

Research Article

A Novel Artificial Neural Network-Based Correlation for Evaluating the Rate of Penetration in a Natural Gas Bearing Sandstone Formation: A Case Study in a Middle East Oil Field

Ahmad Al-AbdulJabbar,¹ Ahmed Abdulhamid Mahmoud,² Salaheldin Elkatatny ,² and Mahmoud Abughaban¹

¹EXPEC Advanced Research Center, Saudi Aramco, Dhahran 31311, Saudi Arabia

²College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, 31261 Dhahran, Saudi Arabia

Correspondence should be addressed to Salaheldin Elkatatny; elkatatny@kfupm.edu.sa

Received 10 July 2021; Revised 22 November 2021; Accepted 21 December 2021; Published 13 January 2022

Academic Editor: Carlos Michel

Copyright © 2022 Ahmad Al-AbdulJabbar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study presented an empirical correlation to estimate the drilling rate of penetration (ROP) while drilling into a sandstone formation. The equation developed in this study was based on the artificial neural networks (ANN) which was learned to assess the ROP from the drilling mechanical parameters. The ANN model was trained on 630 datapoints collected from five different wells; the suggested equation was then tested on 270 datapoints from the same training wells and then validated on three other wells. The results showed that, for the training data, the learned ANN model predicted the ROP with an AAPE of 7.5%. The extracted equation was tested on data gathered from the same training wells where it estimated the ROP with AAPE of 8.1%. The equation was then validated on three wells, and it determined the ROP with AAPEs of 9.0%, 10.7%, and 8.9% in Well-A, Well-B, and Well-D, respectively. Compared with the available empirical equations, the equation developed in this study was most accurate in estimating the ROP.

1. Introduction

Evaluation of the formation drillability is a critical process that is highly dependent on the speed at which the drillbit will be able to drill through the formation or what is called the rate of penetration (ROP) [1]. Estimation and optimizing of the ROP are important since drilling with high ROP could significantly decrease the drilling cost, but on the other hand, the significant increase of the ROP could in many cases lead to several problems such as hole cleaning problems and increasing the drillstring vibration, which could lead to an increase in the nondrilling time and, therefore, raising the drilling cost [2, 3].

Optimization of the ROP requires manipulation with many controllable and uncontrollable parameters. Alteration of uncontrollable parameters like the drilling fluid types or drill bit size is costly; in addition, modification of

any of these parameters affects the others, which complicate predictability of how modification of a single parameter contributes to the change in ROP [4, 5].

Originally, different traditional models were optimized to evaluate the ROP; these models were developed from regression analysis to assess the ROP based on various inputs, the accuracy of these models is significantly affected by the inputs considered [6–8].

The first regression model for ROP predicted was suggested by Maurer [9]; the author developed this model to estimate the ROP for the tricone bit based on the DSR, WOB, and the drill bit size only. The main limitation of this model is that it was developed based on the assumption that the drilled cutting will be lifted through the wellbore directly after the rock is touched by the drilling bit tooth. Later, Bingham [10] conducted several laboratory experiments; based on the results of these experiments, he suggested another ROP model which

TABLE 1: Some of the available ROP correlations.

Reference	Method	Input parameters	Formula
Maurer [9]	Empirical	DSR, unconfined compressive strength (UCS), bit size, and WOB	$R = k \frac{NW^2}{D^2 S^2}$
Bingham [10]	Empirical	Bit size, WOB, and DSR	$R = k \left(\frac{W}{d_b} \right)^{a_s} N^c$
Bourgoyne and Young [11]	Empirical	TVD, WOB, DSR, Q, MW, pore pressure, TFA, and bit wear	$\frac{d}{dt}(R) = e \left(a_1 + \sum_{i=2}^8 a_i x_i \right)$
Warren [12]	Empirical	Bit size, UCS, WOB, and DSR	$R = \left[\frac{aS^2 d_b^2}{N^b W^2} + \frac{c}{N d_b} \right]^{-1}$
Osgouei [5]	Empirical	TVD, MW, WOB, pore pressure, DSR, TFA, Q, wellbore inclination, bit wear, and its type	$\frac{d}{dt}(R) = e \left(a_1 + \sum_{i=2}^{11} a_i x_i \right)$
Al-AbdulJabbar [13]	Empirical	UCS, DSR, PV, Q, WOB, bit size, torque, SPP, and the MW	$ROP = \frac{WOB^a \times RPM \times T \times SPP \times GPM}{d_b^2 \times \rho \times PV \times UCS^b}$

defined the ROP as a function of only WOB and DSR, and it neglected the WOB threshold value.

In 1974, Bourgoyne and Young [11] developed another regression-based model which evaluates the ROP considering the effect of most of the mechanical and physical parameters influencing the drilling process; the effect of these parameters is evaluated using different exponents which are combined to form the full model. Another model to assess the ROP while drilling with tricone bit was developed by Warren [12]; this model evaluates the ROP by considering the optimum conditions to ensure optimum hole cleaning condition which satisfies the requirement that the rate of cutting generation is the same as the rate of cuttings lifting from the wellbore. Recently, Al-AbdulJabbar [13] developed a regression-based ROP model which could estimate the ROP from the drilling fluid properties in addition to the drilling hydraulic and mechanical parameters. Table 1 summarizes all empirical equations developed based on regression analysis to assess the ROP.

With the recent advances in machine learning, several researchers investigated the possibility of applying different data-driven models to evaluate various parameters required in different aspects related to petroleum engineering [14–16]. Artificial neural network (ANN) is the most common technique applied in the petroleum industry, and it approved high performance in evaluating several parameters [17, 18]; other machine learning techniques were also successfully applied in the petroleum industry such as support vector regression (SVR) [19], adaptive network-based fuzzy inference system [20], functional neural networks (FNN) [21], and random forest (RF) [22].

Bilgesu et al. [23] suggested the application of machine learning for ROP prediction. They developed two models based on the ANN to evaluate the ROP in nine formations. Bilgesu et al. [23] optimized their first model to estimate the ROP from the type of the formation, WOB, footage, DSR, drilled formation, drill bit's type, diameter, tooth, bear-

ing wear, mud circulation, and gross hours of drilling (GHD). For their second model, the authors did not include the bit tooth and bearing wear from the inputs. As reported by Bilgesu et al. [23], both models accurately estimated the ROP.

ANN model was also optimized by Bataee and Mohseni [24] to estimate the ROP based on the bit diameter, depth, WOB, DSR, and mud weight; the data considered in this work were collected from Shadegan oil field. From this study, the authors were able to establish the concepts of the effect of each input parameter on ROP prediction, and no empirical equation was derived from the learned ANN model.

In 2018, another ROP model was developed using ANN; this model assessed the ROP from the drilling fluid flowrate (Q), plastic viscosity, and density, in addition to the stand-pipe pressure (SPP), DSR, WOB, torque, and UCS. This model enabled the evaluation of the ROP accurately with an AAPE of 4%.

Later, another ANN-based model was developed by Al-AbdulJabbar et al. [25] to predict the ROP from the inputs used by Elkatatny [26] excluding the drilling fluid plastic viscosity and density to allow for real-time prediction. The optimized model was validated into two wells, where the results showed that ROP was evaluated with a correlation coefficient (R) of higher than 0.94.

Ahmed et al. [27] suggested the use of the support vector regression (SVR) for evaluation of the ROP; they optimized the SVR using ten parameters including drilling fluid properties and drilling mechanical parameters. Validation of this model showed that it predicted the ROP with an AAPE of only 2.83%. After that, another SVR-based model was developed by Gan et al. [28] to predict the ROP from the WOB, DSR, Q, SPP, and the drilling torque. The authors reported that their developed model is not universal, and their suggested modeling process needs to be repeated whenever the model is to be applied in a different area.

TABLE 2: Some of the developed data-driven models for ROP prediction.

Reference	Machine learning models	Input parameters	Remarks
Bilgesu et al. [23]	ANN	Type of the formation, WOB, footage, DSR, drilled formation, drill bit's type, diameter, tooth, and bearing wear, mud circulation, and GHD	—
Bataee and Mohseni [24]	ANN	Bit diameter, depth, WOB, DSR, and mud weight	—
Elkatatny [26]	ANN	Q, MW, PV, SPP, torque, WOB, and DSR	AAPE of 4.0%
Al-AbdulJabbar et al. [25]	ANN	WOB, DSR, SPP, Q, torque	R of 0.96
Ahmed et al. [27]	SVR	Q, MFV, MW, SPP, WOB, DSR, torque, SV, YP, and PV	AAPE is 2.83%
Gan et al. [28]	SVR	WOB, DSR, Q, SPP, and the drilling torque	—
Oyedere and Gray [29]	LDA, QDA, SVR, and RF	WOB, DSR, Q, and UCS	—
Al-AbdulJabbar et al. [30]	ANN	DSR, WOB, torque, RHOB, GR, and DR	R of 0.96
Al-AbdulJabbar et al. [31]	ANN	Q, torque, DSR, WOB, and SSP	R of 0.954, AAPE of 8.85%, and RMSE of 0.44 ft/hr
Alali et al. [32]	ANN	Q, DSR, and WOB	—

In 2020, other five models for optimizing the ROP were developed by Oyedere and Gray [29]; these models were developed using logistic regression, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), SVR, and RF. These models were developed based on five machine learning classification tools to optimize the ROP based on the WOB, DSR, Q, and UCS. Although these models showed good results in predicting the ROP, not all the most influential parameters on ROP are considered to develop these models, where they did not consider the effect of the torque and SPP. Besides, these models are still black box, and no empirical equation was extracted for future use and validation by the readers.

The self-adaptive differential evolution algorithm was applied by Al-AbdulJabbar et al. [30] to optimize the performance of the ANN model for assessment of the ROP during horizontal drilling of carbonate reservoir. The inputs considered to optimize this model include petrophysical and drilling mechanical parameters. The results showed high accuracy for this model which enabled evaluation of the ROIP with an R of 0.96.

In another study, Al-AbdulJabbar et al. [31] learned the ANN to determine the ROP in real-time during horizontally drilling a natural gas-bearing sandstone formation. To allow for ROP prediction in real-time, the ANN model was learned on real-time measurable parameters of Q, DSR, torque, WOB, and SSP. The optimized model was converted into empirical correlation which evaluated the ROP in real-time with R of 0.954, AAPE of 8.85%, and RMSE of 0.44 ft/hr for the validation data.

Recently, Alali et al. [32] proposed a two-phase integrated and data-driven ROP optimization system. The authors developed a heatmap function for identifying the optimal ROP based on the Q, DSR, and WOB; they also extracted an equation from their optimized ANN model for predicting the ROP in real-time. The main limitation for the developed models is that they did not consider most of the parameters influencing the ROP such as the drilling torque and the SPP. Table 2 compares the inputs and the accuracy for some of the developed machine learning models for ROP estimation.

TABLE 3: The ROP duplicates along the drilled wellbore with the increase in the depth.

Depth	ROP (ft/h)	SPP (psi)	Torque (kft.lbf)	WOB (klbf)
XXX1.39	4.3253	2863.6	4.22	19.99
XXX1.42	4.3253	2865.1	5.17	32.78
XXX1.52	12.0027	2860.2	4.49	28.75
XXX1.55	12.0027	2858.9	4.15	25.55
XXX1.59	12.0027	2863.5	3.95	23.91
XXX1.60	12.0027	2858.6	3.68	22.72
XXX2.31	12.1049	2859.0	3.68	21.19
XXX2.32	12.1049	2861.2	3.34	20.44
XXX2.32	12.1049	2862.6	3.34	19.53
XXX2.33	12.1049	2853.1	3.27	17.96
XXX3.45	12.5975	2855.2	3.68	21.55
XXX3.46	12.5975	2856.6	4.15	26.21

The goal of this work is to develop an equation that could be used to estimate the ROP into a vertical sandstone formation from the drilling data such as WOB, DSR, torque, Q, and SPP recorded on time. Conversion of the optimized ANN model to an empirical equation, which is a function of the drilling data only, will allow for real-time prediction.

2. Methodology

For the purpose of this study, drilling mechanical parameters of the WOB, DSR, torque, Q, and SPP gathered from eight wells drilled in the same reservoir were considered to learn the ANN model. All inputs are possible to obtain in real-time from the surface sensors. To ensure that only valid and representative data are used, all inputs were preprocessed through data quality assurance (QA), quality control (QC), and data analytics processes.

2.1. Data Preprocessing. Before preprocessing the data, the input data and their corresponding ROP collected from some wells exceeded 50,000 datasets. Firstly, based on

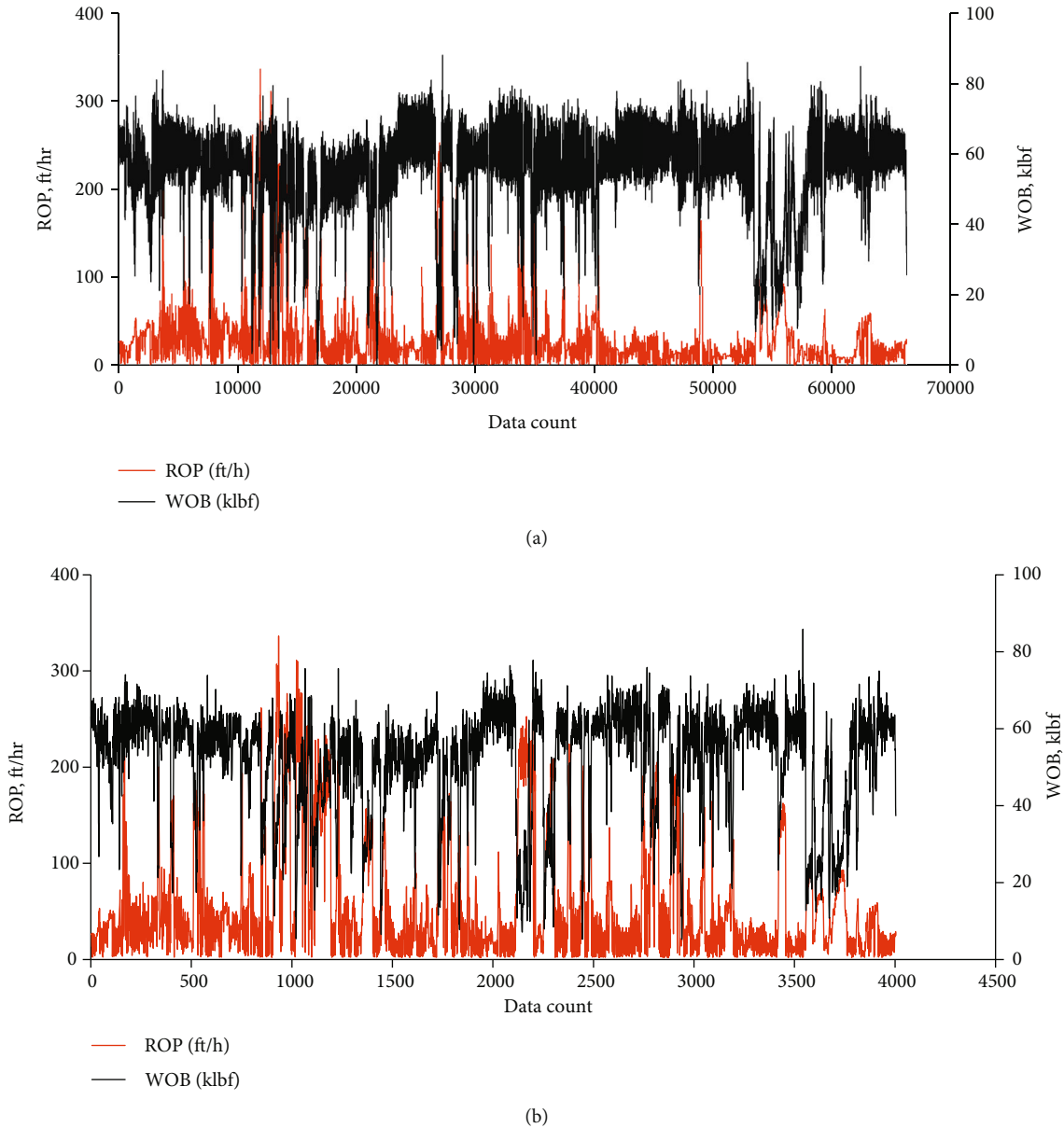


FIGURE 1: The relationship between the ROP and WOB for Well Well-C data (a) before and (b) after QA/QC processing and duplicates removal.

visualizing the data, all anomalies were picked up and removed from the data. In the second stage, minimum limits were set on the training data, all values below these limits are considered not valid for ROP optimization since they are very low (not economical to be used), for example, all ROP values less than 2 ft/hr were excluded from the training data.

It is noted that the data is dominated by many duplicates. As indicated in Table 3, at several depths, the same ROP values are repeated, and these are corresponding to different values of other inputs; this could lead to confusion during the optimization process; therefore, the next stage is to remove all duplicates from the training data.

After performing the previous processes on the training data, these dataset counts were reduced to about 5,000 data points as indicated in Figure 1.

Now, the changes in the ROP as a function of the wellbore depths were investigated; as shown in Table 4, there is a considerable change in the ROP within a small change in the drilled depth which is not normal; this is attributed to sensor fall, bit whirl, or most probably to the stick or slip problems which usually associated with horizontal drilling; therefore, all the data recorded during that time must be removed from the input data. As indicated in Figure 2, the problem of stick and slip also affected the DSR measurement, where the recorded values experienced a considerable variation and fluctuation.

2.2. Using the Confined Compressive Strength to Cap the MSE. The amount of energy needed by the drilling bit to crush the rock is represented by the mechanical specific energy (MSE) [33]. Therefore, optimization of the MSE is

TABLE 4: An extreme variation of the ROP as a result of the stick and slip.

Depth	ROP (ft/h)
XXX1.394	36.15
XXX1.569	63.13
XXX2.092	126.68
XXX2.161	26.83
XXX2.161	13.35
XXX2.559	29.64
XXX2.93	146.51
.	
.	
.	
XXX2.687	99.59
XXX3.147	120.41
XXX3.713	122.39
XXX4.606	135.35
XXX4.973	102.03
XXX5.075	27.09
XXX5.33	69.51
XXX5.74	102.28
XXX5.931	15.07
XXX6.26	59.13

important especially for horizontal drilling since applying high MSE may lead to losing that energy through vibration; on the other hand, applying low MSE will not be enough to optimize the drilling process [34]. According to Teale [35], the required MSE is almost equal to the formation confined compressive strength.

The sandstone formation considered in this study has confined compressive strength between 25,000 to 35,000 psi, and hence, the efficient drilling is possible when the applied MSE is in the range of 25,000 and 35,000 psi; therefore, the training data was filtered again to neglect all inputs with MSE outside the range required for efficient drilling as explained in Figure 3.

To confirm the dependence of the drilling efficiency on the MSE and the relationship between the different inputs and the ROP, R between the WOB and the ROP was determined at various MSE ranges as compared in Table 5. The results showed that the highest correlation between the WOB and ROP was obtained when the MSE was within the range of 25,000 and 35,000 psi.

2.3. Optimizing the Artificial Neural Network Model. The ANN model was trained to predict the ROP using the different input drilling mechanical parameters of the DSR, WOB, SPP, Q , and torque. The training data was gathered from Well-C, Well-E, Well-F, Well-G, and Well-H, and it consists of 630 datasets which is 70% of the total data collected from the training wells (five wells) after preprocessing (897 datasets). Figure 4 shows the data collected from the training wells; these data were utilized to train and test the ANN model.

The training data (630 datasets collected from the training data) was investigated to evaluate their statistical features; this is because after learning the ANN and to ensure high accuracy when it comes to implementing the optimized ANN on new data, it is very important to ensure that this data is with similar statistical characteristics as the training data. The statistical features of the training dataset are listed in Table 6. From this table, the Q is varying between 843 to 1110 gpm, DSR is in the range between 80 and 127 rpm, the SPP is from 1832 to 3277 psi, torque is between 2.9 and 13.1 klb_p, WOB is from 5.3 to 82.3 klb_p and ROP is between 14.8 and 93.3 ft/hr.

A feed-forward network was considered to build the model, where the optimum design parameters of the model were selected based on sensitivity analysis. During this sensitivity analysis, the effect of using a single, two, or three training layers and associated neurons of 3 to 30 per layer on the performance of the ANN model was examined. The performance of different training and transferring functions and the optimum number of the inputs to predict the ROP was also assessed. The results of this analysis were reported in Tables 7–10.

As indicated in Table 7, the use of a single training layer with 5 neurons optimized ROP prediction with AAPE of 7.5%; increasing the number of the training layers did not lead to a decrease in the AAPE. The use of a single layer with only five neurons is important to simplify the matrix of the extracted weights and biases to be used for developing an empirical correlation out of the optimized ANN model.

Out of nine training layers considered for the sensitivity analysis, the performance of the Levenberg-Marquardt (trainlm) function was the best for ROP evaluation with the lowest AAPE of 7.5% as shown in Table 8.

The performance of three transferring was studied in this stage; as shown from Table 9, the use of the pure linear function leads to the optimize the ANN model performance where the ROP was predicted with an AAPE of 7.5%, compared with AAPEs of 7.6% and 7.7% where the tangential sigmoid and log sigmoid functions were used, respectively.

To determine the optimum inputs to be considered, the accuracy of the ANN model in estimating the ROP was compared after excluding everyone of the training inputs and considering the others to predict the ROP. The results of Table 10 indicated that the best performance for the ANN model was when all inputs were used to predict the ROP where the AAPE is 7.5%, excluding the different inputs leads to increasing the AAPE.

Regarding the number of the inputs, the results indicated that for improving the ANN model for ROP prediction, the use of all the five drilling mechanical parameters as inputs is a must, the ANN model should consist of a single training layer associated with five neurons, and the training process will be conducted using trainlm while the data will data transferred from the training layer to the output layer using the pure linear function. Table 11 summarizes the optimum design parameters for the optimized ANN model, and Figure 5 shows a schematic of this optimized model.

2.4. Developing the Empirical Equation for ROP Estimation. The optimized ANN model of Table 11 and Figure 5 was

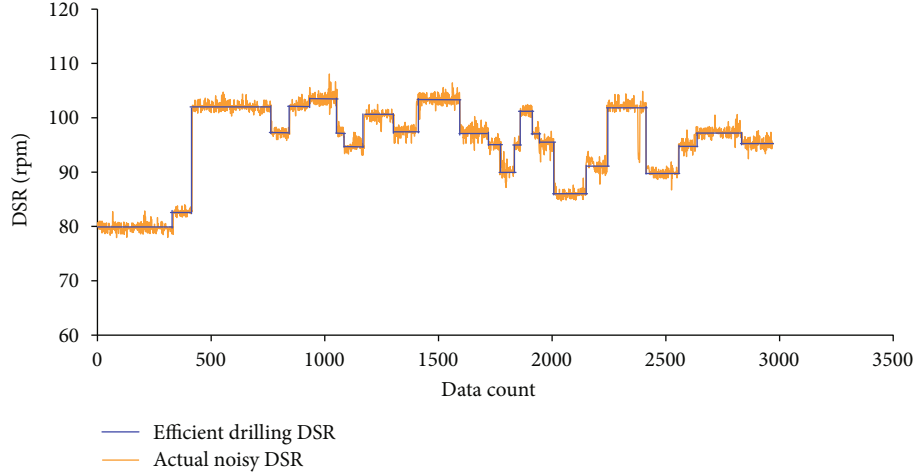


FIGURE 2: Example of the fluctuation of the DSR caused by the stick and slip.

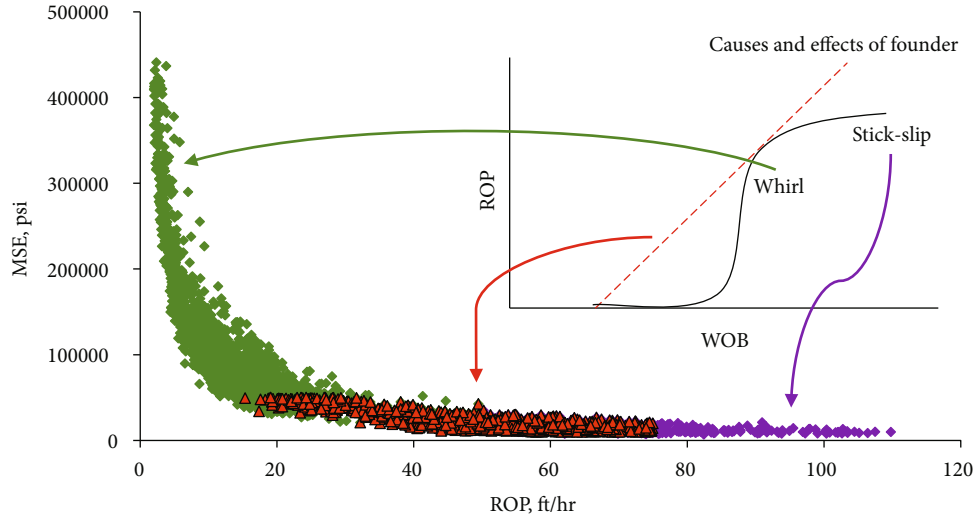


FIGURE 3: The ROP and MSE (including 25000-35000 psi) with wasted energy example.

TABLE 5: The correlation coefficient between the WOB and ROP for Well-B data, at various ranges of the MSE.

MSE range	R
Entire data	-0.61 (negative)
10,000–35,000	-0.24 (negative)
10,000–20,000	0.22
20,000–25,000	0.14
25,000–35,000	0.31

then to be converted into an empirical correlation; this correlation is based on the parameters listed in Table 11 and the extracted weights and biases obtained from the optimized ANN. From Table 11, the optimized ANN model was built using the pure linear transferring function; the generalized form of the empirical correlation representing this model is in

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

where y is the targeted parameter of the output, w denotes the extracted weights, x is the input parameters, and b represents the extracted biases.

Since the optimized ANN model has five inputs and five neurons, substituting these parameters back into Equation (1), it will take the form shown in

$$\text{ROP} = \left(\sum_{j=1}^5 w_{j1} \left(\sum_{i=1}^5 w_{ij} x_i + b_j \right) + b_1 \right), \quad (2)$$

where the extracted w and b are listed in Table 12.

Expanding Equation (2) will lead to getting

$$\text{ROP} = a_1 \text{GPM} + a_2 \text{RPM} + a_3 \text{SPP} + a_4 \text{Torque} + a_5 \text{WOB} + c. \quad (3)$$

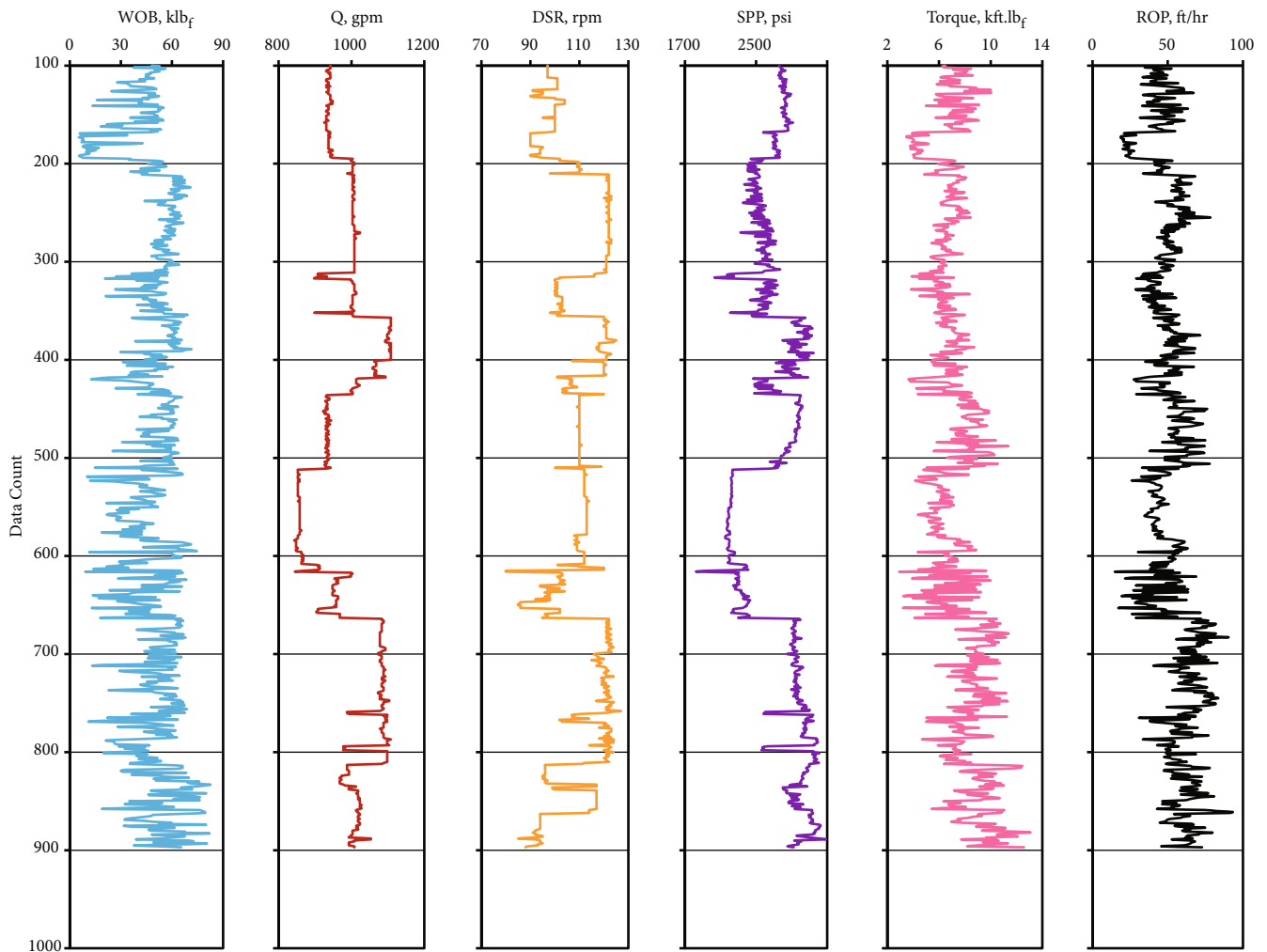


FIGURE 4: The data combination of the data collected from the training wells; these data were utilized to train and test the ANN model.

TABLE 6: The statistical feature of the training data.

	Q (gpm)	DSR (rpm)	SPP (psi)	Torque (kft.lb _f)	WOB (klb _f)	ROP (ft/hr)
Minimum	843	80	1832	2.9	5.3	14.8
Maximum	1110	127	3277	13.1	82.3	93.3
Range	267	47	1445	10.2	77.0	78.5
Standard deviation	76.31	10.7	289.5	1.71	14.4	13.2
Sample variance	5829	114.5	83924	2.93	208.7	175.4

The coefficients a_1 to a_5 needed for Equation (3) could be evaluated as a function of the weights associated with training and output layers, while the constant c was calculated as a function of the matrix of the output layer weights, the matrix of the training layer biases, and the output layer bias as explained in the Appendix.

Substituting for the coefficients a_1 to a_5 and the constant c obtained in the Appendix into Equation (3) leads to the final ROP equation in

$$\text{ROP} = 0.0049 \text{ GPM} + 0.2292 \text{ RPM} + 0.0083 \text{ SPP} - 0.7577 T - 0.0884 \text{ WOB} - 0.0341. \quad (4)$$

2.5. Testing and Validating the Developed Equation. 270 datasets of the data collected from the five training wells were considered to test Equation (4); the testing data represents 30% of the data collected from the training wells. Other unseen data of 280 datapoints from Well-A, 272 data points from Well-B, and 428 datapoints from Well-D were considered to validate in Equation (4).

2.6. Comparing the Performance of Equation (4) to Available Equations. The performance of four previously available equations in estimating the ROP for the 280 datasets of Well-A was compared with that of Equation (4); the four equations compared with Equation (4) were developed by

TABLE 7: Sensitivity analysis for the number of training layers and neurons.

No. of layers	No. of neurons	R	AAPE (%)
1	5	0.934 ¹	7.5 ¹
2	5	0.933	7.6
2	10	0.932	7.7
3	5	0.933	7.6
3	10	0.931	7.9

¹Optimum result.

TABLE 8: Selection of the training function.

Training function	R	AAPE (%)
trainlm	0.934 ¹	7.5 ¹
trainbfg	0.933	7.7
trainrp	0.932	7.9
trainscg	0.933	7.7
traincgb	0.932	7.8
traincgf	0.933	7.7
traincgp	0.931	7.9
trainoss	0.933	7.7
traingdx	0.933	7.7

¹Optimum result.

TABLE 9: Selection of the transferring functions.

Transfer function	R	AAPE (%)
Tangential sigmoid	0.933	7.6 ¹
Log sigmoid	0.932	7.7
Pure linear	0.934 ¹	7.5 ¹

¹Optimum result.

TABLE 10: Selection of the number of inputs.

Inputs	R	AAPE (%)
All	0.934 ¹	7.5 ¹
No Q	0.932	7.7
No DSR	0.90	9.6
No SPP	0.932	7.7
No torque	0.78	13.6
No WOB	0.93	7.9

¹Optimum result.

Maurer [9], Bingham [10], Bourgoyne and Young [11], and Al-Abduljabbar [13] which are listed in Table 1.

3. Results and Discussion

3.1. Training the ANN Model. Figure 6 compares the actual and estimated ROP for the training data collected from the five training wells (630 datasets). The result in Figure 6 indicates the high accuracy of the trained ANN model in predicting the ROP with a high correlation coefficient (R) of 0.94 and an average absolute percentage error (AAPE) of

TABLE 11: The optimized design parameters of the ANN model.

Parameter	Value
Training function	Levenberg-Marquardt
Transferring function	Pure linear
Learning layers (neurons)	1 (5)

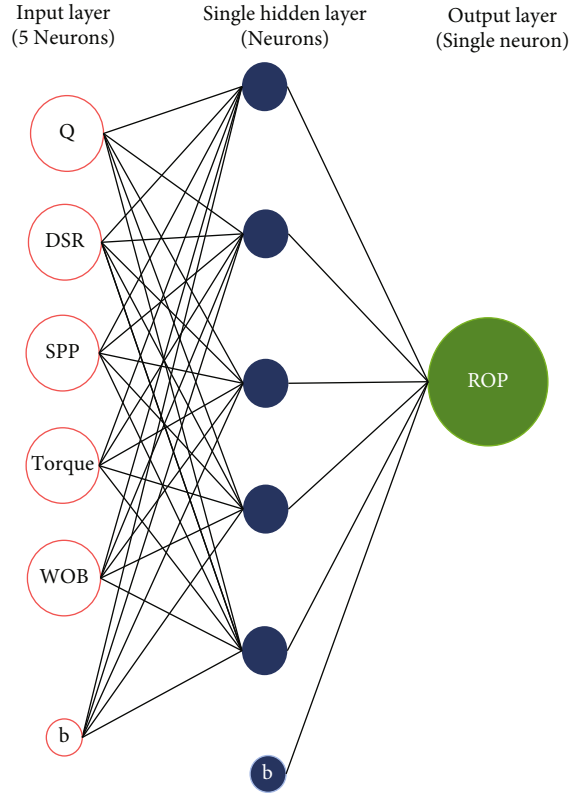


FIGURE 5: Schematic of the optimized ANN model.

7.5%. Comparing the plots of actual and estimated ROP also indicated the high machining between the two plots which confirmed the high accuracy of the ANN model.

3.2. Testing the Developed ROP Equation. Equation (4) was tested on 270 datasets collected from the same training wells. As indicated in Figure 7, the ROP for the testing data was predicted with high accuracy using Equation (4), as confirmed by the good matching of the predicted ROP with the actual ROP curve, and the high R of 0.93 and AAPE of 8.1%.

3.3. Validating the Developed ROP Equation. Now, Equation (4) is ready to be validated on unseen data to confirm its ability to be applied for future assessment of the ROP. The validation data collected from Well-A, Well-B, and Well-D were only processed to remove all data with MSE outside the range of 25,000-35,000 psi; QA/QC processes were not performed on this to evaluate if Equation (4) could be applied for future predictions on the fly.

The results of the validation are compared in Figure 8 which confirmed the high accuracy of Equation (4). ROP was predicted in Well-A with R and AAPE of 0.95 and

TABLE 12: w and b required for Equation (2).

No. of neurons	Training layer					Output layer		
	Weight (w_1)					Biases (b_1)	Weights (w_2)	Bias (b_2)
	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$			
$j = 1$	0.718	-0.740	-0.593	-0.594	0.310	0.476	-0.016	0.564
$j = 2$	0.077	-0.408	0.556	-1.132	-0.777	0.582	-0.407	
$j = 3$	-0.947	0.059	0.778	0.565	0.728	-0.175	-0.095	
$j = 4$	0.276	0.893	-0.351	-0.148	0.969	0.399	-0.319	
$j = 5$	0.107	0.797	0.436	0.686	0.362	-0.566	0.429	

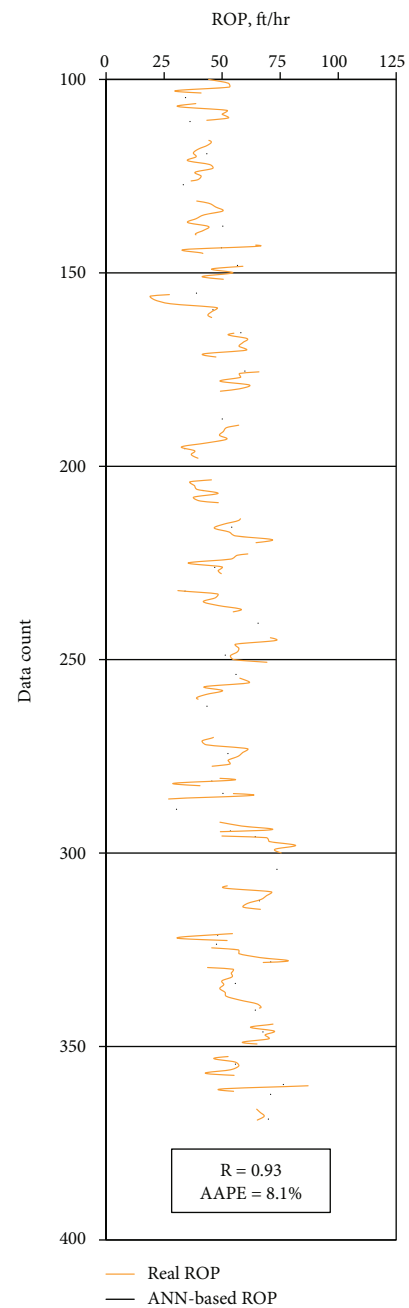
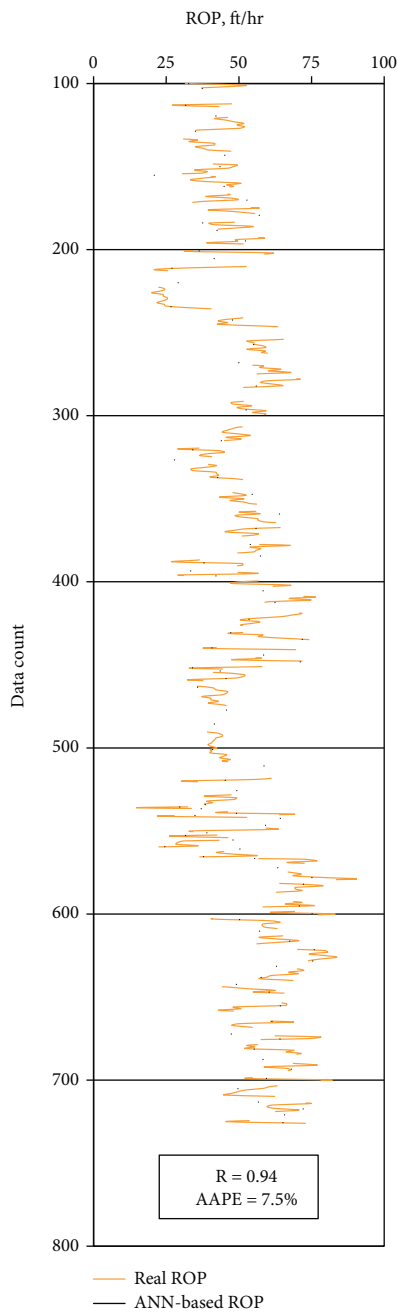


FIGURE 6: Plots of the real and estimated ROP for the training dataset.

FIGURE 7: Plots of the actual and predicted ROP for the testing dataset.

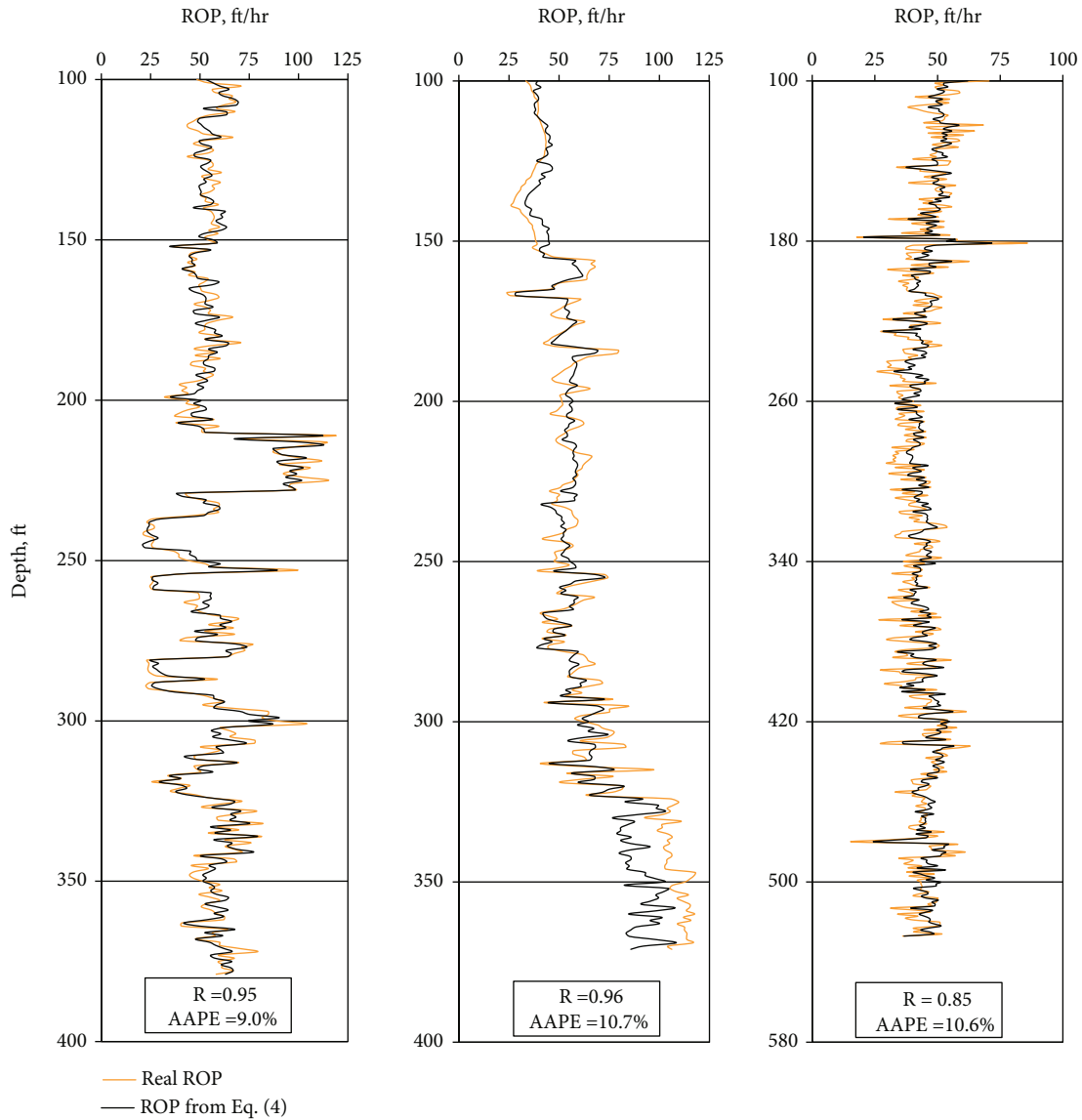


FIGURE 8: Plots of the actual and predicted ROP for the validation datasets.

9.0%, respectively; the ROP for Well-B was predicted with R of 0.96 and AAPE of 10.7%, while the ROP was estimated in Well-D with R of 0.85 and AAPE of 10.6%.

3.4. Comparing the Developed Equation for Predicting the ROP. In this section, the performance of Equation (4) was compared with four previously available equations developed by Maurer [9], Bingham [10], Bourgoyne and Young [11], and Al-Abduljabbar [13] for estimating the ROP in Well-A.

The constants required for all these correlations were calculated in this study based on regression analysis and as a function of the formation properties and drilling parameters. For Maurer's model, the constant k was found to be 10146000. For Bingham's model, the constants k , a_5 , and c were found to be 0.339, -0.269, and 0.636, respectively. For Bourgoyne and Young correlation, the parameters a_1 , a_2 , a_3 , a_4 , a_5 , a_6 , a_7 , and a_8 were found based on the formation

type and drilling parameters to be 6.534, 0.0032, -0.0022, 0.0002, -0.4789, 0.9349, 0, and 0.3314, respectively. For Al-Abduljabbar model, the constants a and b were found to be equal to 0.632 and 0.760, respectively.

The result of comparing Equation (4) with other equations is presented in Figure 9. This result indicated the high accuracy of Equation (4) was able to estimate the ROP with accuracy higher than the available correlations where the R and AAPE for the ROP estimated with Equation (4) are 0.95 and 9.0%, respectively.

Comparing the available correlations, Al-Abduljabbar's model predicted the ROP with the highest R of 0.81 and the lowest AAPE of 14.5% as indicated in Figures 9 and 10. Bingham's model was the second accurate correlation which estimated the ROP with R and AAPE of 0.71 and 15.4%, respectively, followed by Bourgoyne and Young's model which assessed the ROP with R of 0.73 only and AAPE of 17.2%, while Maurer's model is considered the least accurate,

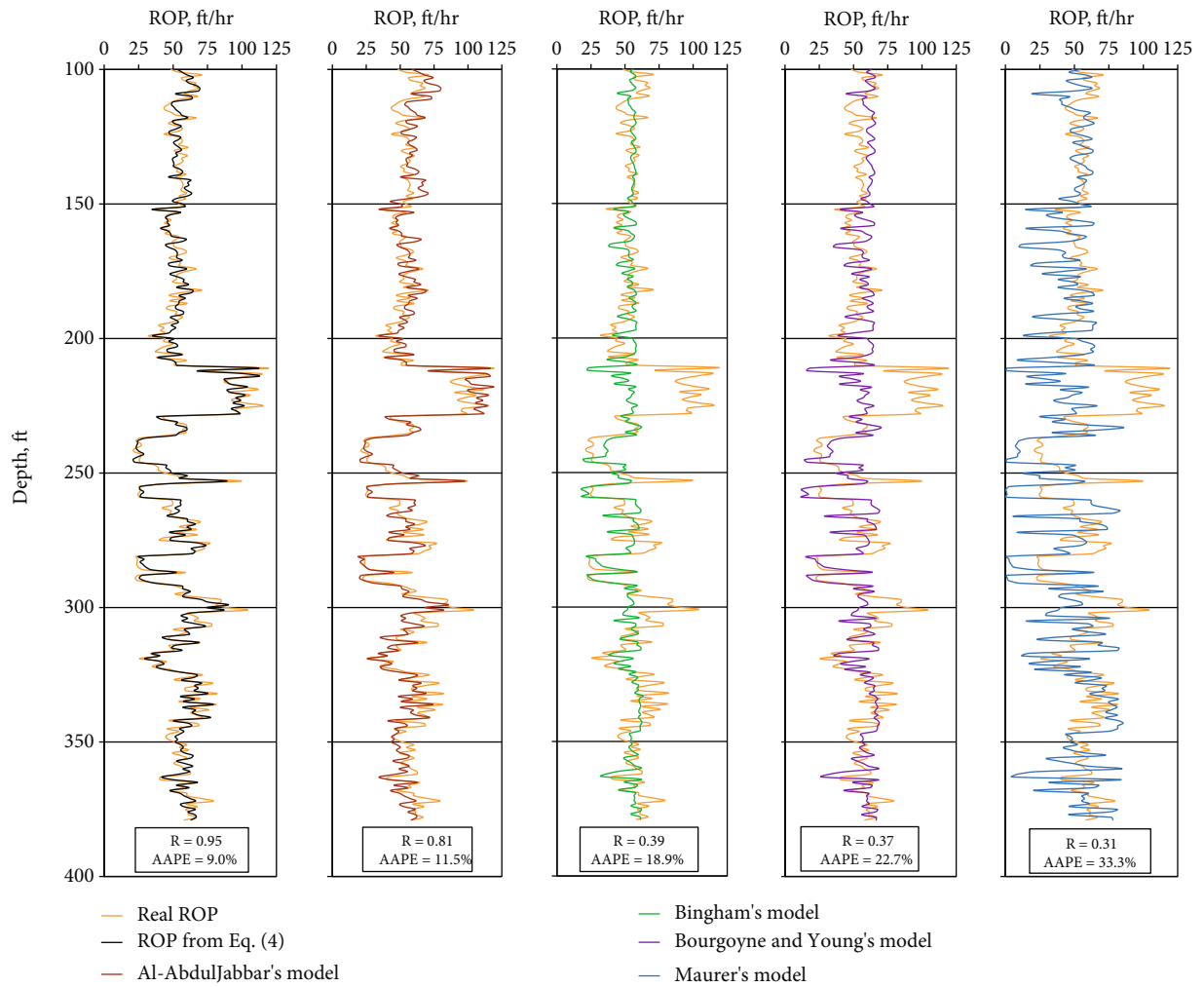


FIGURE 9: Comparison of predictability of Equation (4) and the other ROP models.

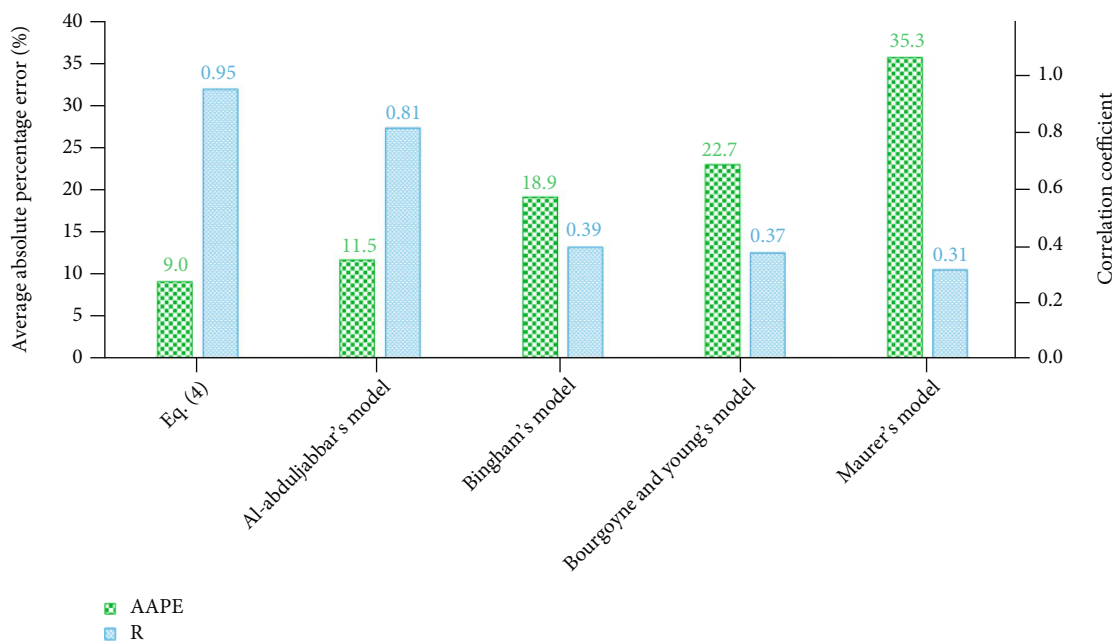


FIGURE 10: Comparison of the accuracy measures for predicting the ROP in Well-A.

and it predicted the ROP with a low R of 0.69 and a high AAPE of 30.9%.

These results of this comparison approved that Equation (4) is highly accurate compared with the other available correlations in assessing the ROP while drilling sandstone formations.

4. Conclusions

This study introduced an empirical equation for predicting the ROP in real time while drilling through sandstone formation; the developed equation was based on the optimized ANN model, and it evaluated the ROP from the drilling mechanical parameters which are measurable at the surface. The results of this study showed that:

- (i) For the training data, the learned ANN model predicted the ROP with an AAPE of 7.5%
- (ii) The extracted equation was tested on data gathered from the same training wells where it estimated the ROP with AAPE of 8.1%
- (iii) The equation was then validated on three wells, and it assessed the ROP with AAPE's of 9.0%, 10.7%, and 8.9% in Well-A, Well-B, and Well-D, respectively
- (iv) Compared with the available empirical equations, the equation developed in this study was most accurate in estimating the ROP

Appendix

Determination of the Coefficients a_1 to a_5 and the Constant c

Firstly, to determine the values of the coefficients a_1 to a_5 , the weights of the output and training layers neurons listed in Table 12 are transformed into matrices with sizes of [1, 5] and [5] and multiplied by each other as explained in Equation (A.1). This multiplication leads to a matrix of [1, 5]; the five components of this matrix in the R.H.S of Equation (A.1) are the coefficients a_1 to a_5 .

$$\begin{aligned}
 & (-0.0158 - 0.4067 - 0.0948 - 0.3190 \ 0.4293) \\
 & \times \begin{pmatrix} 0.718 & -0.740 & -0.593 & -0.594 & 0.310 \\ 0.077 & -0.408 & 0.556 & -1.132 & -0.777 \\ -0.947 & 0.059 & 0.778 & 0.565 & 0.728 \\ 0.276 & 0.893 & -0.351 & -0.148 & 0.969 \\ 0.107 & 0.797 & 0.436 & 0.686 & 0.362 \end{pmatrix} \\
 & = \begin{pmatrix} 0.0049 \\ 0.2292 \\ 0.0083 \\ 0.7577 \\ 0.0884 \end{pmatrix}
 \end{aligned} \tag{A.1}$$

Secondly, to determine the constant c , the output layer matrix with the size of [1, 5] was multiplied by the training layer biases matrix with the size of [1, 5]; the values used in these matrices are extracted from Table 12. The result of this multiplication was then added to the output layer bias of 0.564 as shown in Equation (A.1). The bias of 0.564 is extracted from Table 12.

$$\begin{aligned}
 & (0.0158 - 0.4067 - 0.0948 - 0.3190 \ 0.4293) \\
 & \times \begin{pmatrix} 0.4761 \\ 0.5817 \\ -0.1745 \\ 0.3991 \\ -0.5657 \end{pmatrix} + (0.5637) = -0.0341.
 \end{aligned} \tag{A.2}$$

Nomenclature

AAPE:	Average absolute percentage error
ANN:	Artificial neural networks
DSR:	Drillstring rotation
FNN:	Functional neural networks
LDA:	Linear discriminant analysis
MW:	Mud weight
PV:	Drilling fluid plastic viscosity
Q:	Drilling fluid flowrate
QDA:	Quadratic discriminant analysis
R:	Correlation coefficient
RF:	Random forest
ROP:	Rate of penetration
SPP:	Standpipe pressure
SVR:	Support vector regression
TVD:	True vertical depth
UCS:	Unconfined compressive strength
WOB:	Weight on bit
YP:	Drilling fluid yield point.

Data Availability

Most of the data are available in the manuscript. A detailed sample will be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. T. Bourgoyne Jr., K. K. Millheim, M. E. Chenevert, and F. S. Young, *Applied Drilling Engineering*, vol. 2, Society of Petroleum Engineers Text Book Series, Richardson, TX, 1991.
- [2] A. A. Mahmoud, M. Elzenary, and S. Elkatatny, "New hybrid hole cleaning model for vertical and deviated wells," *Journal of Energy Resources Technology*, vol. 142, no. 3, article 034501, 2020.
- [3] F. Akgun, "How to estimate the maximum achievable drilling rate without jeopardizing safety," in *Abu Dhabi International*

- Petroleum Exhibition and Conference*, Abu Dhabi, United Arab Emirates, 2002.
- [4] M. E. Hossain and A. A. Al-Majed, *Fundamentals of sustainable drilling engineering*, John Wiley & Sons, 2015.
 - [5] R. E. Osgouei, *Rate of Penetration Estimation Model for Directional and Horizontal Wells*, [M.S. thesis], Middle East Technical University, Turkish, 2007.
 - [6] C. Hegde, H. Daigle, H. Millwater, and K. Gray, "Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models," *Journal of Petroleum Science and Engineering*, vol. 159, pp. 295–306, 2017.
 - [7] C. Soares and K. Gray, "Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models," *Journal of Petroleum Science and Engineering*, vol. 172, pp. 934–959, 2019.
 - [8] C. Soares, H. Daigle, and K. Gray, "Evaluation of PDC bit ROP models and the effect of rock strength on model coefficients," *Journal of Natural Gas Science and Engineering*, vol. 34, pp. 1225–1236, 2016.
 - [9] W. Maurer, "The "perfect-cleaning" theory of rotary drilling," *Journal of Petroleum Technology*, vol. 14, no. 11, pp. 1270–1274, 1962.
 - [10] M. G. Bingham, *A New Approach to Interpreting Rock Drillability*, Petroleum Publishing Co., Tulsa, Oklahoma, 1965.
 - [11] A. Bourgoyne and F. S. Young, "A multiple regression approach to optimal drilling and abnormal pressure detection," *Society of Petroleum Engineers Journal*, vol. 14, no. 4, pp. 371–384, 1974.
 - [12] T. M. Warren, "Penetration rate performance of roller cone bits," *SPE Drilling Engineering*, vol. 2, no. 1, pp. 9–18, 1987.
 - [13] A. Al-AbdulJabbar, *Utilizing field data to understand the effect of drilling parameters and mud rheology on rate of penetration in carbonate formations*, King Fahd University of Petroleum & Minerals (KFUPM), 2017, <https://eprints.kfupm.edu.sa/id/eprint/140167>.
 - [14] K. He, C. Xie, S. Chen, and K. K. Lai, "Estimating VaR in crude oil market: a novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network," *Neurocomputing*, vol. 72, no. 16–18, pp. 3428–3438, 2009.
 - [15] P. D. McElroy, H. Bibang, H. Emadi, Y. Kocoglu, A. Hussain, and M. C. Watson, "Artificial neural network (ANN) approach to predict unconfined compressive strength (UCS) of oil and gas well cement reinforced with nanoparticles," *Journal of Natural Gas Science and Engineering*, vol. 88, article 103816, 2021.
 - [16] L. Yu, W. Dai, L. Tang, and J. Wu, "A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting," *Neural Computing & Applications*, vol. 27, no. 8, pp. 2193–2215, 2016.
 - [17] M. Ahmadi and Z. Chen, "Machine learning-based models for predicting permeability impairment due to scale deposition," *Journal of Petroleum Exploration and production technology*, vol. 10, no. 7, pp. 2873–2884, 2020.
 - [18] H. E. Elzain, S. Y. Chung, V. Senapathi, S. Sekar, N. Park, and A. A. Mahmoud, "Modeling of aquifer vulnerability index using deep learning neural networks coupling with optimization algorithms," *Environmental Science and Pollution Research*, vol. 28, no. 40, pp. 57030–57045, 2021.
 - [19] M. Ahmadi and Z. Chen, "Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs," *Petroleum*, vol. 5, no. 3, pp. 271–284, 2019.
 - [20] S. A. Ahmed, A. A. Mahmoud, S. Elkatatny, M. Mahmoud, and A. Abdulraheem, "Prediction of pore and fracture pressures using support vector machine," in *2019 International Petroleum Technology Conference*, Beijing, China, 2019.
 - [21] A. A. Mahmoud, S. Elkatatny, A. Alsabaa, and D. Al Shehri, "Functional neural networks-based model for prediction of the static Young's modulus for sandstone formations," in *54th US Rock Mechanics/Geomechanics Symposium*, Colorado, USA, 2020.
 - [22] H. Osman, A. Ali, A. A. Mahmoud, and S. Elkatatny, "Estimation of the rate of penetration while horizontally drilling carbonate formation using random forest," *Journal of Energy Resources Technology*, vol. 143, no. 9, article 093003, 2021.
 - [23] H. I. Bilgesu, L. T. Tetrick, U. Altmis, and M. S. Ameri, "A new approach for the prediction of rate of penetration (ROP) values," in *Proceedings of the SPE Eastern Regional Meeting*, Lexington, Kentucky, 1997.
 - [24] M. Bataee and S. Mohseni, "Application of artificial intelligent systems in ROP optimization: a case study in Shadegan oil field," in *SPE Middle East Unconventional Gas Conference and Exhibition*, Muscat, Oman, 2011.
 - [25] A. Al-AbdulJabbar, S. Elkatatny, M. Mahmoud, and A. Abdulraheem, "Predicting rate of penetration using artificial intelligence techniques," in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, 2018.
 - [26] S. Elkatatny, "New approach to optimize the rate of penetration using artificial neural network," *Arabian Journal for Science and Engineering*, vol. 43, no. 11, pp. 6297–6304, 2018.
 - [27] S. A. Ahmed, S. Elkatatny, A. Abdulraheem, M. Mahmoud, A. Z. Ali, and I. M. Mohamed, "Prediction of rate of penetration of deep and tight formation using support vector machine," in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, 2018.
 - [28] C. Gan, W. H. Cao, M. Wu et al., "Prediction of drilling rate of penetration (ROP) using hybrid support vector regression: a case study on the Shennongjia area, Central China," *Journal of Petroleum Science and Engineering*, vol. 181, article 106200, 2019.
 - [29] M. Oyedere and K. Gray, "ROP and TOB optimization using machine learning classification algorithms," *Journal of Natural Gas Science and Engineering*, vol. 77, article 103230, 2020.
 - [30] A. Al-AbdulJabbar, S. Elkatatny, A. Abdulhamid Mahmoud et al., "Prediction of the rate of penetration while drilling horizontal carbonate reservoirs using the self-adaptive artificial neural networks technique," *Sustainability*, vol. 12, no. 4, p. 1376, 2020.
 - [31] A. Al-Abduljabbar, A. A. Mahmoud, and S. Elkatatny, "Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations," *Arabian Journal of Geosciences*, vol. 14, no. 2, p. 117, 2021.
 - [32] A. M. Alali, M. F. Abughaban, B. M. Aman, and S. Ravela, "Hybrid data driven drilling and rate of penetration optimization," *Journal of Petroleum Science and Engineering*, vol. 200, article 108075, 2021.
 - [33] F. E. Dupriest and W. L. Koederitz, "Maximizing drill rates with real-time surveillance of mechanical specific energy," in *SPE/IADC drilling conference*, Amsterdam, Netherlands, February 2005.

- [34] M. Armenta, "Identifying inefficient drilling conditions using drilling-specific energy," in *2008 Annual Technical Conference and Exhibition*, Denver, Colorado, USA, 2008.
- [35] R. Teale, "The concept of specific energy in rock drilling," *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 2, no. 1, pp. 57–73, 1965.