Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations

Ahmad Al-AbdulJabbar [1] · Ahmed Abdulhamid Mahmoud [1] · Salaheldin Elkatatny [1]

Abstract

Rate of penetration (ROP) is a critical parameter affecting the total cost of drilling an oil well. This study introduces an empirical equation developed based on the optimized artificial neural networks (ANNs) for estimation of the rate of penetration (ROP) in real-time while horizontally drilling natural gas-bearing sandstone reservoirs based on the surface measurable drilling parameters of the mud injection rate, drillstring rotation speed (DSR), standpipe pressure, torque, and weight on bit (WOB) in combination with ROP_c, which is a new parameter developed in this study based on regression analysis. The ANN model was learned and optimized using 1154 data points; the training parameters were collected while horizontally drilling natural gas-bearing sandstone formations in Well-A. An empirical equation for ROP estimation was developed based on the optimized ANN model. Moreover, 495 unseen data points from Well-A were used to test the developed ROP equation, which was finally validated on 2213 data points from Well-B. The predictability of the new ROP equation was compared with the available correlations. The results showed that, without considering ROP_c, the optimized ANN model estimated the ROP for the training dataset with an average absolute percentage error (AAPE) of 42.6% and correlation coefficient (R) of 0.424, while when ROP_c was considered as an input, the AAPE decreased to 5.11% and R increased to 0.991. The new empirical equation estimated the ROP for the testing data of Well-A with AAPE and R of 5.39% and 0.989 and for the validation data of Well-B with AAPE and R of 8.85% and 0.954, respectively. The new empirical equation overperformed all the available empirical correlations for ROP estimation.

Keywords Rate of penetration · Sandstone formations · Horizontal drilling · Artificial neural networks

Nomenclature

AAPE	Average absolute percentage error
ANN	Artificial neural network
DSR	Drillstring rotation speed
MSE	Mechanical specific energy
Q	Mud injection rate
R	Correlation coefficient
RMSE	Root mean square error
ROP	Rate of penetration
ROP_c	Calculated ROP
SPP	Standpipe pressure
T	Torque
WOB	weight on bit

✉ Salaheldin Elkatatny
elkatatny@kfupm.edu.sa

[1] College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Introduction

One of the most critical factors affecting the total cost of oil well's drilling is the required time to complete the drilling operations (Lyons and Plisga, 2004). Rig time is a function of several parameters including the rate of penetration (ROP), which represents the number of feet drilled per 1 h; therefore, ROP is considered as the most important factor controlling the rig time and the cost of drilling (Barbosa et al., 2019).

There are several parameters controlling the ROP that could be subdivided into two main categories of controllable and uncontrollable parameters (Hossain and Al-Majed, 2015). The mud injection rate (Q), drillstring rotation speed (DSR), standpipe pressure (SPP), torque (T), and the weight on bit (WOB) are all considered as controllable parameters (Eren

and Ozbayoglu, 2010; Mitchell and Miska, 2011; Payette et al., 2017), while the drilling fluid type, rheological properties, and density and the drill bit size are all uncontrollable parameters. Quantification of the effect of the uncontrollable parameters on the ROP is complicated because the change in any of these parameters affects the others (Osgouei, 2007). The ROP is also affected by the hole cleaning conditions, especially for the inclined and horizontal wells (Mahmoud et al., 2020a).

Several previous studies were conducted to develop models for ROP estimation; these models considerably vary in terms of accuracy because of the variation in the parameters considered to calculate the ROP by every one of these models, which significantly limits their applicability (Soares et al., 2016; Soares and Gray, 2019).

There are two types of ROP prediction models: traditional models and data-driven-based models (Hegde et al., 2017). The traditional models are empirical correlations developed based on regression analysis, and the data-driven models are developed based on machine learning techniques.

Empirical equations for prediction of the ROP

Maurer (1962) developed the first empirical correlation for ROP estimation while drilling with the rolling cutter bits. Maurer's (1962) empirical correlation in Eq. (1) estimates the ROP as a function of the WOB, DSR, drill bit size, and rock strength.

$$ROP = \frac{k}{S^2} \left(\frac{WOB}{d_b} - \frac{W_t}{d_b} \right)^2 RPM \tag{1}$$

where K represents the constant of proportionality; S denotes the compressive strength of the formation; WOB is the weight on bit, Klb_f; W_t represents the threshold value of the bit weight; and d_b denotes the drill bit diameter, in. W_t in Eq. (1) is too much smaller than the WOB, and hence, for simplification, the second term in Eq. (1) could be assumed equal to zero.

Another ROP empirical correlation of Eq. (2) was developed by Bingham (1965). In his model, Bingham (1965) combined the rock strength's effect into the constant of proportionality and considered a varying exponent (a_5) to replace the constant exponent of Eq. (1).

$$ROP = k \left(\frac{WOB}{d_b} \right)^{a_5} RPM \tag{2}$$

where K represents the constant of proportionality, which also includes the rock strength's effect, and a_5 denotes the WOB exponent.

Considering Eq. (1) and Eq. (2) developed by Maurer (1962) and Bingham (1965), respectively, it is clear that both

models did not account for the effect of the differential pressure, drill bit's hydraulics and wear, and formation compaction on the change in the ROP, and this considerably reduced the accuracy of ROP estimated using these models.

Bourgoyne and Young (1974) suggested Eq. (3) for ROP estimation and optimization of the drilling process. In their model, Bourgoyne and Young (1974) considered the effect of the formation's compaction, depth, and strength; the bit's diameter, wear, and hydraulics; and the bottom hole's pressure differential, WOB, and DSR on the ROP.

$$\frac{dD}{dt} = e^{\left[a_1 + \sum_{j=2}^{8} a_j x_j \right]} \tag{3}$$

where D represents the well's true vertical depth, ft.; t denotes the time, the constants a_1–a_8 are drilling parameter coefficients; and x_2–x_8 represent the drilling parameters in dimensionless format calculated as a function of the real drilling parameters, where a_1 accounts for the formation strength, $a_2 x_2$ and $a_3 x_3$ consider the compaction of the formation, $a_4 x_4$ accounts for the differential pressure, $a_5 x_5$ models the effect of the bit diameter and WOB, $a_6 x_6$ accounts for the change in the DSR, $a_7 x_7$ considers the variation in the drill bit's tooth wear, and $a_8 x_8$ models the impact of the bit hydraulic jet. Eqs. (4) to (10) could be used to calculate the variable x_j.

$$x_2 = 10,000 - D \tag{4}$$

$$x_3 = D^{0.69}(pore\ pressure\ gradient - 9.0) \tag{5}$$

$$x_4 = D(pore\ pressure\ gradient - equivalent\ circulation\ density) \tag{6}$$

$$x_5 = ln \left[\frac{WOB/d_b - \left(WOB/d_b \right)_t}{4.0 - \left(WOB/d_b \right)_t} \right] \tag{7}$$

$$x_6 = ln \left[\frac{RPM}{100} \right] \tag{8}$$

$$x_7 = -fractional\ tooth\ height\ worn\ away \tag{9}$$

$$x_8 = \frac{\rho\ q}{350\ \mu\ d_n} \tag{10}$$

where ρ represents the drilling mud's density, lb/gal; q is the drilling fluid's flow rate, gal/min; μ represents the viscosity of the drilling mud, cP; and d_n is the diameter of the drill bit, inches.

Applications of machine learning techniques for evaluation of the rate of penetration

Machine learning techniques were applied in different applications in scientific research areas (Hag Elsafi, 2014; Babikir et al., 2019), including the petroleum industry where the AI techniques were used to solve difficult problems such as

prediction of formation tops (Elkatatny et al., 2019), identification of lithology (Ren et al., 2019), evaluation of drill bit wear using drilling parameters (Arehart, 1990), estimation of the drilling fluid rheology (Elkatatny, 2017; Abdelgawad et al., 2018), estimation of total organic carbon (Mahmoud et al., 2017a, 2017b, 2019a, 2019b, 2020b), estimation of the oil recovery factor (Mahmoud et al., 2017c, 2019c), prediction of the pore and fracture pressures (Ahmed et al., 2019a, 2019b), evaluation of the static Young's modulus (Mahmoud et al., 2019d, 2020c, 2020d, 2020e), optimization of the rate of penetration (Al-AbdulJabbar et al., 2018; Mahmoud et al., 2020f), and detection of the downhole anomalies (Alsaihati et al., 2021).

Artificial neural network (ANN) is a machine learning tool inspired by biological neural networks and developed to mimic animal brains. In its simplest form, the ANN model consists of three layers: single input, single training, and single output layers. Any of these layers has a collection of connected neurons that model the biological brain nodes. Every neuron in the input layer represents a single input parameter, while the number of the neurons in the training layer optimized to predict the targeted parameter, and the neuron in the output layer represents the output parameter. Different training and transferring functions are usually evaluated during the model optimization stage to find the optimum weights and biases associated with the input, training, and output layers that will optimize the predictability of the targeted parameter.

To overcome the weakness of the empirical equations developed based on the linear regression analysis on estimating the ROP with low accuracy, Bilgesu et al. (1997) introduced the application of artificial intelligence for estimation of the ROP and suggested two artificial neural network (ANN) models to estimate the ROP for nine formations. In their first model, the authors used the formation type; drill bit's diameter, tooth, bearing wear, and type; gross hours of drilling; mud circulation; WOB; footage; and DSR as inputs to predict the actual ROP, while the bearing wear and bit tooth were excluded from the input variables in the second model. The results of this study showed that both models accurately estimated the ROP.

In another study, Amar and Ibrahim (2012) also suggested another two ANN models to predict the ROP from the formation depth, DSR, WOB, ECD, the formation pore pressure gradient, Reynolds number function, and the drill bit's tooth wear. The results showed that the ANN-based models were able to estimate the ROP with higher accuracy compared with the available empirical equations.

Elkatatny (2018) developed an equation for the estimation of the ROP in vertical wells based on the optimized ANN. The developed equation estimated the ROP based on the surface measurable drilling parameters of the Q, DSR, T, WOB, and standpipe pressure (SPP) in combination with the drilling fluid

properties of the plastic viscosity (PV) and mud weight (MW). The author evaluated his equation in real data, and it evaluated the ROP with a very low average absolute percentage error (AAPE) of 4% compared to AAPE of more than 10% for the estimation with available empirical equations.

Al-AbdulJabbar et al. (2020) optimized an ANN model for ROP estimation on carbonate formation during horizontal drilling. This model is based on the use of the Q, DSR, and T, in combination with the conventional well log data of the gamma ray, formation bulk density, and deep resistivity. This model showed a great improvement in predicting the ROP for the carbonate formations.

In this study, a new model for ROP estimation in sandstone formations during the horizontal drilling process was developed based on the surface measurable parameters of the Q, DSR, SPP, WOB, and T, with a newly developed parameter called calculated ROP (ROP_c), which is defined in the study for the first time from the DSR, the WOB, the drill pipe diameter (D), and the drilled hole area (A).

Methodology

In this study, the ANN technique was applied to develop a model to enable the estimation of the ROP in real-time while horizontally drilling through natural gas-bearing sandstone formations based on only the surface measurable drilling parameters of the Q, DSR, SPP, T, and WOB, with the ROP_c parameter. The expression for the ROP_c is developed in this study, which determines the ROP based on the DSR, the WOB, the drill pipe diameter (D), and the drilled hole area (A).

Data preparation and preprocessing

To train the ANN model, 3082 datasets of the different input drilling parameters and their corresponding actual ROP were collected from an oil well (Well-A) in the Middle East; the input parameters are all surface measurable in the real-time base; this is considered to enable real-time prediction of the ROP will drilling. Another 4662 datasets of the inputs and ROP are collected from another well (Well-B) from the same oil field. The data gathered from both well is collected while horizontally drilling through natural gas-bearing sandstone formations. Before introducing the inputs into the ANN model, the data was evaluated to remove all unrealistic values and outliers. For unrealistic value determination, the mechanical specific energy (MSE), which is a parameter developed by Teale (1965), accounts for the energy required at the surface to drill a specific volume of the rock. According to Teale (1965), the value of the MSE should correlate with the crushing strength of the rock or the rock's compressive strength (UCS) value.
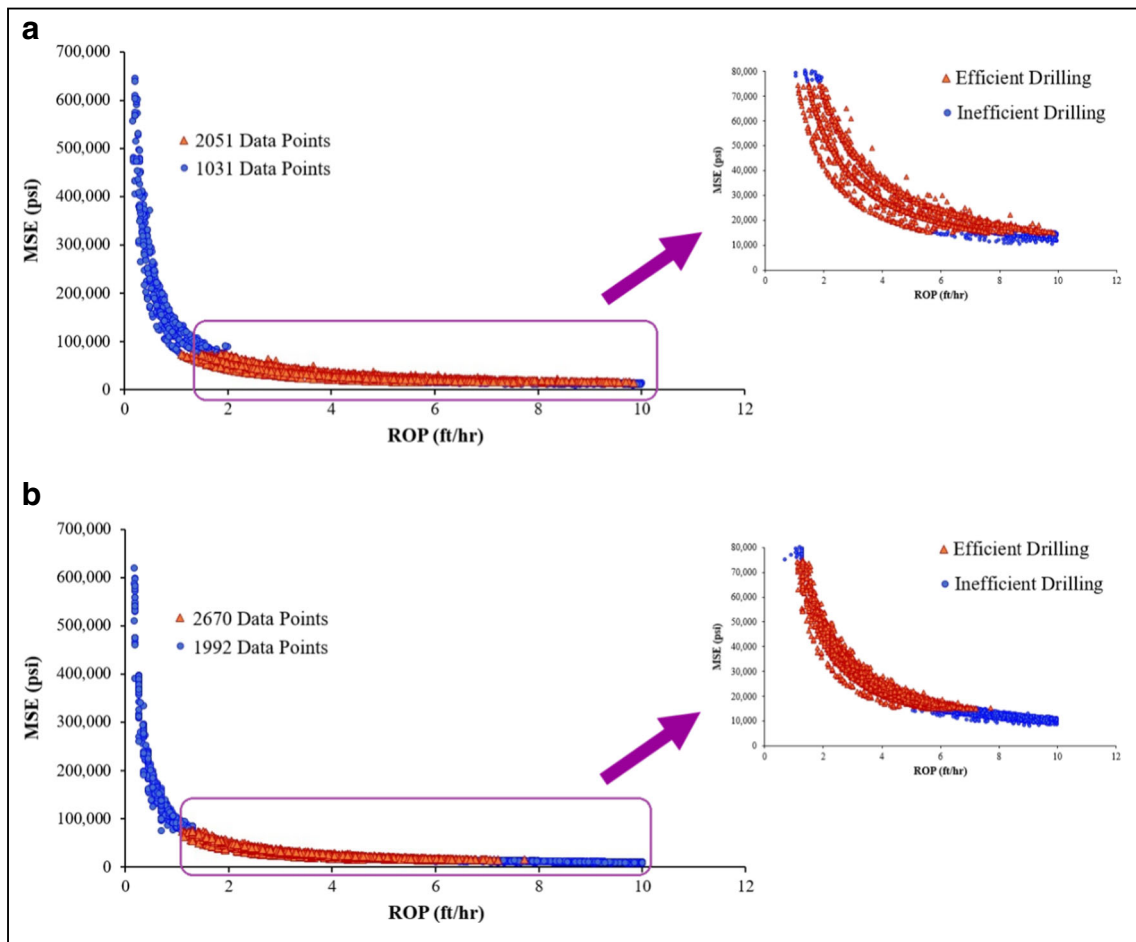
Fig. 1 The relationship between the MSE and the UCS at the points corresponding to the collected input drilling parameters for the sandstone formations considered in this study; these data are from **a** Well-A and **b** Well-B. Many MSE values are significantly different than their corresponding UCS; these data are unrealistic (inefficient drilling) and removed from the inputs

The UCS for the sandstone formations considered in this study is in the range of 25,000 to 45,000 psi. As indicated in Fig. 1, many of the MSE values for the data collected from both Well-A and Well-B are considerably greater or lower than the UCS; all these values are unrealistic and represent data of inefficient drilling, so all the data points with MSE

Fig. 2 The relationship between the MSE and ROP at DSR of 60, 80, and 100 rpm for the training data of Well-A. There is relationship between the MSE and ROP at every specific DSR best fitted by power functions with an exponent (n) of -1.0
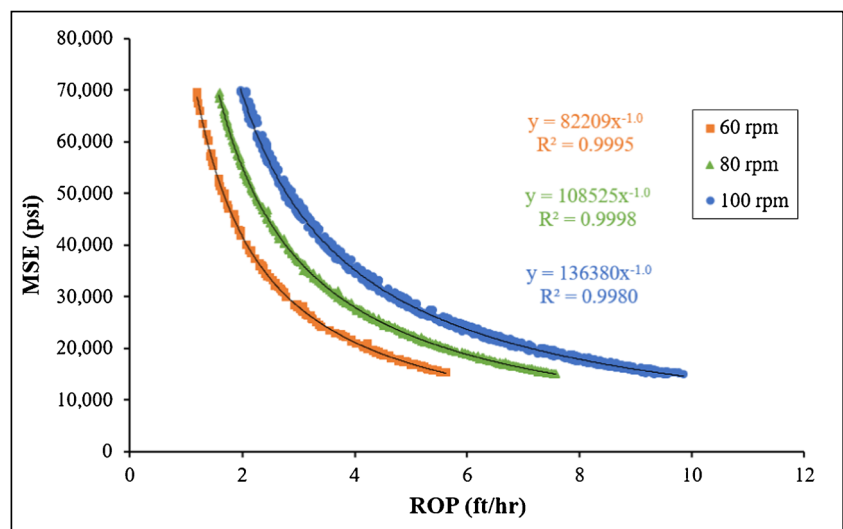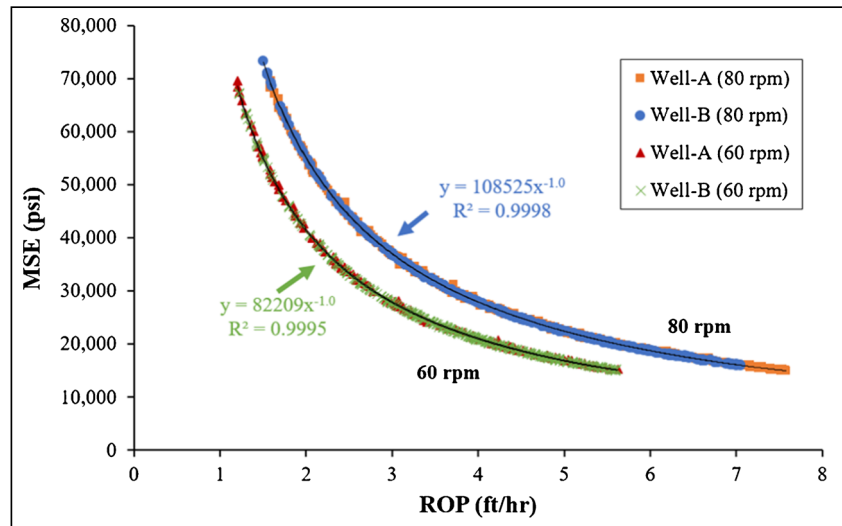
Fig. 3 The plot of the MSE versus ROP at DSR of 60 and 80 rpm for the data of both Well-A and Well-B. There is relationship between the MSE and ROP at every specific DSR best fitted by power functions with an exponent (n) of -1.0

values outside the range of 15,000 to 75,000 psi values are removed from the data considered in this study; this range of the MSE is selected by considering the formation UCS $\pm a$ margin. As shown in Fig. 1, out of the data gathered from Well-A and Well-B, 1031 and 1992 data points, respectively, represent locations of inefficient drilling; in this stage of data preprocessing, all the inefficient drilling data points are removed from the input data. After removing the inefficient (unrealistic) drilling data points, 2051 data points from Well-A and 2670 data points from Well-B were considered realistic.

The second step in the data preprocessing is to define and eliminate all outliers; for this purpose, the standard deviation is considered as the controlling factor in this step. The outliers in all input drilling parameters of Q, DSR, SPP, T, and WOB and their corresponding ROP are identified as the data points outside the range of ± 3.0 standard deviation; every dataset having an outlier is removed from the input data. After outlier removal, 1649 and 2213 of the surface measurable drilling

datasets gathered from Well-A and Well-B, respectively, are considered valid to develop the ANN-based model for ROP estimation.

Developing a new expression for the rate of penetration

To develop the new term for ROP, which is called the calculated ROP (ROP_c), the ROP_c will be considered as an input parameter to train the ANN model, starting from the MSE expression as in Eq. (11).

$$MSE = \frac{WOB}{A} + \frac{8 \times RPM}{d_b \times ROP} \tag{11}$$

The first term in Eq. (11) is much less than the second term, so the main dominant parameters will be the DSR and ROP. The objective at this step is to relate the MSE, DSR, and ROP

Fig. 4 The relationship between the constant "a" and the DSR. The values of the constant "a" are extracted from the plots of Fig. 2
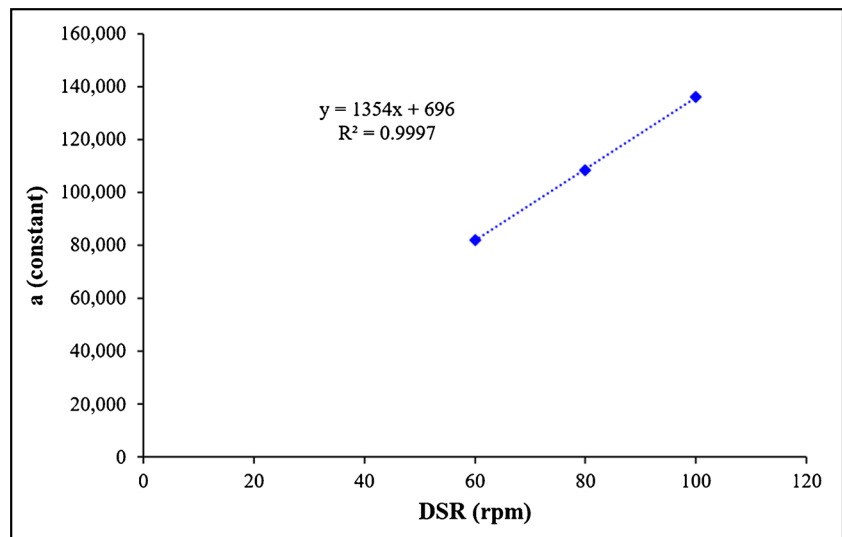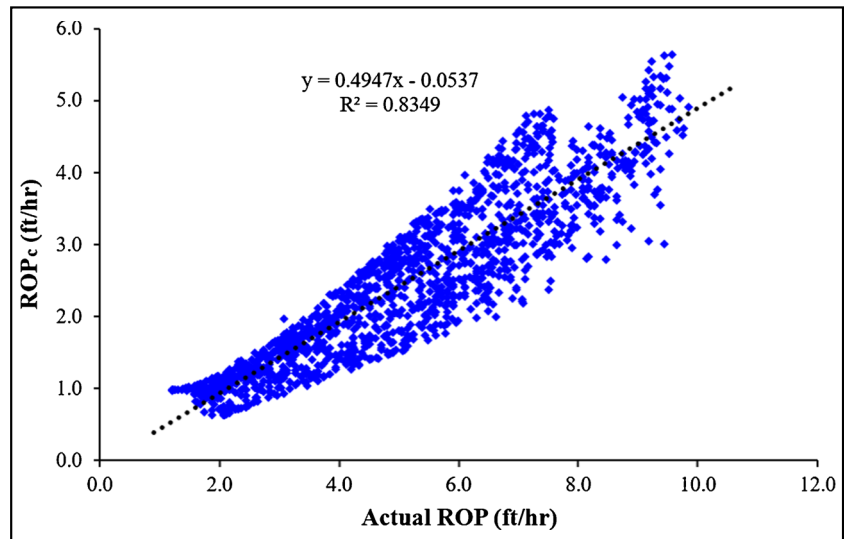
Fig. 5 The cross-plot of the
calculated rate of penetration
(ROP_c) and ROP for the data
collected from Well-A (1649 data
points)

parameters of the training data. By plotting the MSE and ROP
at DSR of 60, 80, and 100 rpm for the training data of Well-A
after preprocessing as in Fig. 2, it is noted that every one of the
three plots in Fig. 2 represents a relationship between the MSE

and ROP at specific DSR, and these three plots are best fitted
by power functions with an exponent (n) of − 1.0 and constant
(a) of 82,209, 108,525, and 136,380 for DSR of 60, 80, and
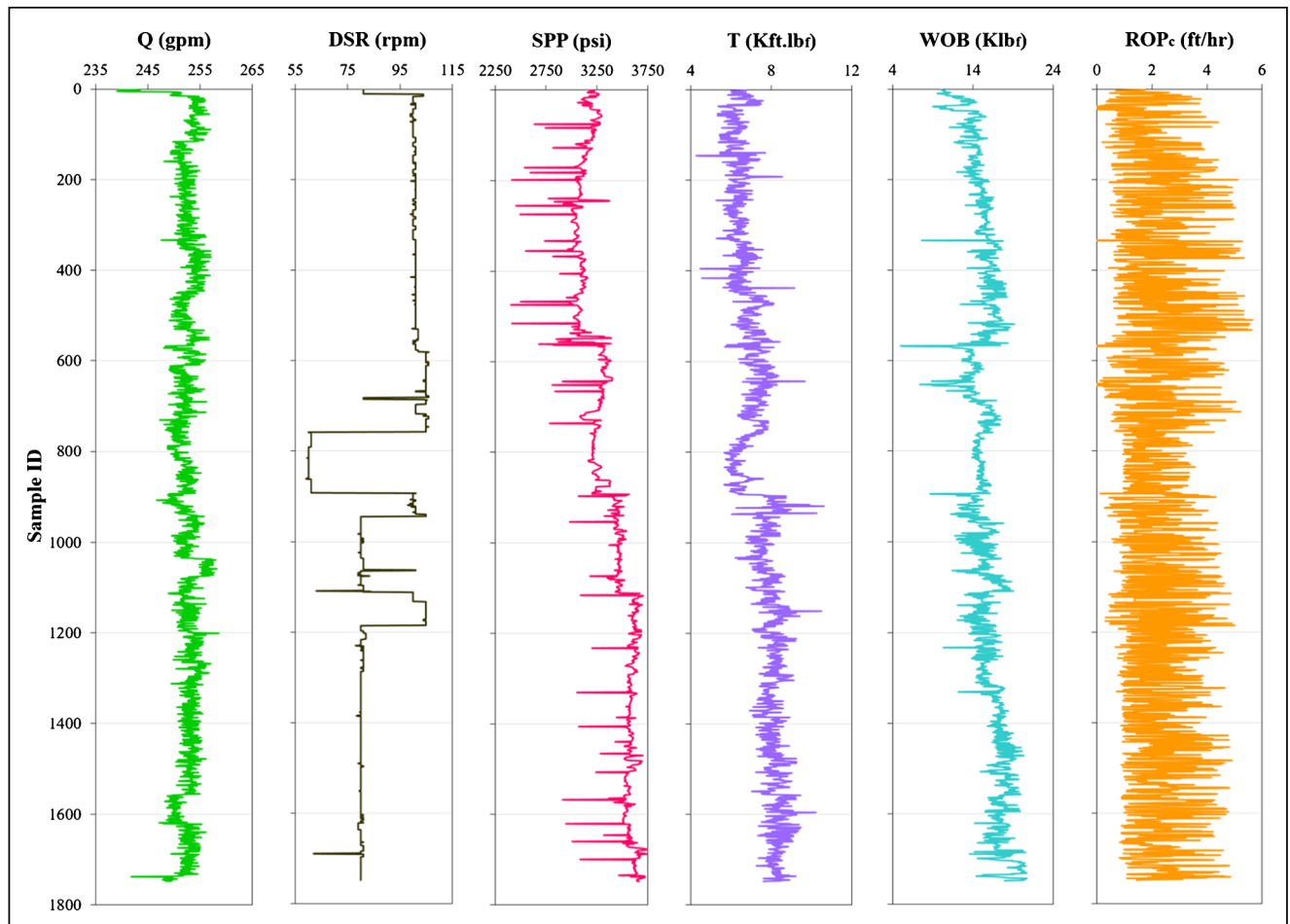100 rpm.

Fig. 6 The input surface measurable drilling parameters of Well-A, from left to right: Q, DSR, SPP, T, WOB, and ROP_c; the ROP_c is calculated using
Eq. (17). This is the data used to train the ANN model

Table 1 The ranges and statistical characteristics for the training drilling parameters and their corresponding ROP collected from Well-A

	Q (gpm)	DSR (rpm)	SPP (psi)	T (kft.lb$_f$)	WOB (klb$_f$)	ROP (ft/h)
Minimum	239	59.0	2401	4.32	5.07	1.20
Maximum	259	106	3746	10.6	20.7	9.85
Mean	253	89.4	3346	7.36	15.6	4.88
Median	253	82.0	3352	7.50	15.6	4.76
Standard deviation	1.95	13.4	229	0.93	1.93	2.09
Sample variance	3.82	179	52,531	0.86	3.73	4.38

The three power functions in Fig. 2 could be expressed in a general form as in Eq. (12) or Eq. (13).

$$MSE = a\, ROP^{-1.0} = \frac{a}{ROP} \tag{12}$$

$$MSE \times ROP = a \tag{13}$$

The exponent in Eq. (12) is simplified to − 1.0 to simplify the derivation of the ROP$_c$ expression.

To carry further analysis on the value of constant "a," multiple analysis was performed. In Fig. 3, in MSE and ROP relations, both Well-A and Well-B were plotted for the DSR of 60 and 80; it is clear in Fig. 3 that the relationships are similar at the same DSR values regardless of the source of the MSE and ROP values (Well-A or Well-B).

Now, let us plot the constant "a" and DSR as shown in Fig. 4. The plot in Fig. 4 confirms that the constant "a" and DSR have a linear relationship, which could be expressed in a general form as in Eq. (14).

$$a = 1354\, RPM + 696 \tag{14}$$

Now, by considering Teale expression for the MSE in Eq. (15), and neglecting the torque and substituting for the MSE from Eq. (12) into Eq. (15), we will have the relationship in Eq. (16):

$$MSE = \frac{WOB}{A} + \frac{2\pi \times RPM \times Torque}{A \times ROP} \tag{15}$$

$$\frac{a}{ROP} = \frac{WOB}{A} + \frac{2\pi \times RPM}{A \times ROP} \tag{16}$$

Table 2 The ANN model's optimum design parameters

Parameter	Value
Training function	Levenberg-Marquardt function
Transferring function	Pure linear function
Training layers	Single training layer
Neurons per layer	5

Rearranging Eq. (16) we will get:

$$ROP = \frac{-\left(2\pi\dfrac{RPM}{A} - a\right)A}{WOB} = ROP_c \tag{17}$$

where the expression for the ROP in Eq. (17) is a new expression and we call it here as calculated ROP of ROP$_c$. Equation (17) is relating the ROP$_c$ to two parameters only: DSR and WOB even after the constant "a" was expanded. The idea behind the previous equations is to create an offset in the equation to make the ROP prediction much easier. The
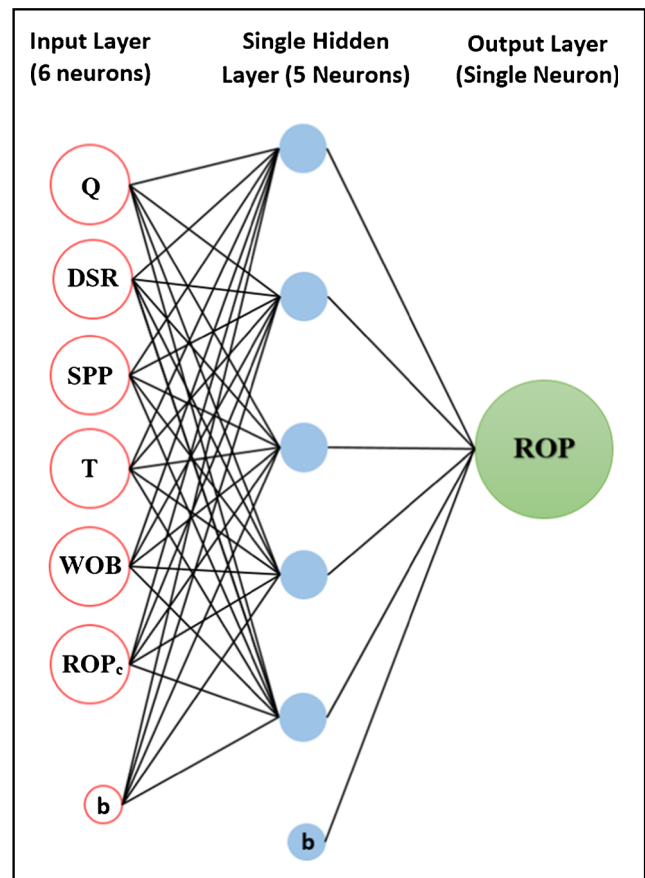

Fig. 7 Schematic of the optimized ANN model for ROP estimation. The letter "b" denotes the bias. This model consists of six neurons in the input layer, five neurons in the training layer, with a single neuron in the output layer

cross-plot of Fig. 5 compares the actual ROP and ROP_c calculated using Eq. (17). Even though the data fit is not perfect, the correlation coefficient is just enough to guide the model toward better ROP prediction. As explained earlier, the newly introduced parameter (ROP_c) will be used as an input to train the ANN model along with the five surface measurable drilling parameters (i.e., Q, DSR, SPP, *T*, and WOB).

Optimizing the artificial neural network model

ANN model was optimized in this study to predict the ROP as a function of six parameters; five surface measurable variables of the Q, DSR, *T*, SPP, and WOB and the sixth parameter are the ROP_c calculated by Eq. (17). The sixth parameter (ROP_c) is included as an input to improve the ROP estimation accuracy.

The ANN model was trained on 1154 datasets of the surface measurable parameters and their corresponding calculated ROP_c to estimate the actual ROP; these are the data collected from Well-A, which represents 70% of Well-A's data after unrealistic value and outlier removal. Figure 6 shows the input surface measurable parameters and ROP_c calculated using Eq. (17) for Well-A.

The statistical features of the training surface measurable drilling parameters and their corresponding ROP values of Well-A are in Table 1. As summarized in this table, the Q values are ranging from 239 to 259 gal per minute (gpm), DSR is ranging between 59.0 and 106 rotations per minute (rpm), SPP is in the range from 2401 to 3746 psi, *T* is ranging between 4.32 and 10.6 kft.lb$_f$, WOB is in the range from 5.07 to 20.7 klb$_f$, and ROP is between 1.20 to 9.85 ft/h. The statistical characteristics listed in Table 1 are very important because they show the applicable range for the optimized ANN model and the empirical correlation to be developed out of this model.

Sensitivity analysis was performed to optimize the ANN model design parameters of the training and transferring functions, the number of training layers, and the optimum number of neurons per every training layer.

During this stage, different training functions such as the gradient descent with adaptive learning rate backpropagation function, Levenberg-Marquardt function, and resilient backpropagation function were evaluated. Predictability of the transferring functions of the pure linear function, logarithmic sigmoid function, and tangential sigmoid function was also studied. The use of one, two, and three training layers with the use of one to 30 neurons in every layer was also evaluated.

Table 2 lists the optimum ANN model's design parameters that according to the sensitivity analysis conducted in this study enabled the prediction of the ROP with the lowest AAPE and root mean square error (RMSE) and the highest *R*. Figure 7 also shows the schematic of the optimized ANN model, which consists of one input layer having six neurons for the six inputs, one training layer having five neurons, and one output layer. This training data was obtained from Well-A.

The effect of excluding one input parameter from the training input variables (i.e., Q, DSR, SPP, *T*, WOB, and ROP_c) was also studied during the sensitivity analysis; at this stage, the aim is to investigate the impact of neglecting a single input variable, especially the newly developed ROP_c, on the predictability of the ROP.

After the sensitivity analysis, and based on the calculated AAPE in Eq. (18), RMSE in Eq. (19), and correlation coefficient (*R*) in Eq. (20), the system with the lowest AAPE and RMSE and the highest *R* was selected as the optimized ANN model for ROP prediction.

$$AAPE = \frac{1}{N} \sum_{i=1}^{N} \left(\left| \frac{(ROP_r)_i - (ROP_p)_i}{(ROP_r)_i} \right| \times 100 \right) \quad (18)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[(ROP_r)_i - (ROP_p)_i \right]^2} \quad (19)$$

$$R = \frac{\sum_{i=1}^{N} \left[\left((ROP_r)_i - \overline{ROP_r} \right) - \left((ROP_p)_i - \overline{ROP_p} \right) \right]}{\sqrt{\sum_{i=1}^{N} \left[(ROP_r)_i - \overline{ROP_r} \right]^2 \sum_{i=1}^{N} \left[(ROP_p)_i - \overline{ROP_p} \right]^2}} \quad (20)$$

Table 3 The weights and biases extracted from the optimized ANN model for ROP estimation, with *i* and *j* indexes accounting for the inputs and neurons, respectively

Training layer									Output layer	
Weight (w1)								Biases (b1)	Weights (w2)	Bias (b2)
		i=1	*i*=2	*i*=3	*i*=4	*i*=5	*i*=6			
No. of Neurons	*j*=1	−0.0211	0.1601	0.7169	−0.3710	−0.1519	1.0517	−0.1014	0.5592	−0.3896
	j=2	0.4345	0.5034	−0.0782	0.7740	0.3895	0.6545	0.0260	0.1292	
	j=3	0.5855	0.0954	0.7083	0.9679	0.7208	0.3322	0.6017	−0.0163	
	j=4	−0.1429	0.1952	−0.8266	−0.0251	−1.1259	1.3125	0.0276	0.6385	
	j=5	−0.1350	0.1133	−0.8087	−0.5706	−0.9110	−0.7650	−0.0256	−0.2254	

where N is the total number of the datasets and the subscripts r and p represent the real and predicted ROP.

Developing the new equation for estimation of the rate of penetration

After optimizing the ANN model, its weights and biases were extracted to develop the new ROP empirical equation. As indicated earlier in Table 2 and Fig. 7, the optimum ANN model has a single training layer associated with five neurons and it calculated the ROP as a function of six inputs using the Levenberg-Marquardt training function and pure linear transferring function. The general form of the equation that represents the ANN model with pure linear transferring function is as in Eq. (21).

$$y = \left(\sum_{j=1}^{m} w_{j1} \left(\sum_{i=1}^{n} w_{ij}x_i + b_j \right) + b_1 \right) \tag{21}$$

where y is the objective parameter; i and j indexes account for the inputs and neurons; w represents the weights; m and n denote the number of neurons and inputs, respectively; x represents the input variables; and b represents the biases.

For the optimized ANN model of Fig. 7, Eq. (21) could be written as in Eq. (22).

$$ROP = \left(\sum_{j=1}^{5} w_{j1} \left(\sum_{i=1}^{6} w_{ij}x_i + b_j \right) + b_1 \right) \tag{22}$$

where the weights and biases required for Eq. (22) were extracted and summarized in Table 3.

By expanding Eq. (22), it will be as in Eq. (23).

$$ROP = a_1 GPM + a_2 RPM + a_3 SPP + a_4 Torque$$
$$+ a_5 WOB + a_6 ROP_c + c \tag{23}$$

To determine the coefficients a_1 to a_6 of Eq. (23), as explained in Appendix A, the output layer and training layer weight matrices, which are [1, 5] and [5, 6], must be multiplied to obtain a matrix of [1, 6]. The constant c could be determined by multiplying the output layer weight matrix by the training layer bias matrices, which are [1, 5] and [1, 5], respectively; this multiplication results in a scaler that is equal to -0.0398; adding this scaler to the output layer bias of -0.3896 results in -0.4294, which equals to the constant c in Eq. (23). More discussion about the matrix multiplication and determination of the coefficients a_1 to a_6 and the constant c is in Appendix A.

Substituting for the coefficients a_1 to a_6 and the constant c from Appendix A into Eq. (23) leads to the final ROP equation in Eq. (24).

$$ROP = -0.0260\ GPM + 0.2521\ RPM$$
$$+ 0.0338\ SPP - 0.0107\ T - 0.5599\ WOB$$
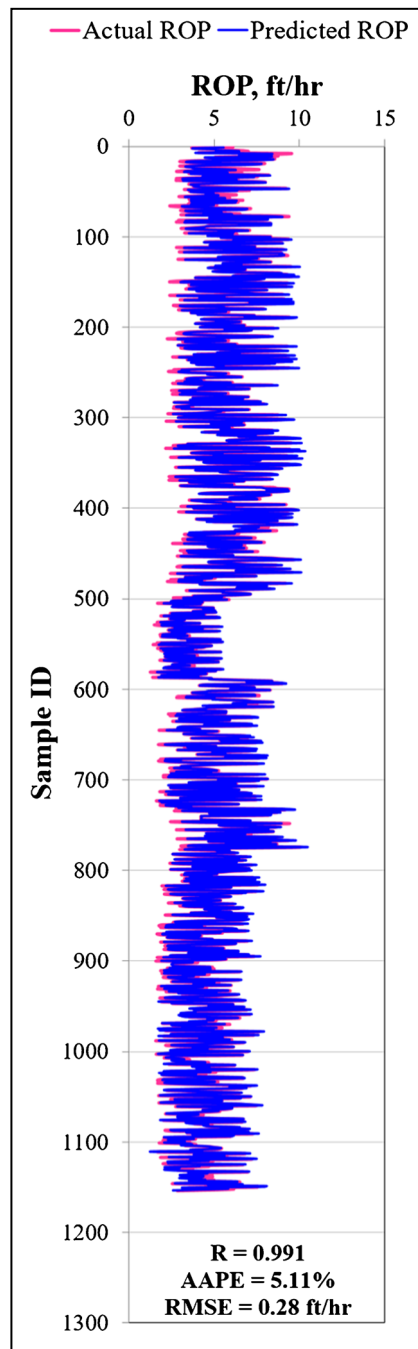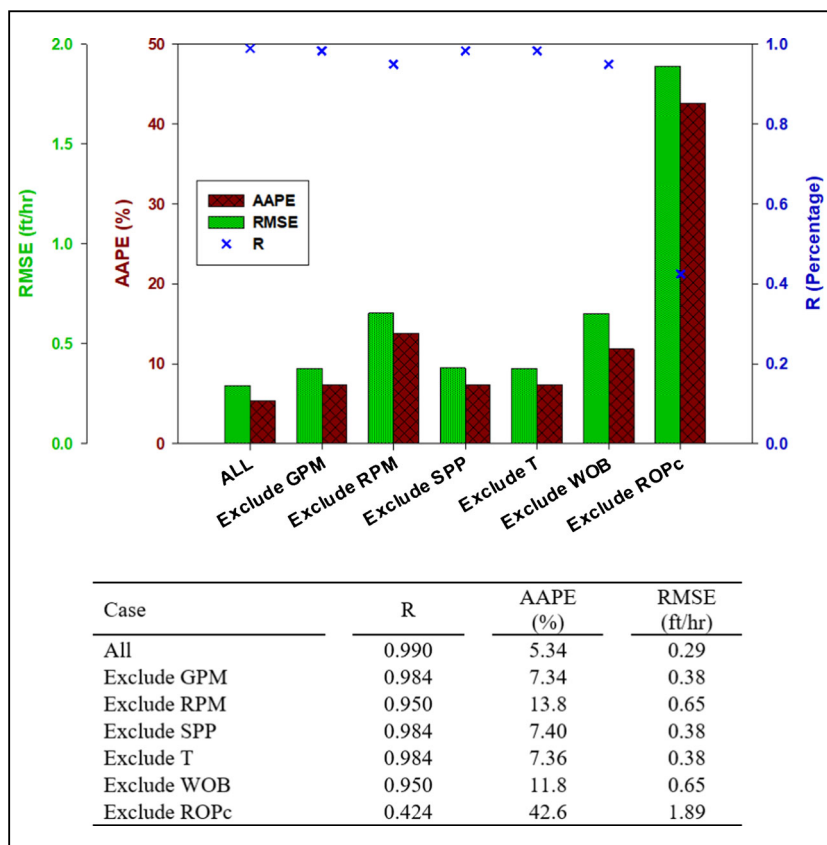$$+ 1.6778\ ROP_c - 0.4294 \tag{24}$$

Fig. 8 Comparison of the actual and predicted ROP for the training datasets collected from Well-A (1154 datasets). The ROP was predicted accurately with R, AAPE, and RMSE of 0.991, 5.11%, and 0.28 ft/h, respectively

Fig. 9 Comparison of the effect of excluding an input parameter from the training data. These results indicate that all inputs are necessary for ROP estimation

Case	R	AAPE (%)	RMSE (ft/hr)
All	0.990	5.34	0.29
Exclude GPM	0.984	7.34	0.38
Exclude RPM	0.950	13.8	0.65
Exclude SPP	0.984	7.40	0.38
Exclude T	0.984	7.36	0.38
Exclude WOB	0.950	11.8	0.65
Exclude ROPc	0.424	42.6	1.89

Testing and validating the new equation for estimation of the ROP

The developed equation for prediction of the ROP (Eq. (24)) was tested on 495 new unseen datasets from Well-A (30% of Well-A) and then validated on 2213 datasets collected from Well-B. The predictability of the developed equation for ROP in Well-B was also compared with the predictability of available empirical equations to evaluate the improvement in ROP prediction using Eq. (24) developed in this study.

Results and discussion

Training the artificial neural network model

The ANN model was firstly trained on 1154 datasets collected from Well-A to predict the ROP from six inputs of the Q, DSR, SPP, T, WOB, and ROP_c. As shown in Fig. 8, the optimized ANN model estimated the ROP accurately as confirmed by the excellent matching between the real and predicted ROP as well as the low AAPE and RMSE of 5.11% and 0.28 ft/h, respectively, and the high R of 0.991.

Sensitivity analysis of the training input variables

As explained earlier, the Q, DSR, SPP, T, WOB, and ROP_c are considered as inputs for the ANN model to estimate the ROP as output for Well-A. Figure 9 compares the AAPE, RMSE, and R in estimating the ROP for the use of all the six input variables and different cases of excluding one of the input variables. As indicated in Fig. 9, the use of all the six inputs enabled the prediction of the ROP accurately with AAPE and RMSE of 5.34% and 0.29 ft/h, respectively, and R of 0.990. Excluding any of the six input parameters from the training input data reduced the accuracy of the ANN model predictability as confirmed by the increase in the AAPE and RMSE and the decrease in R.

Excluding the ROP_c from the input training data significantly reduced the accuracy of the ANN model for ROP estimation. The ANN model estimated the ROP with a very high AAPE of 42.6%, RMSE of 1.89 ft/h, and R of only 0.424, as indicated in Fig. 9. These results confirmed the importance of including the ROP_c parameter calculated using Eq. (17) as an input parameter to train the ANN model for ROP estimation.

Testing the developed equation for the rate of penetration

The developed equation for ROP prediction in Eq. (24) was tested on another 495 datasets collected from Well-A (30% of
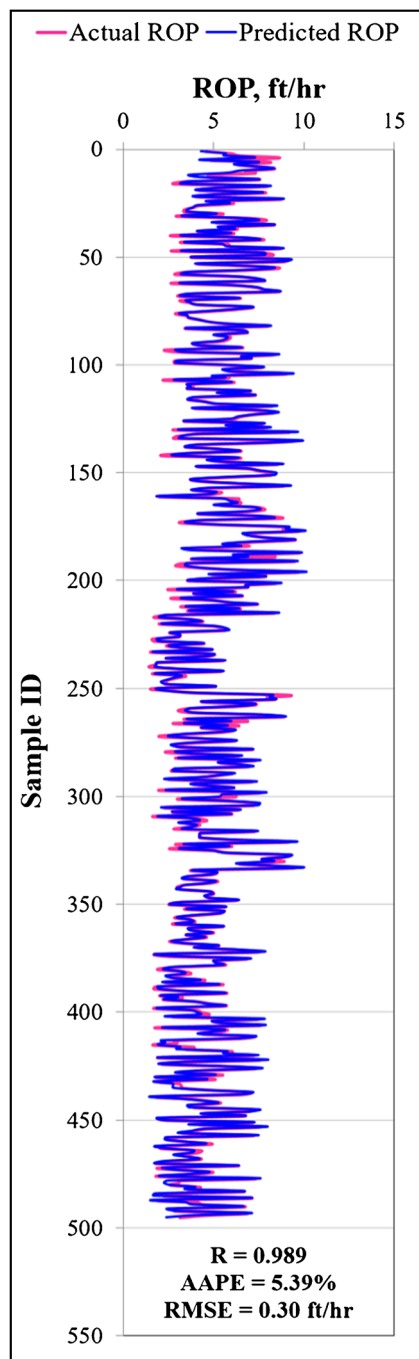
Fig. 10 Comparison of the actual and estimated ROP for the testing datasets collected from Well-A (495 datasets). The predicted ROP values were calculated using the new empirical correlation of Eq. (24). The ROP was predicted accurately with R, AAPE, and RMSE of 0.989, 5.39%, and 0.30 ft/h, respectively

Well-A). As indicated in Fig. 10, Eq. (24) estimated the ROP for the testing data (495 datasets from Well-A) with very low AAPE and RMSE of 5.39% and 0.30 ft/h, respectively. The R between actual and estimated ROP for the testing data is 0.989. Visual check of the real and predicted ROP also indicates the high accuracy of Eq. (24) in estimating the ROP as shown in Fig. 10.

Validating the developed equation for the rate of penetration

The predictability of the optimized ANN-based model developed in this study was evaluated using the 2213 data points collected from Well-B. The predictability of the ANN-based model was also compared to four of the previously available models of Bingham, Maurer, Bourgoyne and Young, and Al-Abduljabbar et al. (2020) model. As shown in Fig. 11, the optimized ANN-based model is most accurate compared with the other models as confirmed by the excellent matching between the actual and estimated ROP. In term of the RMSE, Bingham, Maurer, Bourgoyne and Young, and Al-Abduljabbar et al. model predicted the ROP for Well-B with RMSEs of 1.67, 2.02, 1.29, and 1.39 ft/h, respectively, while the RMSE for the ROP predicted with the ANN-based model is only 0.44 ft/h. All previous models predicted the ROP with a very low R of less than 0.25, while the ANN-based model predicted the ROP with 0.954. As confirmed in Fig. 11 and Fig. 12, all previous models estimated the ROP with very high AAPE and RMSE and low R. The AAPEs for the ROP predicted using Bingham, Maurer, Bourgoyne and Young, and Al-Abduljabbar et al. models are 51.0% 47.57%, 36.64%, and 33.73%, respectively, compared to the AAPE of only 8.85% for the ANN-based model.

These results reflected the accurate predictability of the ANN-based model in predicting the ROP for sandstone formations during the horizontal drilling process based on only the surface measurable drilling parameters.

Conclusions

ANN model was optimized to estimate the ROP in real-time while horizontally drilling through natural gas-bearing sandstone formations based on the Q, DSR, SPP, T, and WOB, in combination with ROP_c, which is a new parameter developed in this study. The ANN-based model was firstly learned and optimized on 1154 data points gathered from Well-A. After that, based on the optimized ANN model, a new empirical equation for ROP estimation was developed. The developed equation was tested on 495 datasets collected from Well-A and validated on the 2213 datasets from Well-B. The predictability of the new ROP equation was compared with the available correlations. The following are concluded out of this study:

a. The ANN-based model estimated the ROP for the training data of Well-A with AAPE and R of 5.11% and 0.991, respectively, when the ROP_c is considered as an input parameter.
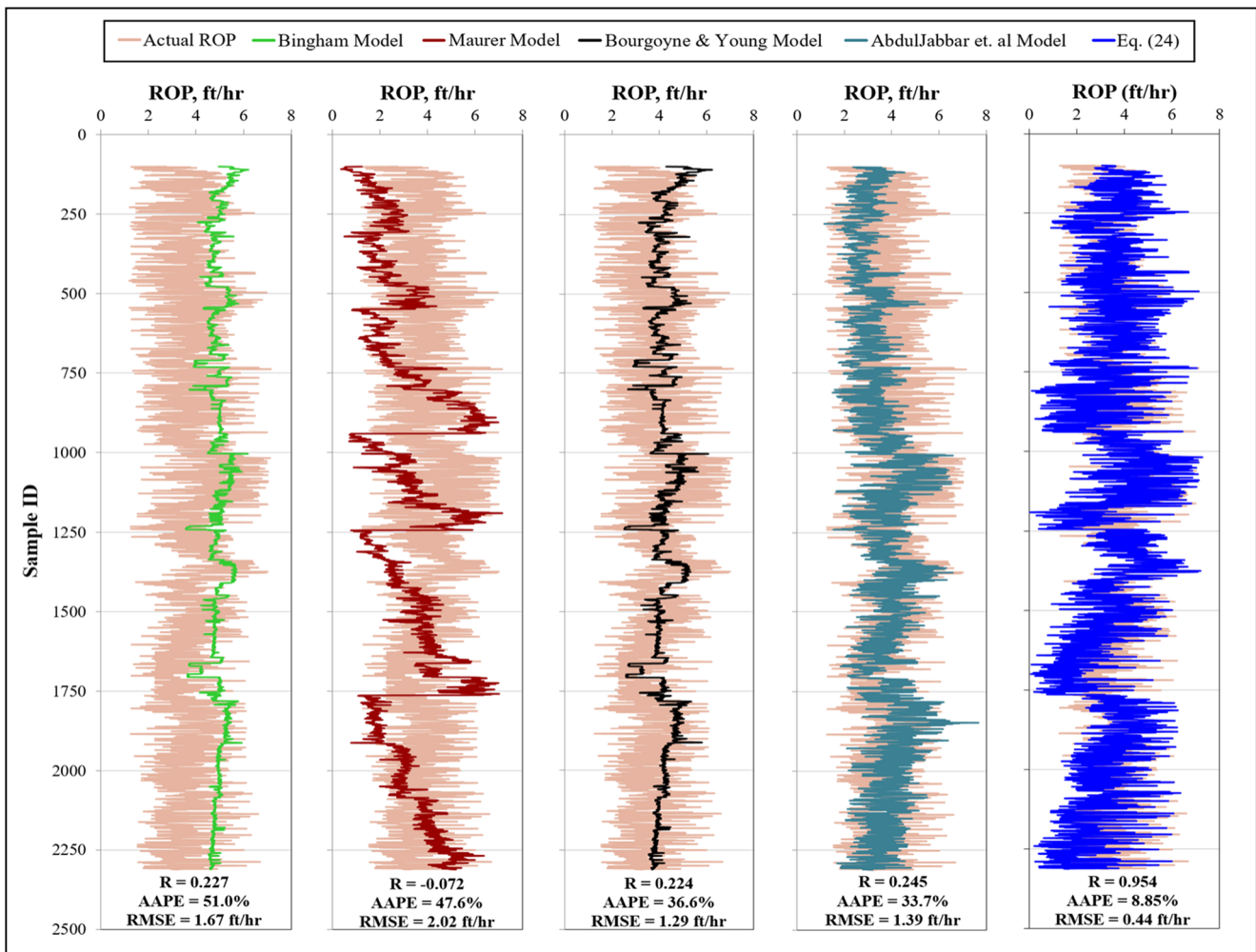
Fig. 11 Comparison of the ROP estimation in Well-B using Bingham model, Maurer model, Bourgoyne and Young model, Al-Abduljabbar et al. model, and the new ROP correlation of Eq. (24) developed in this study. Equation (24) predicted the ROP accurately with the lowest AAPE and RMSE and the highest *R* compared to the previous models
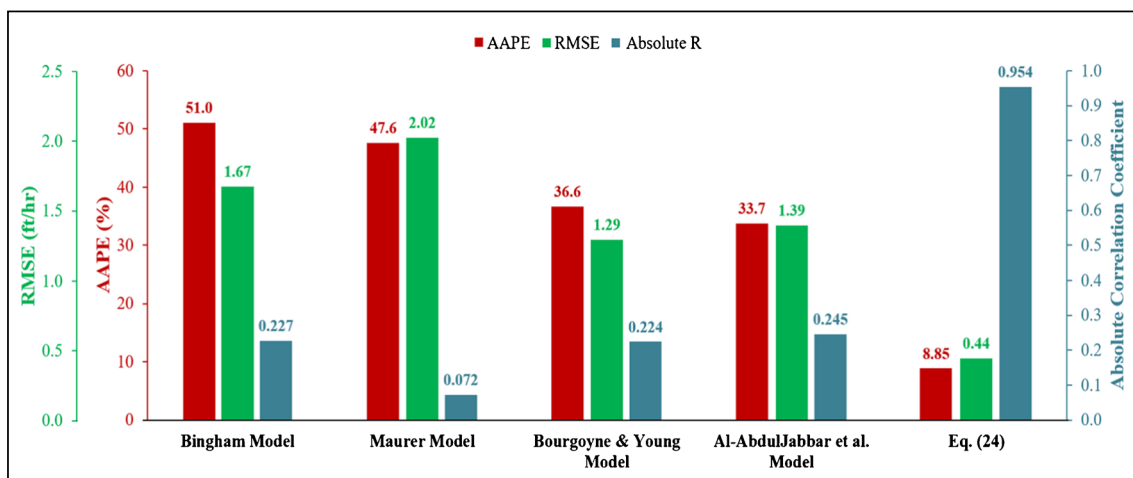

Fig. 12 Comparison of the AAPE, RMSE, and absolute *R* on estimating the ROP for the validation dataset. Equation (24) predicted the ROP accurately with the lowest AAPE and RMSE and the highest *R* compared to the previous models

b. Excluding the ROP_c from the training inputs increased the AAPE for the predicted ROP to 42.6% and reduced R to 0.424.
c. The new empirical equation estimated the ROP for the testing data of Well-A with AAPE and R of 5.39% and 0.989, respectively.
d. For the validation data, the ROP was estimated with AAPE and R of 8.85% and 0.954, respectively, when the new empirical equation developed in this study was used.
e. The optimized ANN-based model overperformed all the available empirical correlations for ROP estimation.

Appendix

Appendix A: Extraction of the coefficients a_1 to a_6 and the constant c of Eq. (23)

To determine the coefficients a_1 to a_6 of Eq. (23), the output layer and training layer weights in Table 3 are transformed into matrices with dimensions of [1, 5] and [5, 6]; multiplication of these matrices will form a matrix with dimensions of [1, 6] as indicated in Eq. (25).

$$
\begin{aligned}
&\begin{pmatrix} 0.5592 & 0.1292 & -0.0163 & 0.6385 & -0.2254 \end{pmatrix} \\
&\times \begin{pmatrix}
-0.0211 & 0.1601 & 0.7169 & -0.3710 & -0.1519 & 1.0517 \\
0.4345 & 0.5034 & -0.0782 & 0.7740 & 0.3895 & 0.6545 \\
0.5855 & 0.0954 & 0.7083 & 0.9679 & 0.7208 & 0.3322 \\
-0.1429 & 0.1952 & -0.8266 & -0.0251 & -1.1259 & 1.3125 \\
-0.1350 & 0.1133 & -0.8087 & -0.5706 & -0.9110 & -0.7650
\end{pmatrix} \\
&= \begin{pmatrix}
-0.0260 \\ 0.2521 \\ 0.0338 \\ -0.0107 \\ -0.5599 \\ 1.6788
\end{pmatrix}
\end{aligned} \tag{25}
$$

The right-hand side matrix of Eq. (25) represents the coefficients a_1 to a_6, where $a_1 = -0.0260$, $a_2 = 0.2521$, $a_3 = 0.0338$, $a_4 = -0.0107$, $a_5 = -0.5599$, and $a_6 = 1.6788$. To extract the constant c, the output layer weight matrix must be multiplied by the training layer bias matrices, which have dimensions of [1, 5] and [1, 5], respectively; then, we add the results to the output layer bias of -0.3896, as explained by Eq. (26).

$$
\begin{aligned}
&\begin{pmatrix} 0.5592 & 0.1292 & -0.0163 & 0.6385 & -0.2254 \end{pmatrix} \\
&\times \begin{pmatrix} 0.5592 \\ 0.1292 \\ -0.0163 \\ 0.6385 \\ -0.2254 \end{pmatrix} + (-0.3896) \\
&= -0.4294
\end{aligned} \tag{26}
$$

References

Abdelgawad K, Elkatatny S, Moussa T, Mahmoud M, Patil S (2018) Real time determination of rheological properties of spud drilling fluids using a hybrid artificial intelligence technique. J Energy Resour Technol 141. https://doi.org/10.1115/1.4042233

Ahmed AS, Mahmoud AA, Elkatatny S (2019a) Fracture pressure prediction using radial basis function. In: Proceedings of the AADE National Technical Conference and exhibition, Denver, CO, USA 9–10 April; AADE-19-NTCE-061

Ahmed AS, Mahmoud AA, Elkatatny S, Mahmoud M, Abdulraheem A (2019b) Prediction of pore and fracture pressures using support vector machine. In: Proceedings of the 2019 international petroleum technology conference, Beijing, China, 26–28 March; IPTC-19523-MS. https://doi.org/10.2523/IPTC-19523-MS

Al-AbdulJabbar A, Elkatatny SM, Mahmoud M, Abdelgawad K, Abdulazeez A (2018) A robust rate of penetration model for carbonate formation. J Energy Resour Technol 141(4):042903–042903-9. https://doi.org/10.1115/1.4041840

Al-AbdulJabbar A, Elkatatny S, Mahmoud AA, Moussa T, Al-Shehri D, Abughaban M, Al-Yami A (2020) Prediction of the rate of penetration while drilling horizontal carbonate reservoirs using the self-adaptive artificial neural networks technique. Sustainability 12(4): 1376. https://doi.org/10.3390/su12041376

Alsaihati A, Elkatatny S, Mahmoud AA, Abdulraheem A (2021) Use of machine learning and data analytics to detect downhole abnormalities while drilling horizontal wells, with real case study. J Energy Resour Technol 143. https://doi.org/10.1115/1.4048070

Amar K, Ibrahim A (2012) Rate of penetration prediction and optimization using advances in artificial neural networks, a comparative study. In: Proceedings of the 4th International Joint Conference on Computational Intelligence, Barcelona, Spain, 5–7 October, pp 647–652. https://doi.org/10.5220/0004172506470652

Arehart RA (1990) Drill-bit diagnosis with neural networks. SPE Comput Appl 2:24–28. https://doi.org/10.2118/19558-PA

Babikir HA, Abd Elaziz M, Elsheikh AH, Showaib EA, Elhadary M, Wu D, Liu Y (2019) Noise prediction of axial piston pump based on

different valve materials using a modified artificial neural network model. Alexandria Eng J 58:1077–1087. https://doi.org/10.1016/j.aej.2019.09.010

Barbosa LFFM, Nascimento A, Mathias MH, de Carvalho JA Jr (2019) Machine learning methods applied to drilling rate of penetration prediction and optimization-a review. J Pet Sci Eng 183:106332. https://doi.org/10.1016/j.petrol.2019.106332

Bilgesu H, Tetrick L, Altmis U, Mohaghegh S, Ameri S (1997) A new approach for the prediction of rate of penetration (ROP) values. In: Proceedings of the SPE Eastern Regional Meeting, Lexington, Kentucky, 22–24 October; SPE-39231-MS. https://doi.org/10.2118/39231-MS

Bingham MG (1965) A new approach to interpreting rock drillability. Petroleum Publishing Co., Tulsa

Bourgoyne A, Young FS (1974) A multiple regression approach to optimal drilling and abnormal pressure detection. Soc Pet Eng J 14:371–384. https://doi.org/10.2118/4238-PA

Elkatatny SM (2017) Real time prediction of rheological parameters of KCl water-based drilling fluid using artificial neural networks. Arab J Sci Eng 42:1655–1665. https://doi.org/10.1007/s13369-016-2409-7

Elkatatny S (2018) New approach to optimize the rate of penetration using artificial neural network. Arab J Sci Eng 43:6297–6304. https://doi.org/10.1007/s13369-017-3022-0

Elkatatny S, Al-AbdulJabbar A, Mahmoud AA (2019) New robust model to estimate the formation tops in real-time using artificial neural networks (ANN). Petrophysics 60:825–837. https://doi.org/10.30632/PJV60N6-2019a7

Eren T, Ozbayoglu ME (2010) Real time optimization of drilling parameters during drilling operations. In: Proceedings of the SPE Oil and Gas India Conference and Exhibition, Mumbai, India, 20–22 January; SPE-129126-MS. https://doi.org/10.2118/129126-MS

Hag Elsafi S (2014) Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Eng J 53:655–662. https://doi.org/10.1016/j.aej.2014.06.010

Hegde C, Daigle H, Millwater H, Gray K (2017) Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. J Pet Sci Eng 159:295–306. https://doi.org/10.1016/j.petrol.2017.09.020

Hossain ME, Al-Majed AA (2015) Fundamentals of sustainable drilling engineering, 1st edn. Wiley-Scrivener, Austin ISBN 10: 0470878177

Lyons WC, Plisga GJ (2004) Standard handbook of petroleum and natural gas. Engineering, 2nd edn. Gulf Professional Publishing, Woburn ISBN 10: 0750677856

Mahmoud AA, Elkatatny S, Abdulraheem A, Mahmoud M, Ibrahim O, Ali A (2017a) New technique to determine the total organic carbon based on well logs using artificial neural network (white box). In: Proceedings of the 2017 SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 24–27 April; SPE-188016-MS. https://doi.org/10.2118/188016-MS

Mahmoud AA, Elkatatny S, Mahmoud M, Abouelresh M, Abdulraheem A, Ali A (2017b) Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network. Int J Coal Geol 179:72–80. https://doi.org/10.1016/j.coal.2017.05.012

Mahmoud AA, Elkatatny S, Abdulraheem A, Mahmoud M (2017c) Application of artificial intelligence techniques in estimating oil recovery factor for water drive Sandy reservoirs. In: Proceedings of the 2017 SPE Kuwait Oil & Gas Show and Conference, Kuwait City, Kuwait, 15–18 October; SPE-187621-MS. https://doi.org/10.2118/187621-MS

Mahmoud AA, Elkatatny S, Ali A, Abouelresh M, Abdulraheem A (2019a) Evaluation of the total organic carbon (TOC) using different artificial intelligence techniques. Sustainability 11:5643. https://doi.org/10.3390/su11205643

Mahmoud AA, Elkatatny S, Ali A, Abouelresh M, Abdulraheem A (2019b) New robust model to evaluate the total organic carbon using fuzzy logic. In: Proceedings of the SPE Kuwait Oil & Gas Show and Conference, Mishref, Kuwait, 13–16 October; SPE-198130-MS. https://doi.org/10.2118/198130-MS

Mahmoud AA, Elkatatny S, Chen W, Abdulraheem A (2019c) Estimation of oil recovery factor for water drive Sandy reservoirs through applications of artificial intelligence. Energies 12:3671. https://doi.org/10.3390/en12193671

Mahmoud AA, Elkatatny S, Ali A, Moussa T (2019d) Estimation of static Young's modulus for sandstone formation using artificial neural networks. Energies 12:2125. https://doi.org/10.3390/en12112125

Mahmoud AA, Elzenary M, Elkatatny S (2020a) New hybrid hole cleaning model for vertical and deviated wells. J Energy Resour Technol 142:034501. https://doi.org/10.1115/1.4045169

Mahmoud AA, Elkatatny S, Abouelresh M, Abdulraheem A, Ali A (2020b) Estimation of the total organic carbon using functional neural networks and support vector machine. In: Proceedings of the 12th International Petroleum Technology Conference and Exhibition, Dhahran, Saudi Arabia, 13–15 January; IPTC-19659-MS. https://doi.org/10.2523/IPTC-19659-MS

Mahmoud AA, Elkatatny S, Ali A, Moussa T (2020c) A self-adaptive artificial neural network technique to estimate static Young's modulus based on well logs. In: Proceedings of the Oman Petroleum & Energy Show, Muscat, Oman, 9–11 March; SPE-200139-MS. https://doi.org/10.2118/200139-MS

Mahmoud AA, Elkatatny S, Al-Shehri D (2020d) Application of machine learning in evaluation of the static Young's modulus for sandstone formations. Sustainability 12(5). https://doi.org/10.3390/su12051880

Mahmoud AA, Elkatatny S, Alsabaa A, Al Shehri D (2020e) Functional neural networks-based model for prediction of the static Young's modulus for sandstone formations. In: Proceedings of the 54rd US Rock Mechanics/Geomechanics Symposium, 28 June - 1 July

Mahmoud AA, Elkatatny S, Abduljabbar A, Moussa T, Gamal H, Al Shehri D (2020f) Artificial neural networks model for prediction of the rate of penetration while horizontally drilling carbonate formations. In: Proceedings of the 54rd US Rock Mechanics/Geomechanics Symposium 28 June - 1 July

Maurer W (1962) The "perfect-cleaning" theory of rotary drilling. J Pet Technol 14:1270–1274. https://doi.org/10.2118/408-PA

Mitchell RF, Miska SZ (2011) Fundamentals of drilling engineering. Society of Petroleum Engineers, Richardson ASIN: B01L0O8WJA

Osgouei RE (2007) Rate of penetration estimation model for directional and horizontal wells. Master's Thesis. Middle East Technical University, Turkish

Payette GS, Spivey BJ, Wang L, Bailey JR, Sanderson D, Kong R, Pawson M, Eddy A (2017) Real-time well-site based surveillance and optimization platform for drilling: technology, basic workflows and field results. In: Proceedings of the SPE/IADC Drilling Conference and Exhibition, The Hague, The Netherlands, 14–16 March; SPE-184615-MS. https://doi.org/10.2118/184615-MS

Ren X, Hou J, Song S, Liu Y, Chen D, Wang X, Dou L (2019) Lithology identification using well logs: a method by integrating artificial neural networks and sedimentary patterns. J Pet Sci Eng 182:106336. https://doi.org/10.1016/j.petrol.2019.106336

Soares C, Gray K (2019) Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. J Pet Sci Eng 172:934–959. https://doi.org/10.1016/j.petrol.2018.08.083

Soares C, Daigle H, Gray K (2016) Evaluation of PDC bit ROP models and the effect of rock strength on model coefficients. J Nat Gas Sci Eng 34:1225–1236. https://doi.org/10.1016/j.jngse.2016.08.012

Teale R (1965) The concept of specific energy in rock drilling. Int J Rock Mechan Min Sci Geomech 2(1):57–73. https://doi.org/10.1016/0148-9062(65)90022-7