Journal of Petroleum Science and Engineering 177 (2019) 236–249

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: A case study from Marun oil field

Mohammad Sabah[a,*], Mohsen Talebkeikhah[a], Farough Agin[a], Farzaneh Talebkeikhah[b], Erfan Hasheminasab[a]

[a] Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
[b] Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Lost circulation
Marun oil field
ANFIS
Artificial neural network
Decision tree

ABSTRACT

One of the most prevalent problems in drilling industry is lost circulation which causes intense increase in drilling expenditure as well as operational obstacles such as well instability and blowout. The aim of this research is to develop smart systems for estimating amount of lost circulation making able to use appropriate prevention and remediation methods. To obtain this aim, a large data set were collected from 61 recently drilled wells in Marun oil field in Iran to be used for developing relevant models. After that, using the extracted data set consisting of 1900 data subset, intelligent prediction models including decision tree (DT), adaptive neuro-fuzzy inference systems (ANFIS), artificial neural networks (ANN) and also hybrid artificial neural network namely genetic algorithm-multi-layer perception (GA-MLP) were developed to make a quantitative prediction on lost circulation. The model outputs are then analyzed by various performance indices such as variance accounted for (VAF), root mean square error (RMSE), performance index (PI) and coefficient of determination (R^2). Eventually, it is found that developed models are highly applicable in lost circulation prediction. Concluding remark is that DT model having determination coefficient of 0.9355 and RMSE of 0.091 is superior comparing to other developed models and hybrid ANN (GA-MLP) exhibits lowest prediction performance among other implemented models.

1. Introduction

Lost circulation is described simply as a portion or whole of the drilling fluid (solids and particles) that is entered into the formations (Messenger JU, 1981). It happens when either the permeability of the formation is high enough, or the voids are too big that they cannot be blocked by solid particles existing in the drilling mud. The amount of lost circulation may be significantly low or high and this drilling problem may occur at any depth. Historically, the lost circulation severity has been categorized based on the amount of the drilling mud that is entered into the formation (seepage, partial and total) (Ghalambor et al., 2014). Severity of lost circulation depends on several variables such as formation pore pressure, formation fracture pressure, rheological properties of drilling fluid, size of fractures and pores in the formation, and some drilling parameters (Moazzeni et al., 2010). Lost circulation can also be categorized into four main possibilities respecting the type of formation listed as: natural/artificial fractured reservoirs, cavernous formations, formations including high permeable zones and formations with low level of consolidation. In all of these possibilities lost circulation cannot be avoided (Fidan et al., 2004).

Lost circulation has been a huge challenge since the beginning of drilling industry (Wang et al., 2008). This phenomenon can cause various problems leading to an increase in drilling non-productive time as well as drilling costs (Whitfill et al., 2007). Some of these problems include stuck pipe, blowout, well abandonment, formation damage and incomplete zonal isolation (Pilehvari and Nyshadham, 2002; Nasiri et al., 2017). Consequently, many oil companies spend millions of dollars annually to deal with the lost circulation (Murchison, 2006).

The most widely used methods to deal with this problem are adding lost circulation material (LCM) into drilling fluid, lowering the mud weight or in serious cases, zonal cementation of problematic location. However, society of drilling industry is already concerning about lack of methods capable to predict the intensity of unfavorable matters before happening. It is important to note that, developing a system being

* Corresponding author.
E-mail address: mohamadsabah@aut.ac.ir (M. Sabah).

https://doi.org/10.1016/j.petrol.2019.02.045
Received 25 February 2018; Received in revised form 13 November 2018; Accepted 14 February 2019
Available online 19 February 2019
0920-4105/

able to predict lost circulation severity before drilling a certain formation can help drilling industry to employ appropriate methods and lost circulation materials to stop or reduce this matter on time. Nowadays, the significant computation and technology progress has introduced newer solutions such as ANN, fuzzy inference systems (FIS), and their hybrids which have led to such a robust performance in making prediction over the desired parameters. Capability of tolerating a wide range of uncertainties in parameters is the most important advantage of these approaches (Yilmaz and Kaynar, 2011).

The objective of this study is to develop intelligent systems with capability of predicting the severity of lost circulation in acceptable accuracy before drilling a certain formation. For this purpose, DT, ANFIS, ANNs, and a hybrid ANN are developed to predict the lost circulation severity in Marun oil field located in southwest of Iran. A large data set are collected from daily drilling reports (DDR) of 61 recently drilled wells and the most influential parameters on the severity of lost circulation are selected to train the proposed machine-learning algorithms. Eventually, prediction performance of developed models are compared to select the most reliable model in terms of lost circulation prediction.

2. Field information and data analysis

Marun field is one of the largest oil fields in Iran. It was discovered in 1963 and is located onshore in the west of Iran. Length and width of the field is about 67 and 7 km, respectively. A scheme of this field is illustrated in Fig. 1. Due to great magnitude of the field, it is divided into 8 sections as shown in Fig. 2. Key Formations drilled to penetrate the reservoir of Marun oil field are Aghajery, Mishan, Gachsaran (GS7, GS6, GS5, GS4, GS3, GS2, Cap rock) and Asmari. Asmari formation is one of the primary reservoirs in this field. There are severe lost circulations during drilling the Asmari formation in comparison with other formations and a lot of money is spent annually to deal with mud loss in this formation.

The required data to predict the lost circulation in Marun field are extracted from daily drilling reports of 61 recently drilled wells. The parameters that are considered as inputs are depth, northing, easting, hole size, weight on bit, pump rate, pump pressure, viscosity, shear stresses at shear rates of 600 and 300 rpm (θ_{600}, θ_{300}), gel strength, drilling time, drilling meterage, solid percent obtained from retort test, bit rotational speed, formation type, pore pressure, drilling mud pressure, and formation fracture pressure. Table 1 shows statistical details of these parameters. The amount of the lost circulation is considered as output. The overall data set consists of 1900 records divided randomly into two parts. 70% (1330 data sets) of the overall data are used to construct the models and the remaining 30% (570 data sets) are used to test the developed models. After gathering and selecting the required data, they are normalized to enhance accuracy of the systems. The

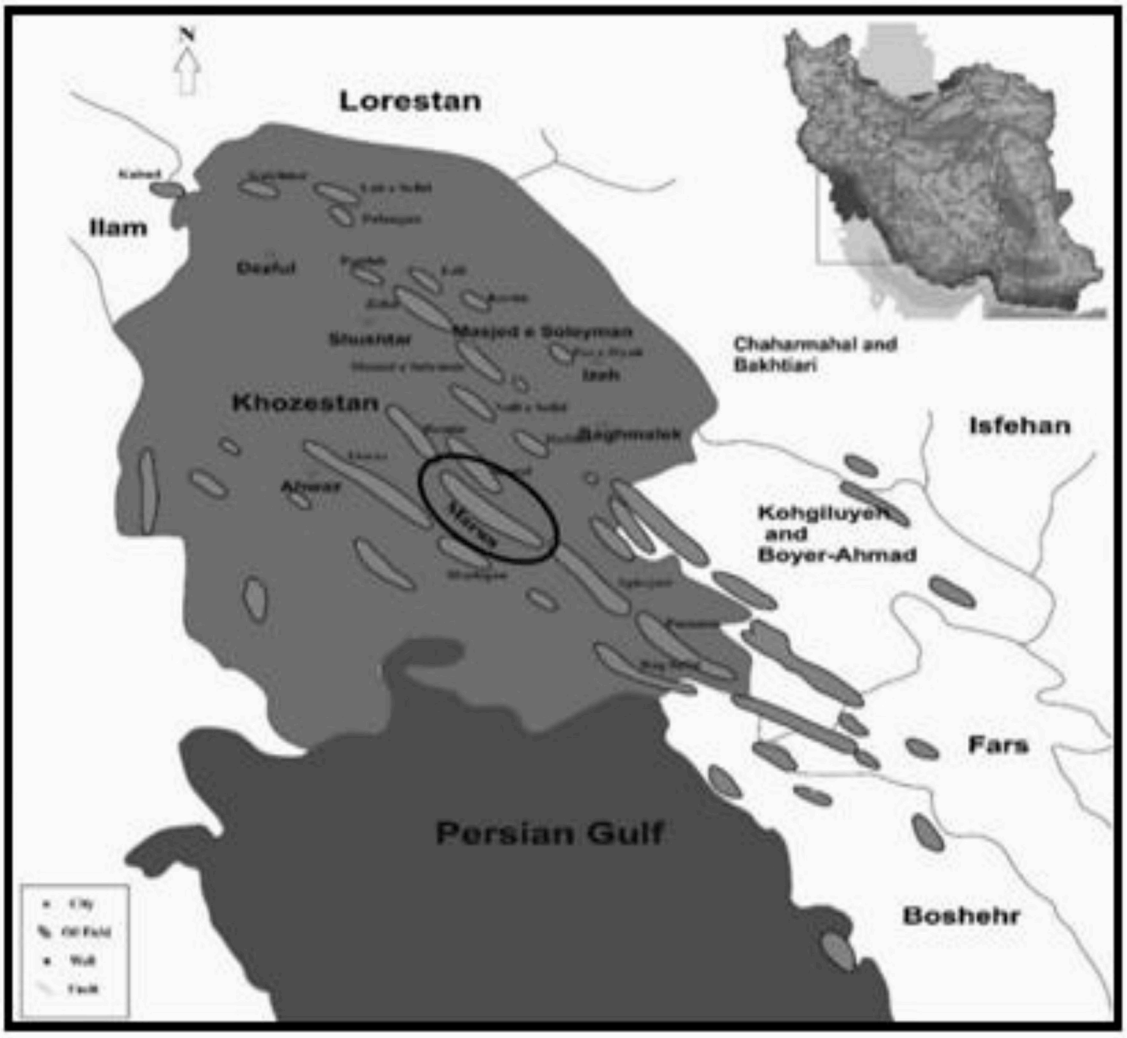

Fig. 1. Location of Marun oil field.

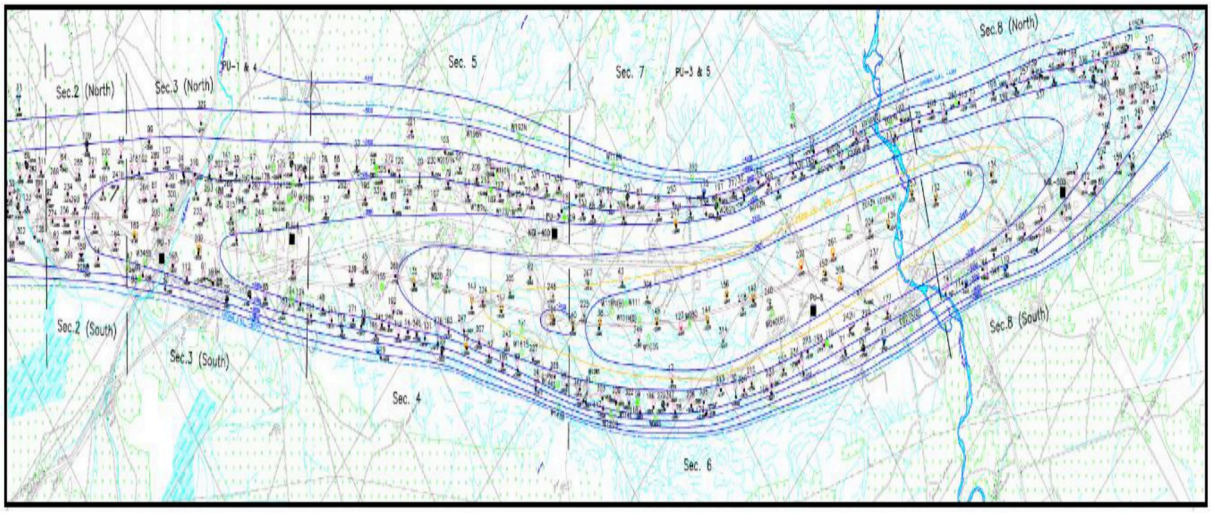

Fig. 2. Different sections of Marun oil field (3).

Table 1
Data used as inputs and output.

#	Parameter	Unit	Minimum	Maximum	Average
1	Drilling Meterage	m	0.32	650	47.27
2	Northing	m	1887146	1936065	1907587
3	Easting	m	1005840	1049982	1026675
4	Hole Size	m	0.103	0.65	0.3
5	Weight On Bit	kg	1000	70000	20650
6	Pump Rate	m^3/sec	0.005	0.063	0.033
7	Pump Pressure	MPa	0.34	20.06	13.34
8	Viscosity	cp	27	100	47.33
9	θ_{600}	lb/(100 ft)^2	3	293	79.84
10	θ_{300}	lb/(100 ft)^2	2	163	46.87
11	Gel Strength	lb/(100 ft)^2	1	49	5.54
12	Drilling Time	sec	360	86400	59148
13	Depth	m	5.1	1698.6	855.21
14	Solid Percent	%	0	61	23.2
15	Bit rotational speed	RPM	50	250	137.31
16	Formation Type	–	1	15	5.94
17	Pore Pressure	MPa	0.05	23.1	11.34
18	Mud Pressure	MPa	0.052	33.47	12.79
19	Fracture Pressure	MPa	0.078	30.41	16.59
20	Severity of lost circulation	bbl/hr	0	600	80.90

following equation is used for normalization of the input and output data between -1 and 1 (Deosarkar and Sathe, 2012):

$$x_i^n = 2 \times \frac{x_i - x_{min}}{x_{max} - x_{min}} - 1 \quad (1)$$

In above equation, i is number of parameters, x_{max} and x_{min} represent maximum and minimum values of x_i, respectively.

3. Decision tree

Classification and regression tree (CART) technique belongs to a class of non-parametric supervised learning tools implemented in a wide range of fields of occupation and industries, proved the effectiveness in execution of a cause-and-effect analysis. DT methods in data mining are essentially divided in two categories: "The classification tree" and "The regression tree". The type of tree is decided from the output variable type whether naturally categorical, which then the classification tree is formed; or continuous that is sent into regression tree (Breiman, 2017). In this study, since the output variable is the amount of lost circulation, which is considered a continuous variable, a regression tree analysis is used.

The underlining procedure for any of these two DTs adheres to the same instruction. DTs come out of separating observations into subgroups by creating splits on predictors (Maucec et al., 2015). This process is also known as binary recursive partitioning following a binary splitting process in which parent nodes are descended into two child nodes and subsequently the rows of the tree will go down until the eventual one which are terminal nodes with no further splitting (Singh, 2017). The beginning of the DT is with finding splitting criteria on the basis of input variables and minimizing the square error between the observed and calculated value of output variable. Then it produces one root node and two child nodes (Singh, 2015). The similar pattern is applied for each child node repeatedly in order to further splitting. At last, the DT generates a logical sequence of splitting criteria based on input variable and a DT diagram to provoke a scheme of the procedure. Developed regression tree for prediction of lost circulation is shown in details in appendix.

4. Artificial neural network

ANN is similar to the ways that the brain uses for learning. They are inspired from the brain structure (Okpo et al., 2016) and are the non-linear mathematical models which are highly considered because of their simplicity, flexibility and availability (Khoshjavan et al., 2010; Al-Naser et al., 2016). ANNs can model complex non-linear systems through the establishment of relationships between input data and output data without using a mathematical model for the system (Soofastaei et al., 2016). The two principle components of them are neurons (node, processing elements) and interconnections (weights). The role of the former is to process information, whereas the latter

makes connections among different neurons (Mohaghegh, 2000). A pair of remarkable neural networks are presented named as multilayer perceptron (MLP) and radial basis function (RBF) networks. The main difference of the two aforementioned networks is the way through which the nodes process the information.

In oil industry, ANN method has been used in a variety of different sections. Characterizing reservoir-related properties (Ashoori et al., 2010; Torabi et al., 2011; Derakhshanfard and Mehralizadeh, 2018; Abedini and Abedini, 2011; Abedini and Torabi, 2013), determining drilling parameters (Zhang et al., 2018; Elkatatny et al., 2016), defining lithological properties and analyzing source rock characteristics (Martin et al., 2017; Bolandi et al., 2017), and enhancing oil recovery and production (Le Van and Chon, 2017; Xue et al., 2014; Shan et al., 2018) are examples of ANN implementations.

4.1. Multi-layer perceptron neural network (MLP-NN)

An MLP network consist of an input layer, one or more hidden layer, and one output layer. Each layer has several processing neurons and each neuron is fully connected to subsequent layers by weighted interconnections. The number of input parameters is equal to the number of neurons in input layer and for output layer one neuron is used corresponding to the output of the model. The relationship between inputs and outputs of the model are appointed in hidden layer(s) (Lashkarbolooki et al., 2012). Both the number of hidden layers and the number of neurons in them significantly influence the efficiency of a MLP network (Hemmati-Sarapardeh et al., 2016). Value of each node in hidden layers/output layer is calculated through multiplying the weight of each node from previous layer to the desired node in hidden layers/output layer and these values are summed. Following this, a bias value is added to the obtained results and the calculated value is passed to the activation level through a transfer function to produce the output. There are different activation functions such as identity function, binary step function, binary sigmoid, bipolar sigmoid, Gaussian, and linear functions which can be used in hidden and output layers (Fausett, 1994). The output of model can be expressed as follows:

$$y_k = F_k\left(\sum_{i=1}^{m} w_{kj}x_j + b_k\right) \quad (2)$$

Where y_k is output, w_{kj} are the link weights, x_j is input, b_k is bias vector, and F_k represents the activation transfer function. Training process of MLP neural network is performed through a back propagation algorithm such as Levenberg-marquardt (LM), gradient descent (GD), scaled conjugate gradient (SCG), and resilient back propagation (RP) (Ghoreishi and Heidari, 2013; Heidari et al., 2016).

In this study, among the all training algorithms, LM algorithm is selected as a training function to employ for model. In the literature, LM algorithm has been implemented by many scholars in approximating geotechnical engineering problems (Meulenkamp and Grima, 1999; Armaghani et al., 2017; Ceryan et al., 2013). As highlighted by several researchers, an MLP-NN with two or more hidden layers can provide greater accuracy in extremely complex problems (Pham et al., 2017; Surkan and Singleton, 1990; Openshaw and Openshaw, 1997; Li and Yeh, 2002). Therefore, two hidden layers were considered in MLP-NN to provide sufficient accuracy. Tangent sigmoid and Purelin as most widely used transfer functions (Yilmaz and Kaynar, 2011) are implemented in hidden layers and output layer, respectively. The number of neurons in input and output layer were taken into account according to the input and output variables. As stated before, the network

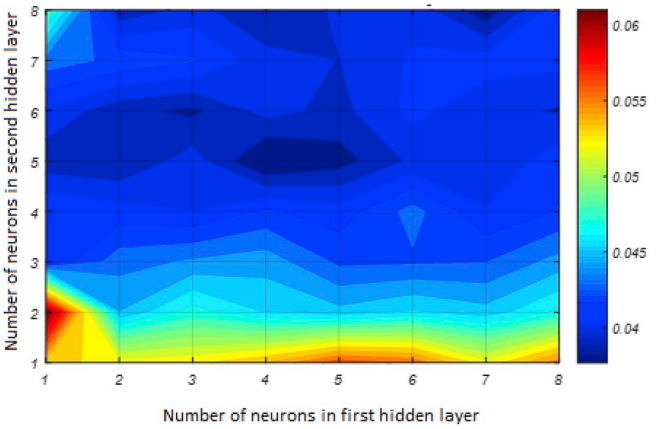

Fig. 3. Number of neurons in hidden layers versus MSE of testing data subset.

performance relies heavily on the number of the neurons in their hidden layer(s). Here, a trial and error approach was used to determine number of neurons in both hidden layers. The number of neurons in each hidden layer were changed and then efficiency of the different MLP-NN structures were analyzed based on mean square error (MSE) of testing data subset. MSE value of testing data subset versus number of neurons in the both hidden layers are shown in Fig. 3. According to the figure, the optimum case has 4 neurons in the first hidden layer and 5 neurons in the second hidden layer as this kind of network yields minimum MSE for the testing data subset.

4.2. Radial basis function neural network (RBF-NN)

RBFs emerged as a remedy in the late 1980s. Feasibility of this type of neural network in treating arbitrary scattered data, easily generalizing to several dimensional space and providing spectral accuracy has made it particularly a popular alternative to MLP (Elsharkawy, 1998; Lashkenari et al., 2013). In addition, RBF neural networks surpass MLP models for the distinguished accuracy in modeling non-linear data and the feasibility of being trained in a single direct procedure as opposed to an iterative solution in MLP (Venkatesan and Anitha, 2006). Structure of RBF is similar to the MLP. The main difference between these neural networks is that RBF has only one hidden layer consisting of a number of nodes called RBF units. Architecture of RBF-NN is as a two-layer

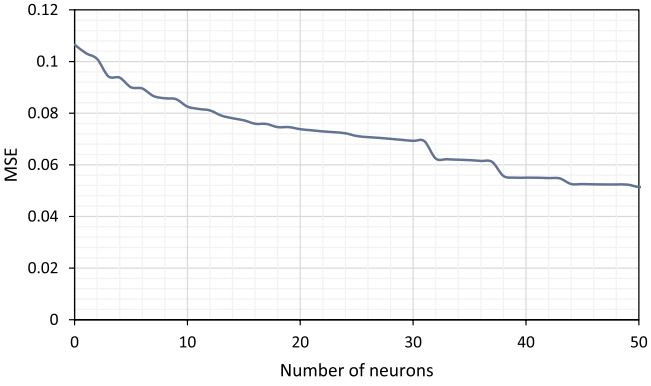

Fig. 4. MSE of the lost circulation prediction achieved by the RBF network versus the number of neurons in the hidden layer.

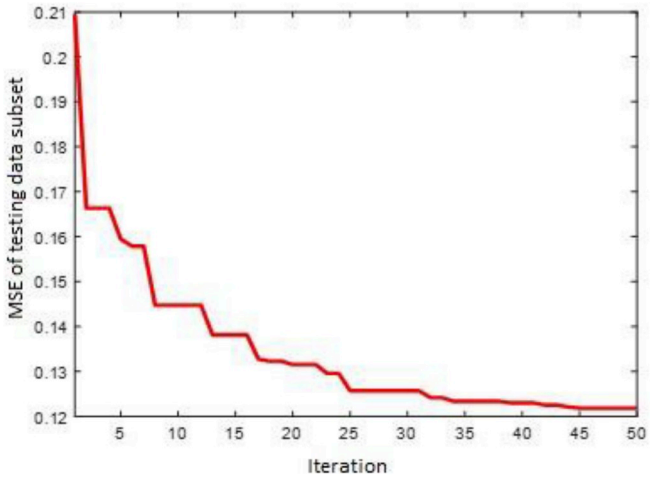

Fig. 5. Convergence of GA-MLP model to the optimum solution.

feed-forward neural network in which the inputs are transferred to the output layer by the neurons at the hidden layer (Wu et al., 2012; Park and Sandberg, 1991). Every RBF network has two important parameters describing center location of functions and its deviation. Finding unit centers and determining optimum values of weights connecting RBF units to the output units are two main steps in training process of RBF neural networks. Different methods such as random selection of centers, clustering, and density estimation can be used to find the centers in RBF networks (Orr, 1996). The system outputs can be shown as Eq. (3):

$$f(x_i) = w^T \varphi(x_i) \tag{3}$$

Here w^T is the transposed output of layer vector and $\varphi(x_i)$ is kernel function, which is typically in Gaussian form (Haykin and Network, 2004). In this study, a clustering method based on K-means algorithms was used for finding unit centers and gradient descent back propagation algorithm was selected as training function of RBF-NN. Number of neurons in hidden layer is a substantially determinant tuning parameter in developing these neural network structures. However, different optimization algorithms can be used for this purpose (Najafi-Marghmaleki et al., 2017), here a trial and error approach was employed to determine the optimum value of this parameter. Different structure of RBF-NN were developed by changing this parameter and the performance of each RBF-NN was monitored based on MSE value of testing data subset. Fig. 4 shows the behavior of this approach by considering MSE values responses versus number of neurons. According to this figure, the optimum value is 50 for number of neurons in the hidden layer.

Table 2
Architecture of MLP neural network trained with genetic algorithm.

Network type	Hybridized Neural Network
Training function	Genetic Algorithm
Number of layers	2
Nodes in hidden layer	10
Transfer function of hidden layer	TANSIG
Neurons in output layer	1
Transfer function of output layer	PURELIN
Performance objective function	MSE

Table 3
Properties of genetic algorithm.

Population Size	150
Max Generation	50
Selection Mode	Random
Recommendation Percent	20
Cross over Percent	50
Mutation Percent	30

4.3. Genetic algorithm-multilayer perception neural network (GA-MLP)

Weights and biases are two important parameters of MLP neural networks which need to be optimized through an optimization algorithm. Here, these tuning parameters are optimized using genetic algorithm in an attempt to improve lost circulation prediction accuracy compared with the MLP-NN trained by the LM back-propagation algorithm. The GA was employed due to its capability in solving complex and nonlinear optimization problems. The first step in applying GA is generating an initial population of feasible solutions. Each member of generated population is indicative of a solution for the desired problem. Fitness of these members is determined by a function called objective function (fitness function). Following this, all members are ranked based on their MSE between measured and predicted values. There are specific genetic operators including cross over, reproduction, and mutation that are used to keep the balance between exploration areas and information acquired from exploitation areas. In other words, the fittest members of initial population are subjected to aforementioned operators to form new population and guide the search to the area including global minimum (Ab Wahab et al., 2015). Fig. 5 displays the convergence of tuning parameters to the expected optimum condition. The optimum structure of GA-MLP neural network and optimized values of GA are listed in Table 2 and Table 3, respectively.

5. Adaptive neuro-fuzzy inference system (ANFIS)

This Fuzzy Logic (FL) based mean was first afforded by Zadeh (1988). ANFIS model is capable of solving non-linear problems and modeling physics by taking advantage of qualitative methods in the place of operating quantitative approaches (Khosravanian et al., 2016) through transforming input data into specific terms known as linguistic terms or fuzzy sets (Sugeno and Kang, 1988). There are five layers in the structure of neuro-fuzzy system. These layers are Fuzzification, rules, normalization, defuzzification and output (Zahmatkesh et al., 2017).

In the first layer, input data are converted to fuzzy input by defining certain functions named membership functions (MFs) for each input variable. The calculated values of membership degrees for each input parameter are multiplied and the resultant is called firing strength and it is presented as follows:

$$w_i = \prod_{i=1}^{m} \mu_{ij}(x_j) \tag{4}$$

w_i is the firing strength, μ_{ij} is the membership degree of j-th MF for i-th input parameter and m is the number of input parameters. In this layer, the firing strength of each rule is calculated via multiplication and a

Table 4
Parameters and their corresponding values used for developing the ANFIS model.

ANFIS parameters	Value
Number of MFs	28
Radius of influence	0.5
Squash factor	0.5
Accept ratio	0.15
And method	Prod
Or method	Probor
Implication	Min
Aggregation	Max
Defuzzification	Wtaver

rule which contains the highest firing strength matches the input data. Next layer calculates the ratio of firing strength of each rule to sum of all firing strength of rules as following:

$$\overline{w_i} = \frac{w_i}{\sum_i w_i} \tag{5}$$

$\overline{w_i}$ Represents normalized firing strength. Final output of ANFIS model is calculated by multiplying the normalized firing strength with first order Takagi-Sugeno-Kang (TSK–FIS):

$$\overline{w_i} \times f_i = \overline{w_i} \times \left(\sum_{j=1}^{m} n_{ij} x_j + r_{ij} \right) \tag{6}$$

In above equation, f_i can be a polynomial function or a constant number. The n_{ij} and r_{ij} values are tuning parameters of TSK-FIS. The

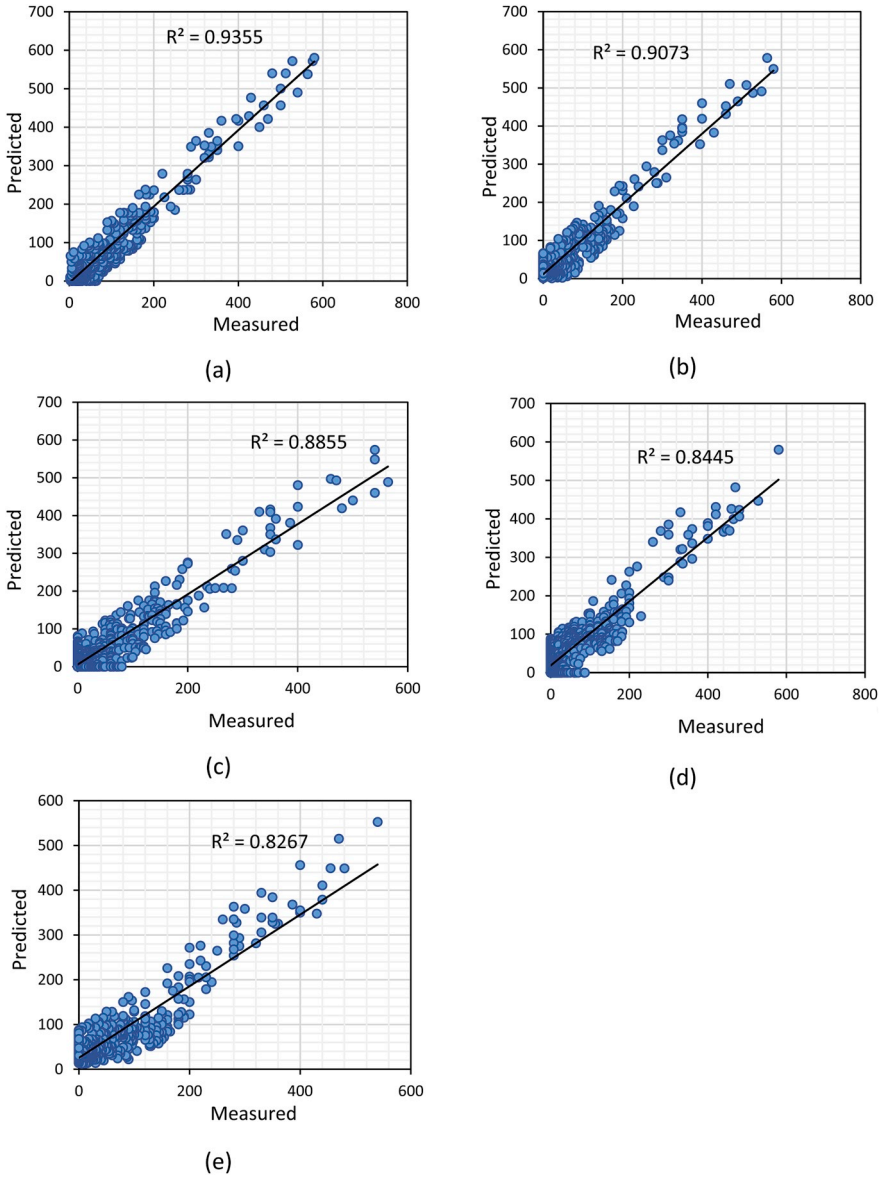

Fig. 6. The cross plot for (a) DT, (b) MLP, (c) ANFIS, (d) RBF and (e) GA-MLP.

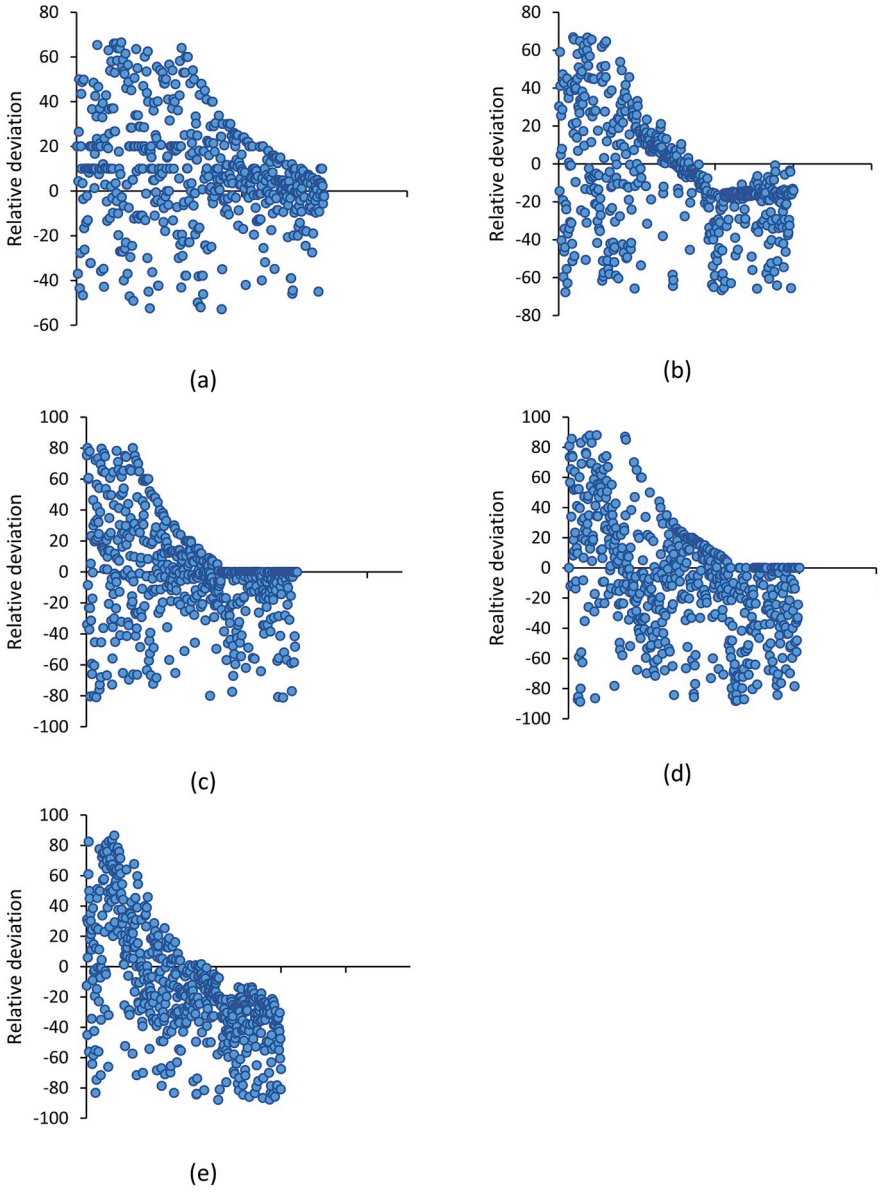

Fig. 7. Relative deviations of developed models: (a) DT, (b) MLP, (c) ANFIS, (d) RBF, and (e) GA-MLP.

optimum values of these parameters should be determined by an optimization algorithm to achieve better prediction performance (Jang, 1993). The last layer calculates sum of the output of all nodes from the previous layer to produce overall ANFIS output as below:

$$overall\ output = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (7)$$

In this study, Gaussmf membership function was selected as a membership function to fuzzify predictor variables, which consists of two constants (σ and c) and the rule base was adapted from normalized training data set. For developing initial FIS, different methods such as grid partitioning, subtractive clustering, and fuzzy c-means clustering are provided in MATLAB software (Salehinia et al., 2016). In this study, subtractive clustering method and hybrid rule algorithm were utilized to generate and train initial FIS, respectively. Parameters involved in the developing process of ANFIS model are presented in Table 4.

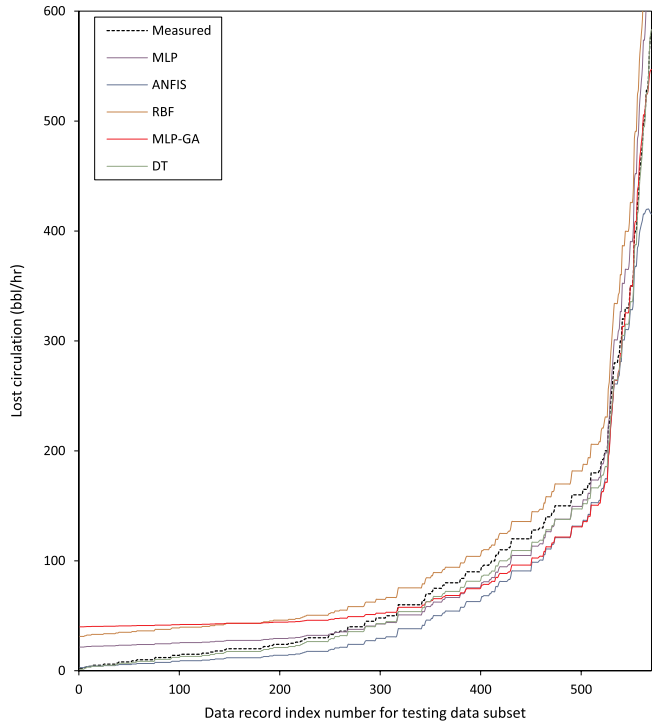

Fig. 8. Measured versus predicted ROP versus data record index numbers for five developed machine-learning models applied to the testing data subset: (a) DT, (b) MLP, (c) ANFIS, (d) RBF, and (e) GA-MLP.

Table 5
Performance indices.

Model	Data set	RMSE	R^2	VAF (%)	PI
Decision tree	Test	0.091	0.93	93.21	1.8
	Train	0.052	0.9725	97.23	1.9
MLP	Test	0.099	0.9073	90	1.75
	Train	0.094	0.9213	92.12	1.78
ANFIS	Test	0.1087	0.8855	88.32	1.71
	Train	0.1163	0.9018	90.12	1.73
RBF	Test	0.1315	0.8445	84.43	1.63
	Train	0.1172	0.85	85	1.65
GA-MLP	Test	0.137	0.826	82.57	1.59
	Train	0.132	0.8307	82.78	1.6

6. Results and discussions

In this section, the performance of implemented models is analyzed using various graphical and statistical methods to choose the most reliable model. Regression plots of predicted values versus measured data for developed models are shown in Fig. 6. In this figure, the vertical and horizontal axes are predicted and measured values, respectively. As it is obvious, the values of regression coefficients for all developed models are highly acceptable indicating that the proposed models are sufficiently accurate for prediction of lost circulation. However, this figure reveals that the DT model has the highest regression coefficient and best prediction performance than MLP, ANFIS, RBF and GA-MLP models. Moreover, the relative deviations of developed models are presented in Fig. 7. It is clear that the smallest deviations belongs to DT and MLP-NN models and highest deviation belongs to RBF and GA-MLP models. In other words, deviation of the predicted values by DT and MLP-NN from measured amounts are relatively less than other developed models. Fig. 8 presents a physical comparison between results of five developed machine-learning algorithms and measured lost circulation. It is noticeable that DT exhibits a very good prediction performance and its results are well matched with measured values of lost circulation. MLP neural network comes in second in terms of prediction performance and like DT model successfully models trends of changes in lost circulation measurement. The third and fourth most accurate predictive models are ANFIS and RBF-NN, respectively. Interestingly, GA-MLP model does not appropriately tracks trends of measured lost circulation data comparing with other implemented intelligent systems. There are also various statistical performance indices to assess the performance of mentioned models. In this study, VAF, RMSE, PI, and R^2 are used to compare prediction capability of developed models as employed by Yılmaz and Yuksek, 2008, Basarir et al., 2014 and Boyacioglu and Avci, 2010

$$RMSE = \left(\frac{1}{p}\sum_{r=1}^{p}(y_r - z_r)^2\right)^{\frac{1}{2}} \tag{8}$$

$$PI = \left(r + \frac{VAF}{100}\right) - RMSE \tag{9}$$

$$VAF = \left(1 - \frac{var(y_r - z_r)}{var(y_r)}\right) \times 100 \tag{10}$$

$$R^2 = 1 - \frac{\sum_{r=1}^{p}(y_r - z_r)^2}{\sum_{r=1}^{p}(y_r - y_{r,mean})^2} \tag{11}$$

where z is the model output, y is the actual output and p is the number of data in data set. The accuracy of the proposed models is distinctive according to the small values of RMSE and high values of R^2, PI and VAF. Table 5 indicates the RMSE, VAF, PI and R^2 values for the developed models. In order to have a better illustration, Fig. 9 shows graphical comparison of implemented models in terms of different performance indices. Since DT model has the highest value of R^2, VAF and PI and the lowest value of RMSE, it has the best performance among other models in lost circulation prediction. The second most accurate model is MLP-NN, which presents slightly lower prediction efficiency compared to the DT model. It indicates that both models are showing well prediction performance. The lowest prediction accuracy among the implemented models refers to GA-MLP.

As it is obvious from results, decision tree shows additional progress in prediction performance comparing to ANNs and ANFIS. This better

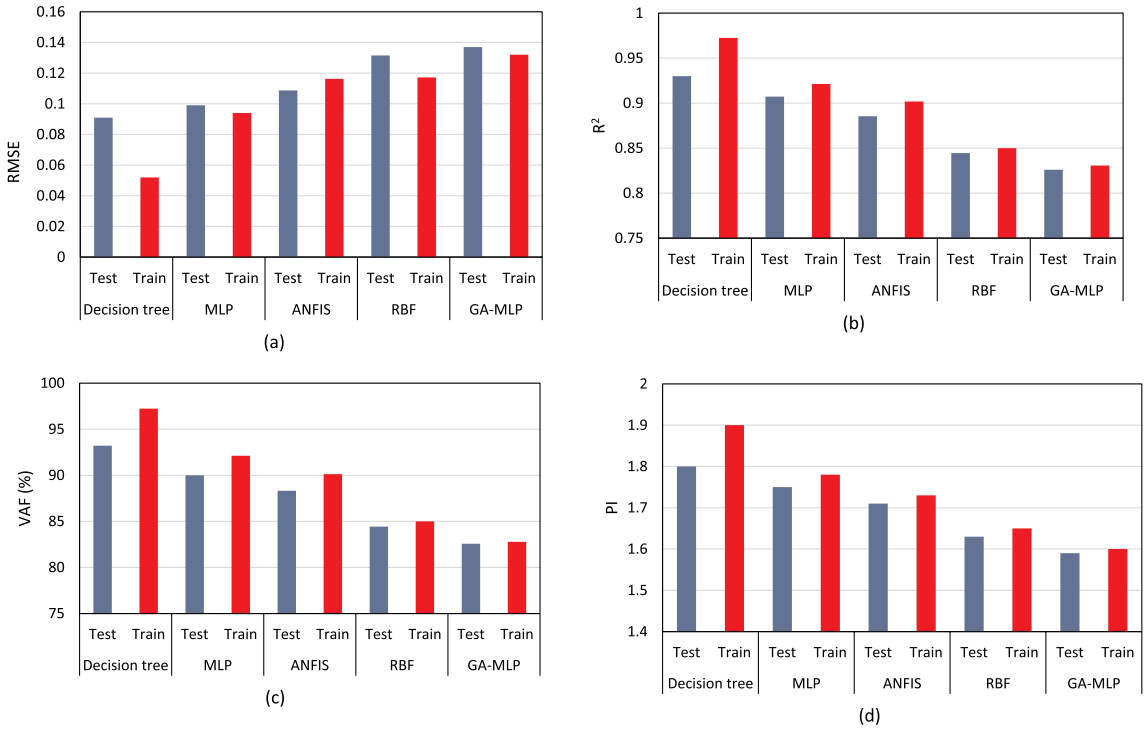

Fig. 9. Graphical comparison between (a) RMSE, (b) R^2, (c) VAF, and (d) PI value of implemented models.

performance shows itself especially when a large number of variables are used. For a small number of variables, ANNs and ANFIS can be competitive, but as the number of the variables increases, the DT results are more efficient and superior than ANN and ANFIS techniques (Roe et al., 2005). Regarding to the GA-MLP model results, using genetic algorithm to adjust weight and bias of MLP-NN failed in enhancing prediction performance of this type of neural network. This low prediction accuracy can be attributed to performance of genetic algorithm in searching global minimum. By using GA as a training function, the probability of convergence at local minimum increases which results in unsatisfied adjustments of weights and biases.

7. Conclusion

In this study, several inference systems including DT, MLP-NN, RBF-NN, ANFIS and GA-MLP were utilized to predict lost circulation. A large data set was extracted from drilling reports of Marun oil field. The overall data set with 1900 data points, were divided randomly into two parts. 70% of the overall data were used to construct the models and the remaining 30% were used to assess the prediction performance of developed models. The parameters that were considered as inputs were depth, northing, easting, hole size, weight on bit, pump rate, pump pressure, viscosity, shear stresses at shear rates of 600 and 300 rpm (θ_{600}, θ_{300}), gel strength, drilling time, drilling meterage, solid percent obtained from retort test, bit rotational speed, formation type, pore pressure, drilling mud pressure, and formation fracture pressure. Amount of lost circulation was considered as the demanding output. Statistical and graphical analysis showed that the prediction performance of developed models are sufficiently acceptable in prediction of lost circulation. However, DT is the most accurate method relative to other four developed models. This tremendous performance of DT in prediction manifests itself when a large number of variables need to be defined. For a small number of variables, ANNs and ANFIS lead to better results, but as the number of the variables increases, the DT model becomes more efficient and superior to ANN and ANFIS techniques. Furthermore, using GA as training function of MLP-NN cannot be as effective approach for enhancing prediction capability of these types of neural networks.

Appendix

In this study, a complex regression tree with 5 fold cross validation is used to fit an optimized regression tree for training data set and to achieve an accurate predictive model. Developed regression tree with statistical information for each node is shown in three sections (a, b, and c) in Fig. 10. Classification tree has one root node, 35 intermediate nodes, and 37 terminal nodes. The text box at each node provides statistical information for that node. The first split at root node is performed based on viscosity and splitting value is 35.5 cp. In this node, there are a total of 1867 observations with mean value of 78.51 and standard deviation of 151.57. The left hand side of the branch indicates that if viscosity becomes lower than aforementioned splitting value, splitting befalls depending on flow rate of drilling fluid. On the other side, if viscosity exceeds the 35.5 cp, the nodes splits further based on the bit rotational speed. This process continues based on different splitting criteria until reaching terminal nodes with no further splitting.

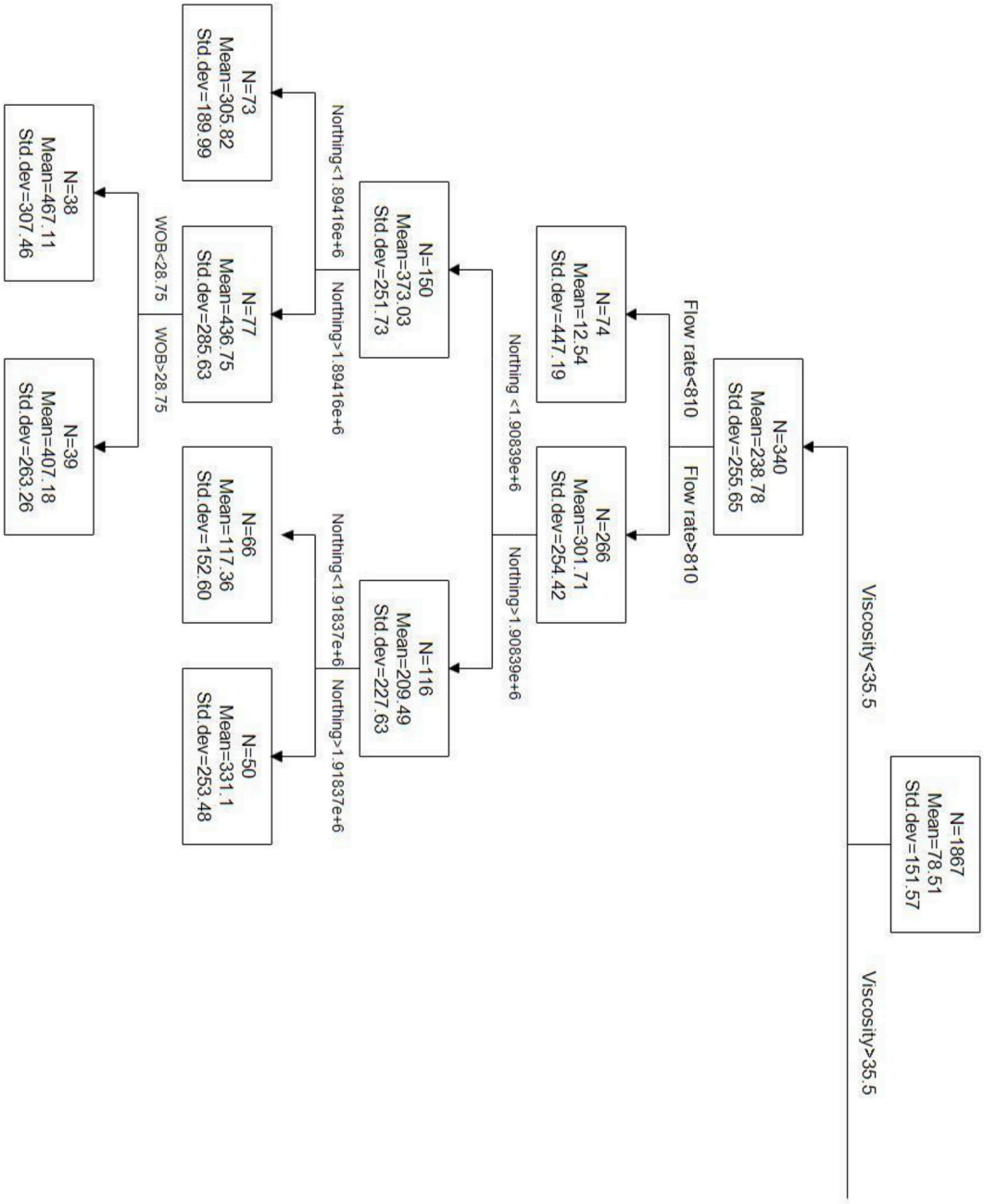

Fig. 10a. Optimal regression tree with statistical information for each node shown in the text box.

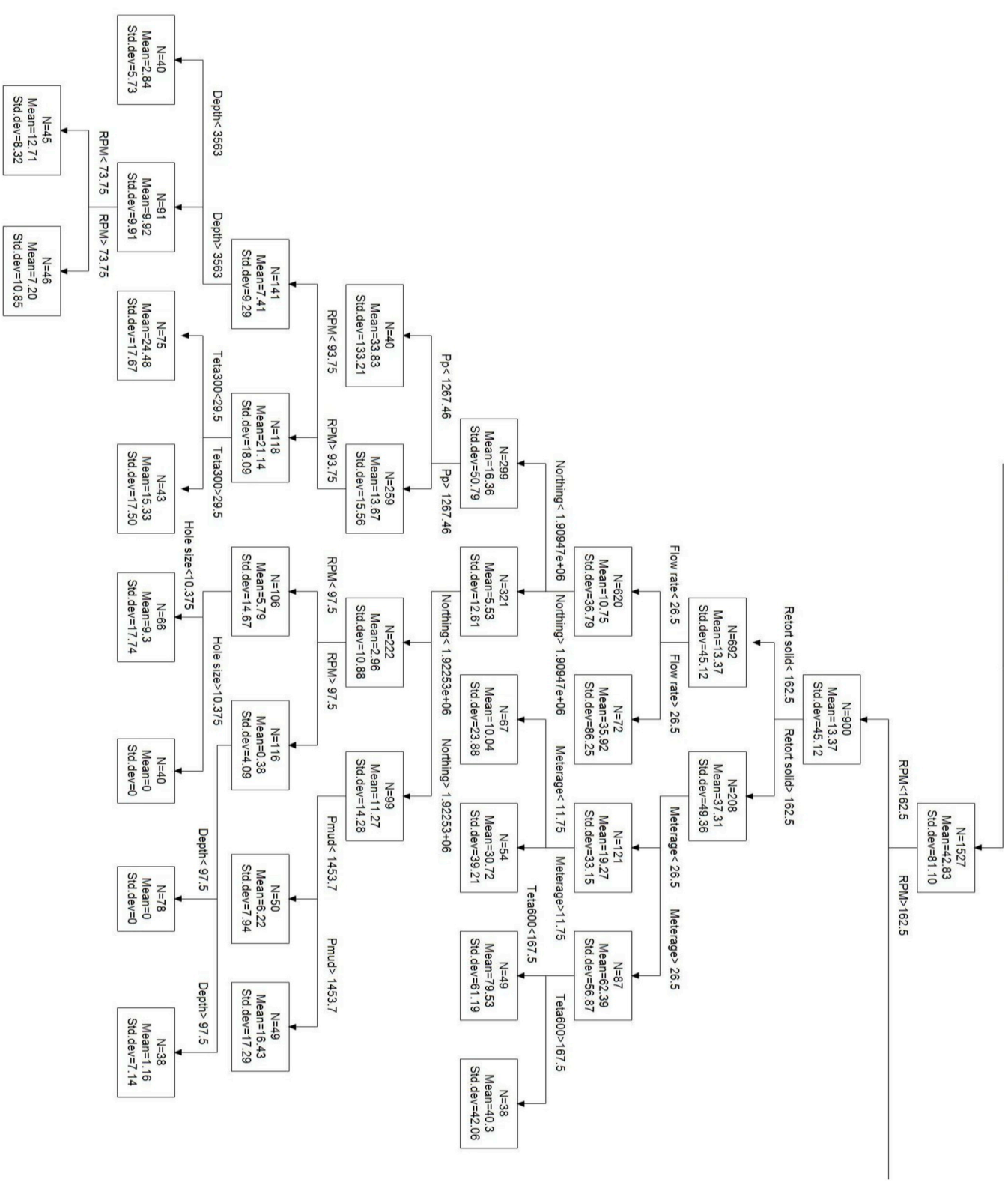

Fig. 10b. Optimal regression tree with statistical information for each node shown in the text box.

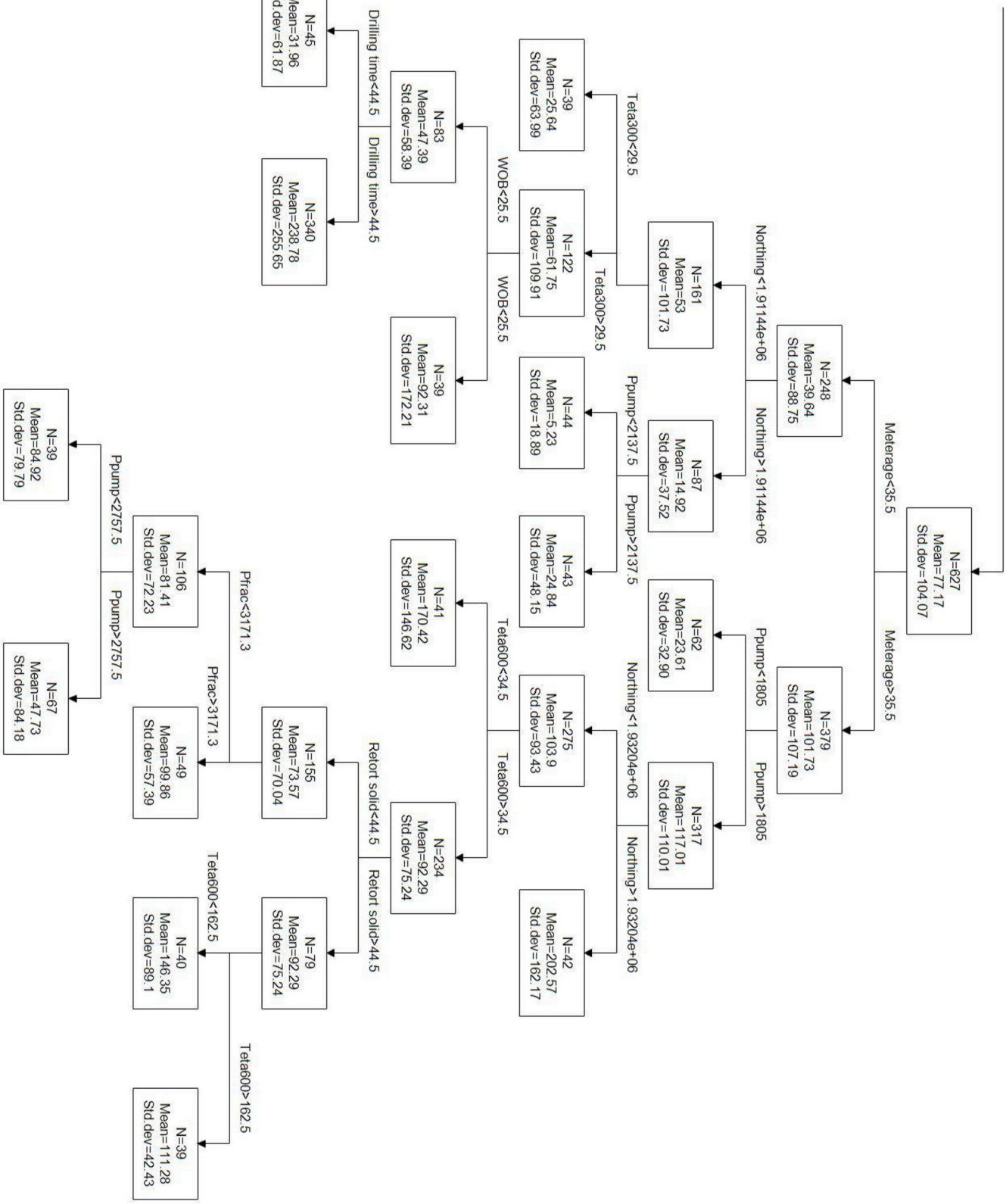

Fig. 10c. Optimal regression tree with statistical information for each node shown in the text box.

References

Ab Wahab, M.N., Nefti-Meziani, S., Atyabi, A., 2015. A comprehensive review of swarm optimization algorithms. PLoS One 10 (5), e0122827.

Abedini, R., Abedini, A., 2011. Development of an artificial neural network algorithm for the prediction of asphaltene precipitation. Petrol. Sci. Technol. 29 (15), 1565–1577.

Abedini, A., Torabi, F., 2013. Implementing artificial neural network for predicting capillary pressure in reservoir rocks. Spec. Top Rev. Porous Media Int. J. 4 (4).

Al-Naser, M., Elshafei, M., Al-Sarkhi, A., 2016. Artificial neural network application for multiphase flow patterns detection: a new approach. J. Petrol. Sci. Eng. 145, 548–564.

Armaghani, D.J., Mohamad, E.T., Narayanasamy, M.S., Narita, N., Yagiz, S., 2017. Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. Tunn. Undergr. Space. Technol. 63, 29–43.

Ashoori, S., Abedini, A., Abedini, R., Nasheghi, K.Q., 2010. Comparison of scaling equation with neural network model for prediction of asphaltene precipitation. J. Petrol. Sci. Eng. 72 (1–2), 186–194.

Basarir, H., Tutluoglu, L., Karpuz, C., 2014. Penetration rate prediction for diamond bit drilling by adaptive neuro-fuzzy inference system and multiple regressions. Eng. Geol. 173, 1–9.

Bolandi, V., Kadkhodaie, A., Farzi, R., 2017. Analyzing organic richness of source rocks from well log data by using SVM and ANN classifiers: a case study from the Kazhdumi formation, the Persian Gulf basin, offshore Iran. J. Petrol. Sci. Eng. 151, 224–234.

Boyacioglu, M.A., Avci, D., 2010. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. Expert Syst. Appl. 37 (12), 7908–7912.

Breiman, L., 2017. Classification and Regression Trees. Routledge.

Ceryan, N., Okkan, U., Kesimal, A., 2013. Prediction of unconfined compressive strength of carbonate rocks using artificial neural networks. Environ. Earth Sci. 68 (3), 807–819.

Deosarkar, M.P., Sathe, V.S., 2012. Predicting effective viscosity of magnetite ore slurries by using artificial neural network. Powder. Technol. 219, 264–270.

Derakhshanfard, F., Mehralizadeh, A., 2018. Application of artificial neural networks for viscosity of crude oil-based nanofluids containing oxides nanoparticles. J. Petrol. Sci. Eng. 168, 263–272.

Elkatatny, S., Tariq, Z., Mahmoud, M., 2016. Real time prediction of drilling fluid rheological properties using artificial neural networks visible mathematical model (white box). J. Petrol. Sci. Eng. 146, 1202–1210.

Modeling the properties of crude oil and gas systems using RBF network. In: Elsharkawy, A.M. (Ed.), SPE Asia Pacific Oil and Gas Conference and Exhibition. Society of Petroleum Engineers.

Fausett, L.V., 1994. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Prentice-Hall Englewood Cliffs.

Use of cement as lost circulation material-field case studies. In: Fidan, E., Babadagli, T., Kuru, E. (Eds.), IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition. Society of Petroleum Engineers.

Integrated workflow for lost circulation prediction. In: Ghalambor, A., Salehi, S., Shahri, M.P., Karimi, M. (Eds.), SPE International Symposium and Exhibition on Formation Damage Control. Society of Petroleum Engineers.

Ghoreishi, S., Heidari, E., 2013. Extraction of epigallocatechin-3-gallate from green tea via supercritical fluid technology: neural network modeling and response surface optimization. J. Supercrit. Fluids 74, 128–136.

Haykin, S., Network, N., 2004. A comprehensive foundation. Neural. Network. 2 (2004), 41.

Heidari, E., Sobati, M.A., Movahedirad, S., 2016. Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). Chemometr. Intell. Lab. Syst. 155, 73–85.

Hemmati-Sarapardeh, A., Ghazanfari, M.H., Ayatollahi, S., Masihi, M., 2016. Accurate determination of the CO2-crude oil minimum miscibility pressure of pure and impure CO2 streams: a robust modelling approach. Can. J. Chem. Eng. 95, 253–261.

Jang, J.-S., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685.

Khoshjavan, S., Heidary, M., Rezai, B., 2010. Estimation of coal swelling index based on chemical properties of coal using artificial neural networks. Iran. J. Mater. Sci. Eng. 7 (3).

Khosravanian, R., Sabah, M., Wood, D.A., Shahryari, A., 2016. Weight on drill bit prediction models: sugeno-type and Mamdani-type fuzzy inference systems compared. J. Nat. Gas Sci. Eng. 36, 280–297.

Lashkarbolooki, M., Hezave, A.Z., Ayatollahi, S., 2012. Artificial neural network as an applicable tool to predict the binary heat capacity of mixtures containing ionic liquids. Fluid. Phase. Equilib. 324 (0), 102–107.

Lashkenari, M.S., Taghizadeh, M., Mehdizadeh, B., 2013. Viscosity prediction in selected Iranian light oil reservoirs: artificial neural network versus empirical correlations. Petrol. Sci. 10 (1), 126–133.

Le Van, S., Chon, B.H., 2017. Evaluating the critical performances of a CO2–Enhanced oil recovery process using artificial neural network models. J. Petrol. Sci. Eng. 157, 207–222.

Li, X., Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int. J. Geogr. Inf. Sci. 16 (4), 323–343.

Martin, K.G., Totten, M.W., Raef, A., 2017. Characterization of a reservoir ooid shoal complex and Artificial Neural Networks application in lithofacies prediction: mississippian St. Louis formation, Lakin fields, western Kansas. J. Petrol. Sci. Eng. 150, 1–12.

Maucec, M., Singh, A.P., Bhattacharya, S., Yarus, J.M., Fulton, D.D., Orth, J.M., 2015. Multivariate analysis and data mining of well-stimulation data by use of classification-and-regression tree with enhanced interpretation and prediction capabilities. SPE Econ. Manag. 7 (02), 60–71.

Messenger JU, 1981. Lost Circulation.

Meulenkamp, F., Grima, M.A., 1999. Application of neural networks for the prediction of the unconfined compressive strength (UCS) from Equotip hardness. Int. J. Rock Mech. Min. Sci. 36 (1), 29–39.

Prediction of lost circulation using virtual intelligence in one of Iranian oilfields. In: Moazzeni, A., Nabaei, M., Jegarluei, S.G. (Eds.), Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.

Mohaghegh, S., 2000. Virtual intelligence and its applications in petroleum engineering. J. Petrol. Technol. 52 (9), 1–24 Distinguished Author Series.

Murchison, W., 2006. Lost Circulation for the Man on the Rig. Murchison Drilling Schools. Inc, Albuquerque, NM.

Najafi-Marghmaleki, A., Barati-Harooni, A., Tatar, A., Mohebbi, A., Mohammadi, A.H., 2017. On the prediction of Watson characterization factor of hydrocarbons. J. Mol. Liq. 231, 419–429.

Nasiri, A., Ghaffarkhah, A., Moraveji, M.K., Gharbanian, A., Valizadeh, M., 2017. Experimental and field test analysis of different loss control materials for combating lost circulation in bentonite mud. J. Nat. Gas Sci. Eng. 44, 1–8.

Artificial neural network model for predicting wellbore instability. In: Okpo, E., Dosunmu, A., Odagme, B. (Eds.), SPE Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.

Openshaw, S., Openshaw, C., 1997. Artificial Intelligence in Geography.

Orr, M.J., 1996. Introduction to Radial Basis Function Networks. Technical Report, Center for Cognitive Science. University of Edinburgh.

Park, J., Sandberg, I.W., 1991. Universal approximation using radial-basis-function networks. Neural. Comput. 3 (2), 246–257.

Pham, B.T., Bui, D.T., Pourghasemi, H.R., Indra, P., Dholakia, M., 2017. Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theor. Appl. Climatol. 128 (1–2), 255–273.

Effect of material type and size distribution on performance of loss/seepage control material. In: Pilehvari, A.A., Nyshadham, V.R. (Eds.), International Symposium and Exhibition on Formation Damage Control. Society of Petroleum Engineers.

Roe, B.P., Yang, H.-J., Zhu, J., Liu, Y., Stancu, I., McGregor, G., 2005. Boosted decision trees as an alternative to artificial neural networks for particle identification. Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip. 543 (2–3), 577–584.

Salehinia, S., Salehinia, Y., Alimadadi, F., Sadati, S.H., 2016. Forecasting density, oil formation volume factor and bubble point pressure of crude oil systems based on nonlinear system identification approach. J. Petrol. Sci. Eng. 147, 47–55.

Shan, L., Cao, L., Guo, B., 2018. Identification of flow units using the joint of WT and LSSVM based on FZI in a heterogeneous carbonate reservoir. J. Petrol. Sci. Eng. 161, 219–230.

Root-cause identification and production diagnostic for gas wells with plunger lift. In: Singh, A. (Ed.), SPE Reservoir Characterisation and Simulation Conference and Exhibition. Society of Petroleum Engineers.

Singh, A., 2017. Application of data mining for quick root-cause identification and automated production diagnostic of gas wells with plunger lift. SPE. Prod. Oper. 32 (03).

Soofastaei, A., Aminossadati, S.M., Arefi, M.M., Kizil, M.S., 2016. Development of a multilayer perceptron artificial neural network model to determine haul trucks energy consumption. Int. J. Min. Sci. Technol. 26 (2), 285–293.

Sugeno, M., Kang, G., 1988. Structure identification of fuzzy model. Fuzzy. Sets Syst. 28 (1), 15–33.

Neural networks for bond rating improved by multiple hidden layers. Neural Networks. In: Surkan, A.J., Singleton, J.C. (Eds.), 1990 IJCNN International Joint Conference on; 1990. IEEE.

Torabi, F., Abedini, A., Abedini, R., 2011. The development of an artificial neural network model for prediction of crude oil viscosities. Petrol. Sci. Technol. 29 (8), 804–816.

Venkatesan, P., Anitha, S., 2006. Application of a radial basis function neural network for diagnosis of diabetes mellitus. Curr. Sci. 91 (9), 1195–1199.

Wang, H., Sweatman, R.E., Engelman, R., Deeg, W.F., Whitfill, D.L., Soliman, M.Y., et al., 2008. Best practice in understanding and managing lost circulation challenges. SPE. Drill. Complet. 23 (02), 168–175.

Preventing lost circulation requires planning ahead. In: Whitfill, D.L., Jamison, D.E., Wang, M., Angove-Rogers, A. (Eds.), International Oil Conference and Exhibition in Mexico. Society of Petroleum Engineers.

Wu, Y., Wang, H., Zhang, B., Du, K.-L., 2012. Using radial basis function networks for function approximation and classification. ISRN. Appl. Math. 2012.

Xue, Y., Cheng, L., Mou, J., Zhao, W., 2014. A new fracture prediction method by combining genetic algorithm with neural network in low-permeability reservoirs. J.

Petrol. Sci. Eng. 121, 159–166.
Yilmaz, I., Kaynar, O., 2011. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. Expert. Syst. Appl. 38 (5), 5958–5966.
Yılmaz, I., Yuksek, A., 2008. An example of artificial neural network (ANN) application for indirect estimation of rock parameters. Rock. Mech. Rock Eng. 41 (5), 781–795.
Zadeh, L.A., 1988. Fuzzy logic. Computer 21 (4), 83–93.
Zahmatkesh, I., Soleimani, B., Kadkhodaie, A., Golalzadeh, A., Abdollahi, A.-M., 2017. Estimation of DSI log parameters from conventional well log data using a hybrid particle swarm optimization–adaptive neuro-fuzzy inference system. J. Petrol. Sci. Eng. 157, 842–859.
Zhang, X., Zhang, H., Guo, J., Zhu, L., 2018. Auto measurement while drilling mud pulse signal recognition based on deep neural network. J. Petrol. Sci. Eng. 167, 37–43.