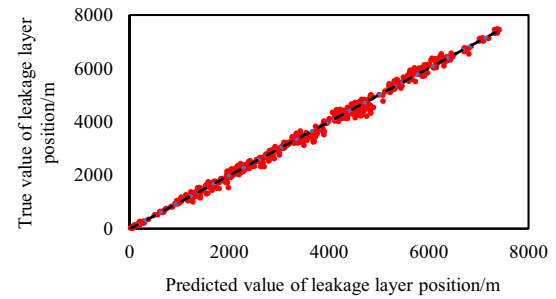

a. Prediction results of the GA-BP model

b. Prediction trend of the GA-BP model

c. Prediction results of the random forest model

d. Prediction trend of the random forest model

Fig. 8. Trend chart of leakage layer position prediction for the GA-BP and random forest models.

is determined to be high or low, final plugging will be affected. With a density of 1.15 g/cm³, the leakage layer is sensitive to pressure. When pressure increases to 1.21 g/cm³, leakage and even backflow of drilling fluid will occur.

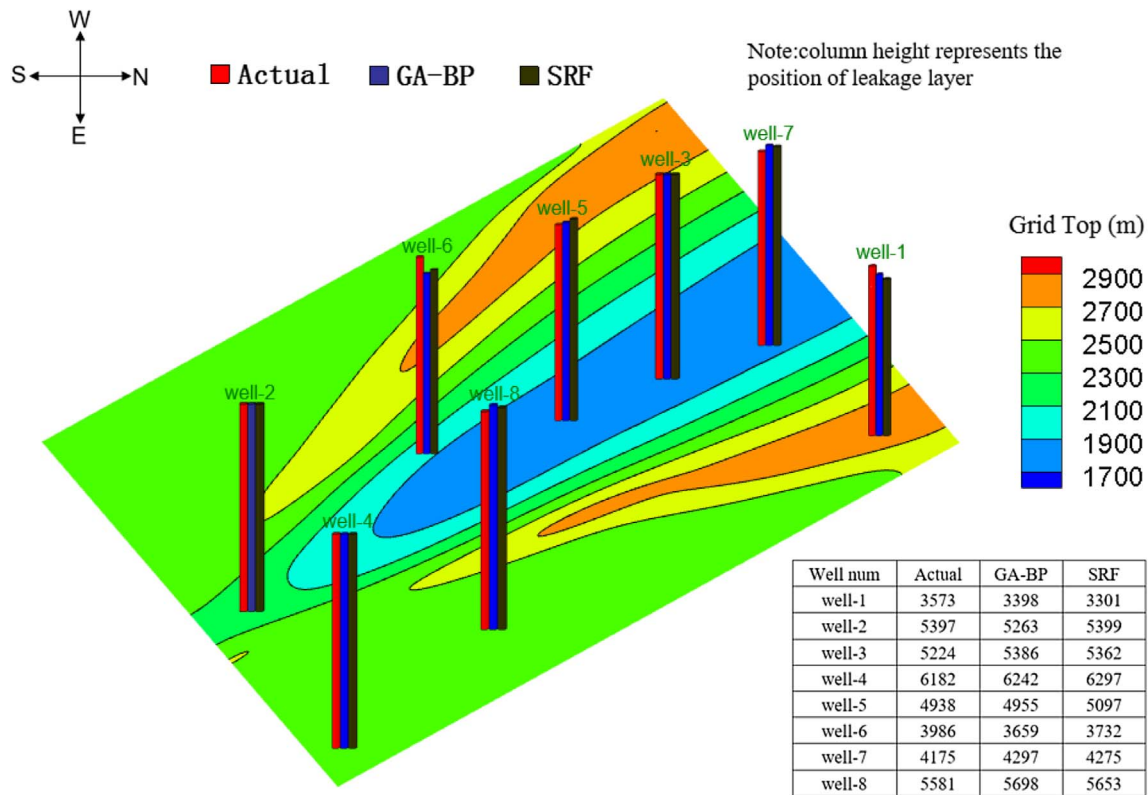| Well num	Actual	GA-BP	SRF
well-1	3573	3398	3301
well-2	5397	5263	5399
well-3	5224	5386	5362
well-4	6182	6242	6297
well-5	4938	4955	5097
well-6	3986	3659	3732
well-7	4175	4297	4275
well-8	5581	5698	5653

Fig. 9. Prediction results of leakage layer position of GA-BP model and random forest model.

In the target block, eight wells have lost circulation in the past three months. The established GA-BP neural network and random forest models were then used to predict real-time drilling data, and these prediction results were used to support decision-making during the plugging efforts. At last, five wells (numbers 1, 4, 5, 6, and 8) were successfully plugged once. Figure 9 shows the prediction results of the two models.

The results show that the prediction accuracy of the two models constructed by this method is basically consistent with the actual value, which can be used to predict the location of lost circulation zone and transmit it to the surface in real time, thus effectively guiding the lost circulation operation and providing scientific basis for the determination of lost circulation formula and plugging material. Even according to the change of prediction results of lost circulation location in the drilling process, combined with drilling conditions and operation parameters, the downhole formation fracture can be monitored to prevent the loss of drilling fluid caused by the unknown length and width of induced fracture caused by natural fracture or improper operation, so as to standardize drilling operation and reduce downhole accidents.

7 Conclusion

This study has presented a method for quickly predicting the locations of drilling leakage layers using AI technology and has established two models based on the GA-BP neural network and random forest algorithms. These models thus established were confirmed by drilling data gathered from various blocks in southwestern China; the results are compared as follows:

1. The two models showed excellent performance in predicting the locations of drilling leakage layers. Regarding accuracy, both the GA-BP neural network and random forest models performed well in mining potential information from drilling data and, based upon that, predicting the parameters of the drilling leakage layers.
2. Through variance analysis, the location parameters of circulation loss were related to drilling fluid parameters, drilling geological parameters, and drilling mechanical parameters, among which five parameters (*i.e.*, drilling fluid density, horizon, lithology, riser pressure, and WOB) had the greatest impact on the location of circulation loss.
3. Based on field application, the framework proposed in this study is expected to be placed in practice and provide the most effective solution for leakage events in other field examples. However, these models are only valid for other datasets within the range of datasets used in the training process.

References

1 Feng Y., Arlanoglu C., Podnos E., Becker E., Gray K.E. (2015) Finite-element studies of hoop-stress enhancement for wellbore strengthening, *SPE Drill. Complet.* **30**, 38–51.

2 Xu C.Y., Kang Y., Chen F., You Z. (2017) Analytical model of plugging zone strength for drill-in fluid loss control and formation damage prevention in fractured tight reservoir, *J. Pet. Sci. Eng.* **149**, 686–700.

3 Zhang J.B., Wang Z.Y., Liu S., Zhang W.G. (2019) Prediction of hydrate deposition in pipelines to improve gas transportation efficiency and safety. *Appl. Energy* **253**, 1.

4 Zhang L.S., Bian Y.H., Zhang S.Y., Yan Y.F. (2019) A new analytical model to evaluate uncertainty of wellbore collapse pressure based on advantageous synergies of different strength criteria, *Rock Mech. Rock Eng.* **52**, 2649–2664.

5 Rojas J.C., Bern P.A., Ftizgerald B.L., Modi S., Bezant P.N. (1998) Minimizing down hole mud losses, in: *Paper No. IADC/ SPE 39398, IADC/SPE Drilling Conference, Dallas, TX, March 3–6*, p. 7.

6 Zhang L., Wang Z., Du K., Xiao B., Chen W. (2020) A new analytical model of wellbore strengthening for fracture network loss of drilling fluid considering fracture roughness, *J. Nat. Gas Sci. Eng.* **77**, https://doi.org/10.1016/j.jngse.2019.103093.

7 Vipulanandan C., Mohammed A. (2020) Effect of drilling mud bentonite contents on the fluid loss and filter cake formation on a field clay soil formation compared to the API fluid loss method and characterized using Vipulanandan models, *J. Petrol. Sci. Eng.* **189**. https://doi.org/10.1016/j.petrol.2020.107029.

8 Majidi R., Miska S.Z., Yu M., Thompson L.G. (2008a) Quantitative analysis of mud losses in naturally fractured reservoirs: the effect of rheology, in: *SPE 114130 presented at the SPE Western Regional and Pacific Section AAPG Joint Meeting, Bakersfield, 31 March–2 April.*

9 Majidi R., Miska S.Z., Yu M., Thompson L.G., Zhang J. (2008b) Modeling of drilling fluid losses in naturally fractured formations. Paper No. SPE 114630, in: *SPE Annual Technical Conference and Exhibition, Denver, CO, September 21–24*, p. 11.

10 Albattat R., Hoteit H. (2019) Modeling yield-power-law drilling fluid loss in fractured formation, *J. Petrol. Sci. Eng.* **182**.

11 Xu C., Kang Y., Chen F., You Z. (2016) Fracture plugging optimization for drill-in fluid loss control and formation damage prevention in fractured tight reservoir, *J. Nat. Gas Sci. Eng.* **35**.

12 Bjorndalen H.N., Jossy W.E., Alvarez J.M., Kuru E. (2014) A laboratory investigation of the factors controlling the filtration loss when drilling with Colloidal Gas Aphron (CGA) fluids. *J. Pet. Sci. Eng.* **117**.

13 Dias F.T.G., Souza R.R., Lucas E.F. (2015) Influence of modified starches composition on their performance as fluid loss additives in invert-emulsion drilling fluids, *Fuel* **140**.

14 Zukui L., Xinxu Z., Jingai Z., Daixu T., Yingsong Y., Jing Y., Yanqing L., Shengli Petroleum Administration Bureau, Dongying, Shandong, PR China (2001) The experiment investigation of the correlation of acoustic logging and rock mechanical and engineering characteristics, in: *Chinese Society for Rock Mechanics and Engineering. Frontiers of Rock Mechanics and Sustainable Development in the 21st Century Proceedings of the 2001 ISRM International Symposium Asian Rock Mechanics Symposium (ISRM 2001-2nd ARMS). Chinese Society for Rock Mechanics and Engineering*, pp. 105–107.

15 Chen P., Gupta P., Dudukovic M.P., Toseland B.A. (2006) Hydrodynamics of slurry bubble column during dimethyl ether (DME) synthesis: Gas–liquid recirculation model and radioactive tracer studies, *Chem. Eng. Sci.* **61**, 19, 6553–6570.

16 LeCun Y., Bengio Y., Hinton G. (2015) Deep learning, *Nature* **521**, 436–444.

17 Ghahramani Z. (2015) Probabilistic machine learning and artificial intelligence, *Nature* **521**, 452–459.

18 Littman M.L. (2015) Reinforcement learning improves behavior from evaluative feedback, *Nature* **521**, 445–451.

19 Shelley B., Grieser B., Johnson B.J., Fielder E.O., Heinze J.R., Werline J.R. (2008) Data analysis of Barnett shale completions, *SPE J.* **13**, 366–374.

20 Awoleke O.O., Lane R.H. (2011) Analysis of data from the Barnett shale using conventional statistical and virtual intelligence techniques, *SPE Reserv. Eval. Eng.* **14**, 544–556.

21 Shaheen M., Shahbaz M., Rehman Z., Guergachi A. (2011) Data mining applications in hydrocarbon exploration, *Artif. Intell. Rev.* **35**, 1–18.

22 Ma Z., Leung J.Y., Zanon S., Dzurman P. (2015) Practical implementation of knowledge-based approaches for steam-assisted gravity drainage production analysis, *Expert Syst. Appl.* **42**, 7326–7343.

23 Wang S., Chen S. (2019) Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling, *J. Petrol. Sci. Eng.* **174**, 682–695.

24 Moazzeni A., Nabaei M., Jegarluei S.G. (2012) Decision making for reduction of nonproductive time through an integrated lost circulation prediction, *Petrol. Sci. Technol.* **30**, 20, 2097–2107.

25 Moazzeni A., Ali Haffar M. (2015) Artificial intelligence for lithology identification through real-time drilling data, *Earth Sci. Clim. Change* **6**, 3, 265.

26 Ahmadi M.A., Shadizadeh S.R., Shah K., Bahadori A. (2018) An accurate model to predict drilling fluid density at wellbore conditions, *Egyptian J. Petrol.* **27**, 1, 1–10.

27 Kato K., Sakawa M., Ishimaru K., Ushiro S., Shibano T. (2019) Heat load prediction through recurrent neural network in district heating and cooling systems, in: *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC), Singapore, 12–15 October 2008*, IEEE, Piscataway, NJ, USA, pp. 1401–1406. ISBN 978-1-4244-2383-5.

28 Izadyar N., Ong H.C., Shamshirband S., Ghadamian H., Tong C.W. (2015) Intelligent forecasting of residential heating demand for the district heating system based on the monthly overall natural gas consumption, *Energy Build.* **104**, 208–214.