



# Evaluation of the wellbore drillability while horizontally drilling sandstone formations using combined regression analysis and machine learning models

Ahmed Abdulhamid Mahmoud<sup>1</sup> · Hany Gamal<sup>1</sup> · Salaheldin Elkatatny<sup>1</sup>

Received: 12 April 2022 / Accepted: 12 April 2023 / Published online: 24 April 2023  
© The Author(s) 2023

## Abstract

The rate of penetration (ROP) is an influential parameter in the optimization of oil well drilling because it has a huge impact on the total drilling cost. This study aims to optimize four machine learning models for real-time evaluation of the ROP based on drilling parameters during horizontal drilling of sandstone formations. Two well data sets were implemented for the model training–testing (Well-X) and validation (Well-Y). A total of 1224 and 524 datasets were implemented for training and testing the model, respectively. A correlation for ROP assessment was suggested based on the optimized artificial neural network (ANN) model. The precision of this equation and the optimized models were tested (524 datapoints) and validated (2213 datapoints), and their accuracy was compared to available ROP correlations. The developed ANN-based equation predicted the ROP with average absolute percentage errors (AAPE) of 0.3% and 1.0% for the testing and validation data, respectively. The new empirical equation and the optimized fuzzy logic and functional neural network models outperformed the available correlations in assessing the ROP. The support vector regression accuracy performance showed AAPE of 26.5%, and the correlation coefficient for the estimated ROP was 0.50 for the validation phase. The outcomes of this work could help in modeling the ROP prediction in real time during the drilling process.

**Keywords** Rate of penetration · Horizontal drilling · Sandstone formations · Machine learning

## Abbreviations

AAPE	Average absolute percentage error
ANN	Artificial neural network
COA	Cuckoo optimization algorithm
FIS-SC	Fuzzy inference system with subtractive clustering
FNN	Functional neural network
FS	Forward selection
GRNN	General regression neural network
MPNN	Multilayer perceptron neural network
PSA	Particle swarm optimization algorithm
QA	Quality assurance
QC	Quality control
R	Correlation coefficient
RMSE	Root-mean-square error

ROP <sub>regression</sub>	Regression-based rate of penetration
SVR	Support vector regression

## List of symbols

$A$	Area of drilled hole, in <sup>2</sup>
$a$	Constant
$a_1$ to $a_8$	The constants for Bourgoyne and Young's correlation
$aa_1$ to $aa_{11}$	The constants for Osgouei's correlation
$a_B$ and $c_B$	The exponents for Bingham's correlation
$a_W$ and $c_W$	The constants for Warren's correlation
$b_1$ and $b_2$	Biases
$b_W$	The exponent for Warren's correlation
$D$	Well depth, ft
DSR	Drillstring rotation speed, rpm
$k_B$	Bingham's correlation constant
$k_M$	Maurer's correlation constant
$M$	Number of neurons
MSE	Mechanical specific energy, psi
MW	Mud weight, ppg
$N$	Number of input parameters
PV	Drilling fluid plastic viscosity, cP
$Q$	Drilling fluid flow rate, gpm

✉ Salaheldin Elkatatny  
elkatatny@kfupm.edu.sa

<sup>1</sup> College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ROP	Rate of penetration, ft/hr
SPP	Standpipe pressure, psi
$T$	Torque, kft lb <sub>f</sub>
UCS	Unconfined compressive strength
$w_1$ and $w_2$	Weights of the ANN neurons
WOB	Weight on bit, klbs
$x$	The input parameters
$x_2$ to $x_8$	The drilling parameters for Bourgoyne and Young's correlation
$xx_2$ to $xx_{11}$	The drilling parameters for Osgouei's correlation
$y$	The output parameter
$p$	Thickness of rock layer, in

## Introduction

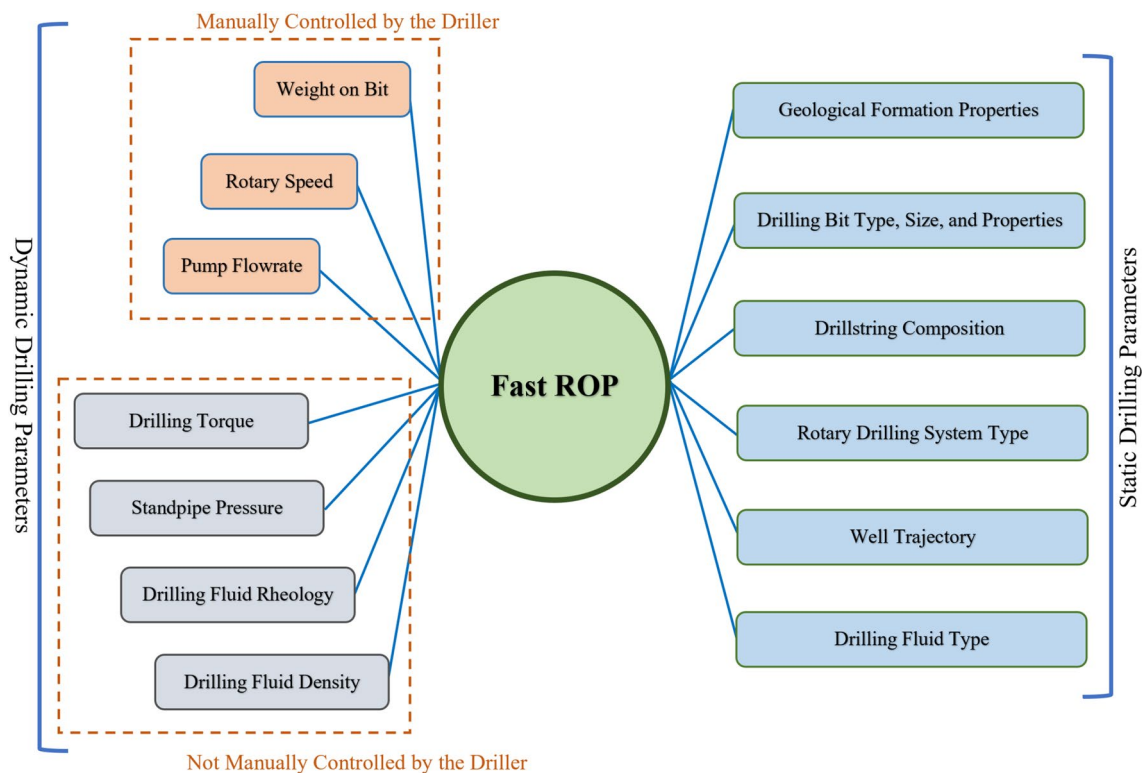
The rate of penetration (ROP) is a critical drilling parameter indicating how fast the drill bit is penetrating the formations (Bourgoyne et al. 1991). Although it is necessary to decrease the drilling time by increasing the ROP and, hence, decreasing the drilling cost, the speed at which the drillbit is penetrating the downhole formations (i.e., ROP) is limited by the cuttings lifting capacity of the drilling fluid which is required to maintain the wellbore clean of the cutting (Mahmoud et al. 2020a; b, c, d). The use of high ROP could

also lead to other problems during the drilling process such as drillstring vibration which is accounted in many situations to the wellbore instability or loss of bottom hole assembly (Akgun 2002).

ROP is affected by many other drilling parameters, the drillstring composition, drilling fluid properties, well trajectory, and others as indicated in Fig. 1, as the figure shows the statics and dynamic drilling parameters that greatly affect the ROP. Most of these parameters are affecting each other, in other words, changing one of these parameters in many cases affects the other parameters contributing to the ROP, this fact makes the possibility of predicting the ROP more complicated, and it also complicated the possibility of evaluating the impact of only a single factor on the ROP (Mhosain and Al-Majed 2015; Osgouei 2007).

Nowadays, many ROP models and empirical correlations are available, each of these models was developed after setting specific assumptions and it is also using specific input parameters to assess the ROP. Because of the variety of assumptions and inputs, these ROP models are considerably different in terms of their accuracy (Lyons and Plisga 2004; Mitchell and Miska 2011; Rabia 2001). Table 1 summarizes the available ROP correlations.

The most commonly used empirical correlation in the last decade was the one developed by Bourgoyne and Young (1974); this correlation defined the ROP as based on eight



**Fig. 1** The parameters affecting ROP optimization

**Table 1** The list of some of the available ROP correlations

Authors	Inputs	Formula
Maurer (1962)	Drillstring rotation speed (DSR), weight on bit (WOB), unconfined compressive strength (UCS), the drillbit diameter ( $d_b$ ), and the drillability constant for Maurer correlation ( $k_M$ )	$ROP = k_M \frac{DSR \times WOB^2}{(d_b)^2 UCS^2}$
Bingham (1965)	Drillstring rotation speed, weight on bit, the drillbit diameter, the drillability constant for Bingham's correlation ( $k_B$ ), and the exponents $a_B$ and $c_B$	$ROP = k_B \left( \frac{WOB}{d_b} \right)^{a_B} DSR^{c_B}$
Bourgoyne and Young (1974)	Well depth ( $D$ ), Pore pressure, mud weight, WOB, mud flowrate, DSR, and bit wear	$\frac{dD}{dt} = e^{\left( a_1 + \sum_{i=2}^8 a_i \times x_i \right)}$
Warren (1987)	DSR, UCS, WOB, and the $d_b$	$ROP = \left[ \frac{a_w UCS^2 \times d_b^2}{DSR^{b_w} \times WOB^2} + \frac{c_w}{DSR \times d_b} \right]^{-1}$
Osgoei (2007)	Well depth, pore pressure, mud weight, weight on bit, mud flowrate, drillstring rotation speed, bit type, and wear, and the wellbore inclination	$\frac{dD}{dt} = e^{\left( aa_1 + \sum_{i=2}^{11} aa_i \times x_i \right)}$
Al-AbdulJabbar, (2017)	Drillstring rotation speed, UCS, WOB, drilling fluid plastic viscosity (PV), the drillbit diameter, standpipe pressure (SPP), torque, drilling fluid flowrate ( $Q$ ), and the drilling fluid weight (MW)	$ROP = \frac{WOB^a \times DSR \times Torque \times SPP \times Q}{d_b^c \times MW \times PV \times UCS^b}$

functions as indicated in Table 1. Any of these functions is based on specific drilling parameters. Although the relationship between the ROP and these eight functions is very complex and not linear, Bourgoyne and Young (1974) simply considered the ROP as the multiplication and addition of these functions, which limits the accuracy of ROP prediction.

Recently, different machine learning models were successfully applied to different aspects of science and engineering (Najafzadeh 2019; Saberi-Movahed et al. 2020; Thanh et al. 2020; Elzain et al. 2021; Thanh et al. 2022a; Thanh and Lee 2022; Thanh et al. 2022b), and petroleum engineering is not an exception (Barbosa et al. 2019; Elkattatny et al. 2019; Mahmoud et al. 2020b; Mahmoud et al. 2020d; Alsaihati et al. 2021; Siddig et al. 2021). Recent research was performed to enhance ROP prediction using machine learning capabilities, and in these studies, different machine learning techniques, input parameters, and other technical aspects related to the drilling operations and well planning were considered for ROP prediction while drilling carbonate formation (Mahmoud et al. 2020a; Osman et al. 2021), natural gas-bearing sandstone formation (Al-AbdulJabbar et al. 2022a), and complex lithology formations (Gamal et al. 2020).

In 2011, Bahari et al. (2011) hierarchically employed the general regression neural network (GRNN) to uncover the complex and nonlinear relationship between the ROP and the eight functions defined by Bourgoyne and Young (1974) to improve ROP prediction. The results indicated that the GRNN was powerful to define the relationship between the ROP and these eight functions, and it improved ROP prediction.

Anemangely et al. (2018) utilized two hybrid models for ROP prediction based on mud log data and petrophysical

logs. The hybrid models are composed of a multilayer perceptron neural network (MPNN) coupled with the particle swarm optimization algorithm (PSA) in the first model and with a Cuckoo optimization algorithm (COA) in the second model. The authors found that the optimized models were superior in ROP estimation. The results also indicated that although increasing the number of inputs is important to enhance ROP prediction, the use of five or more inputs did not significantly improve ROP predictability using their optimized models.

Recently, Al-AbdulJabbar et al. (2022b) suggested an artificial neural network (ANN)-based empirical correlation for estimating the ROP during horizontal drilling of carbonate formation. To generalize their correlation, the authors optimized the ANN model using data collected from five different wells, the inputs used are real-time available drilling parameters of  $Q$ , SPP, WOB, torque, and DSR, which enabled real-time prediction while drilling another three wells in the same reservoir, and the empirical correlation accurately predicted the ROP.

In another study, Al-AbdulJabbar et al. (2022a) proposed an estimation of the ROP into sandstone formation using the ANN model based on the same inputs considered by Al-AbdulJabbar et al. (2022b). The results also indicated the high precision of the ANN in evaluating the ROP for this sandstone formation. A summary of some available data-driven-based ROP models is presented in Table 2.

In this study, machine learning models of ANN, fuzzy inference system with subtractive clustering (FIS-SC), SVR, and functional neural network (FNN) were optimized for ROP estimation while drilling through sandstone formations in a horizontal model. The four machine learning models were optimized to estimate the ROP from the DSR, standpipe pressure (SPP), WOB, and torque, as well

**Table 2** Some of the machine learning-based ROP models

Authors	Method	Inputs
Bahari et al. (2011)	GRNN	The eight functions defined by Bourgoyne and Young (1974)
Anemangely et al. (2018)	MPNN-PSA and MPNN-COA	Mud log data and petrophysical logs
Elkatatny (2018)	ANN	Drilling fluid flowrate, PV, and MW, standpipe pressure, DSR, WOB, and the torque
Ahmed et al. (2018)	Support vector regression (SVR)	The drilling fluid flowrate, MW, PV, yield point, and solids percentage, in addition to the standpipe pressure, DSR WOB, march funnel viscosity, and the drilling torque
Al-AbdulJabbar et al. (2018)	ANN	DSR, WOB, standpipe pressure, torque, and the drilling fluid flowrate
Al-AbdulJabbar et al. (2020)	ANN	DSR, WOB, torque, gamma-ray, formation deep resistivity, and the formation bulk density
Al-AbdulJabbar et al. (2022b)	ANN	DSR, WOB, standpipe pressure, torque, and the drilling fluid flowrate

as a new regression-based rate of penetration parameter or  $ROP_{\text{regression}}$  which was calculated from the DSR and WOB. Based on the trained ANN model, a correlation for assessment of the ROP was derived in this work.

The current study provides novel contributions over the published work by presenting the full approach for developing four machine learning developed models for predicting the penetration rate during the drilling operation from the surface drilling parameters that are available during the drilling operation through horizontal drilling sandstone formations. In addition, this study presented a new method for enhancing the machine learning capability for ROP prediction by presenting a new calculated parameter  $ROP_{\text{regression}}$  based on mathematical derivation for the mechanical specific energy (MSE) equation to relate the MSE and the DSR to the ROP. Besides, the current research proposed a newly developed empirical equation for ROP estimation which is based on the ANN model. All these novel contributions will enhance the ROP estimation and provide real-time guidance for the drilling engineers to optimize the controllable drilling parameters for the best penetration rate for cost savings during the drilling operation.

## Methodology

Four different machine learning techniques were used to develop various real-time ROP models from only the surface measurable drilling parameters and the  $ROP_{\text{regression}}$  parameter. Before training the models, the data collected for this work were studied to eliminate the non-real values and outliers. Based on linear regression, the expression for the  $ROP_{\text{regression}}$  was determined to account for the ROP as a function of the drilled hole area, the drillpipe diameter, WOB, and DSR.

## Training data preparation and preprocessing

In this study, two wells were selected (Well-X and Well-Y). It is worth mentioning that the two wells penetrated the same geology scheme, so drilling the same formations during the drilling phase. The drilling data of WOB, SPP, DSR, and the torque recorded on a real-time base by the surface sensors were obtained from the two wells under consideration; both wells were drilled using the conventional bottom hole assembly. These data were obtained while drilling sandstone formations; both wells were drilled using a top drive rotary system. Originally, 3082 datasets from Well-X were collected to learn the machine learning models and test the learned models, while 4662 datasets from Well-Y were used to validate the learned models. Different processes of data quality control (QC) and quality assurance (QA) such as non-real values and outlier removal were performed; these processes were considered to ensure that only the valid data was considered to optimize the models.

During the QA/QC stage, the MSE term developed by Teale (1965) to describe the applied energy by the drill bit to penetrate the formation (Dupriest and Koederitz 2005) was considered for non-real values determination. The MSE should be optimized to have values similar to the UCS (Teale 1965).

The sandstone formation considered to obtain the data needed in this work has UCS between 25,000 and 45,000 psi. Figure 2 shows the plot of MSE versus ROP for all Well-X data. As shown in Fig. 2, several MSE values are considerably lower or greater than the range of the UCS, this huge difference confirms that at all these points the collected data are unrealistic and the data corresponding to these points must be removed since they represent inefficient drilling. The MSE values for the data used in this work are between 15,000 and 75,000 psi, which was determined based on the formation  $UCS \pm$  a margin; all points with MSE out of this range were removed from the training data.

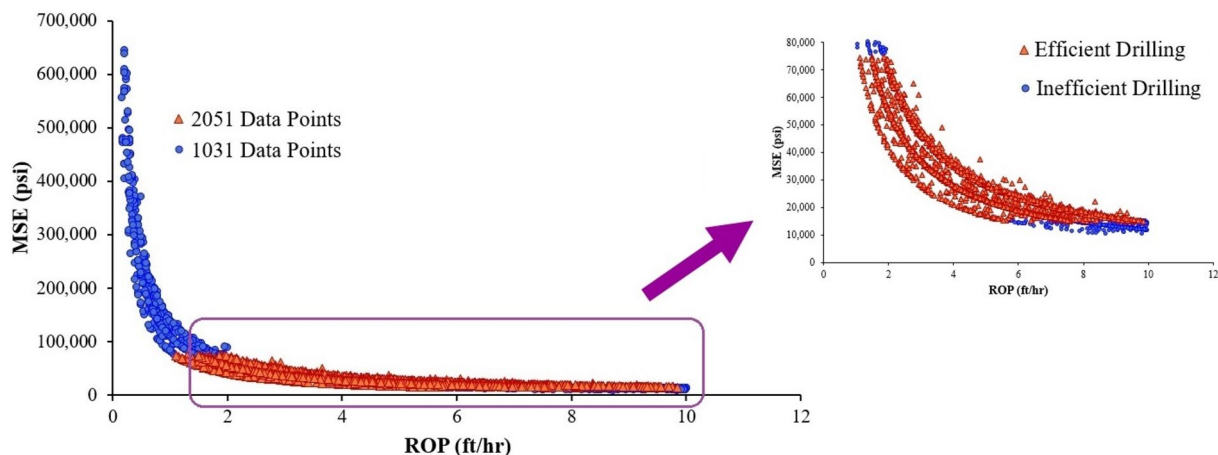


Fig. 2 The MSE versus ROP for all Well-X data

1031 of the datasets gathered from Well-X represent locations of inefficient drilling (from Fig. 2), these datasets were eliminated from the inputs at this stage. Then, 2051 of the data collected from Well-X are considered realistic.

Before considering the data for training the machine learning models, the outliers in the inputs were also removed. For this purpose, only the input values within  $\pm 3.0$  standard deviation are considered non-outliers, while the other values were removed from the training data. After this process, 1748 datasets of Well-X data were considered to validate the machine learning models.

### Developing an expression for the regression-based ROP

To improve ROP predictability using the machine learning models, a new input parameter was developed based on regression analysis and it was called regression-based ROP ( $ROP_{\text{regression}}$ ); the expression for this parameter was developed based on Teale (1965) equation for the MSE, by neglecting the torque in Teale (1965) equation for the MSE, this will lead to express MSE as presented in (1), and this expression was used to develop the expression for  $ROP_{\text{regression}}$ .

$$MSE = \frac{WOB}{A} + \frac{2\pi \times T}{A \times p} \tag{1}$$

where  $T$  is the required torque to remove a layer of rock with thickness  $p$  in a single revolution. The first step in this derivation is to relate the MSE and DSR to the ROP. The MSE and ROP values of the training data of Well-X are plotted in Fig. 3 that shows the MSE-ROP plots for different levels of DSR [60, 80, and 100 rpm], as indicated in this figure, the MSE and ROP values at every single DSR value fitted

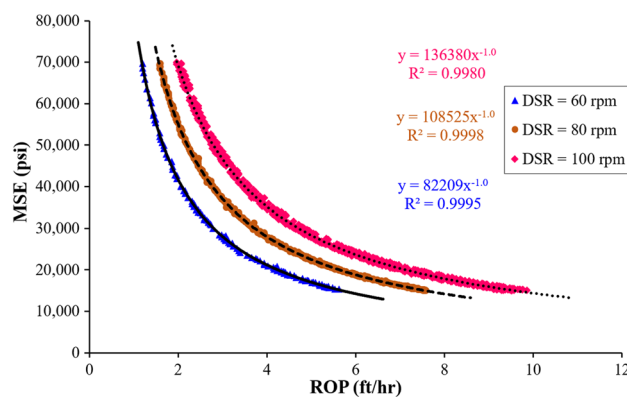


Fig. 3 MSE and ROP plots for training data corresponding to DSRs of 60, 80, and 100 rpm

to a single curve of a power function. The governing equations for the three curves in Fig. 3 have the same exponent of  $-1.0$  and various constants ( $a$ ) of 82,209, 108,525, and 136,380 for DSRs of 60, 80, and 100 rpm, respectively.

The three functions that relate the MSE and ROP to DSR as indicated in Fig. 3 could be represented generally as in (2).

$$MSE = a ROP^{-1.0} = \frac{a}{ROP} \tag{2}$$

Now if the three values of ‘ $a$ ’ (extracted from Fig. 3) and their corresponding DSR are plotted as in Fig. 4 as the plot shows the relationship between ‘ $a$ ’ and DSR is representable by a straight line which could be represented by (3).

$$a = 1354 \text{ DSR} + 696 \tag{3}$$

Now by calling the MSE from (2), after substitution into (1) and replacement of  $T$  and  $p$  with DSR and ROP as

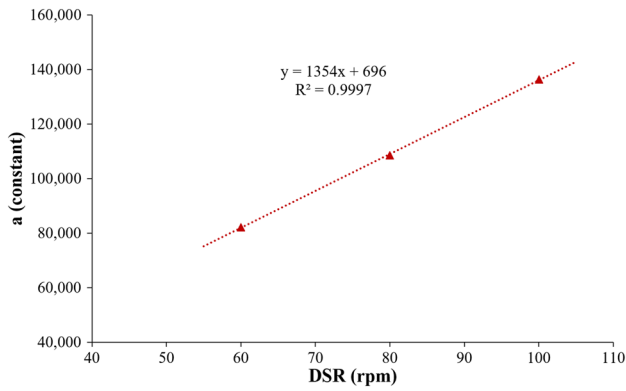


Fig. 4 The plot of the constant ‘a’ values and the corresponding DSR

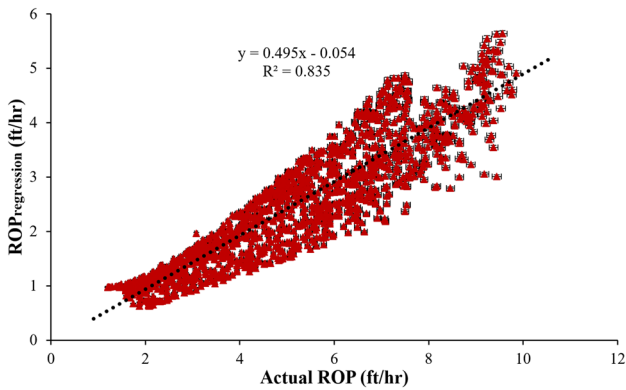


Fig. 5 The cross-plot of the ROP<sub>regression</sub> and ROP for Well-X data (1748 data points)

suggested by (Al-Abduljabbar et al. 2021), this leads to the expression in (4).

$$\frac{a}{ROP} = \frac{1354 \text{ DSR} + 696}{ROP} = \frac{WOB}{A} + \frac{2\pi \times \text{DSR}}{A \times ROP} \tag{4}$$

Rearranging (4) we will get:

$$ROP = \frac{-\left(2\pi \frac{\text{DSR}}{A} - a\right)A}{WOB} = ROP_{regression} \tag{5}$$

$$ROP_{regression} = \frac{-\left(2\pi \frac{\text{DSR}}{A} - (1354 \text{ DSR} + 696)\right)A}{WOB} \tag{6}$$

where the ROP<sub>regression</sub> parameter derived in (6) is the new expression for the ROP which we call the regression-based ROP or ROP<sub>regression</sub>. (6) is defining the ROP<sub>regression</sub> based on DSR and WOB only. Figure 5 shows the cross-plot of actual ROP versus ROP<sub>regression</sub> that is calculated from (6). The good correlation coefficient between ROP and ROP<sub>regression</sub> of 0.835 is enough to guide the machine learning models

toward better ROP prediction. As discussed previously, ROP<sub>regression</sub> was considered as an input to learn the machine learning models in addition to the four surface measurable drilling parameters.

### Optimizing the machine learning models

The machine learning models were optimized to assess the ROP based on five parameters; four surface measurable parameters of the WOB, torque, DSR, and SPP, and the fifth parameter is the ROP<sub>regression</sub> calculated using (6). The machine learning models were learned using 1224 datapoints of the surface measurable parameters and the ROP<sub>regression</sub> to estimate the real ROP; the training data are 70% of Well-X’s. Figure 6 shows the plot of the input parameters considered for learning the machine learning techniques. Table 3 lists the applicability range for the optimized models and the statistical properties of the learning data.

The design parameters of the four models considered in this work were optimized using sensitivity analysis. The first model considered in this study is the ANN. The effect of the training function, transferring function, the number of hidden (training) layers, and the number of neurons were studied during this stage of sensitivity analysis. For the FIS-SC, the cluster radius was optimized between 0.2 and 0.9 and the number of iterations between 10<sup>2</sup> and 3 × 10<sup>3</sup> was studied.

The third machine learning model considered is the functional neural network (FNN). For this technique, the effect of the training method and function type on estimating the ROP was evaluated. Different training methods such as the forward selection (FS), forward–backward selection, backward–forward selection, and backward elimination methods were evaluated. Different training function types were studied in this work such as the linear function without iteration term, nonlinear function without iteration term, and nonlinear function with iteration term (NLFIT).

The effect of different design parameters of the SVR model such as the kernel, kernel options, lambda, epsilon, and C was studied. The effect of different kernels such as the Gaussian and multi-quadratic, various kernel options from 1 to 10, lambda between 10<sup>−7</sup> and 10<sup>−5</sup>, epsilon from 10<sup>−5</sup> to 10<sup>−1</sup>, and C between 10<sup>2</sup> and 3 × 10<sup>3</sup> was evaluated. A summary of the optimum design parameters is provided in Table 4.

### Extracting the new ROP correlation

The ROP empirical correlation developed in this work is based on the weights and biases of the different neurons of the ANN model after optimization. As summarized earlier

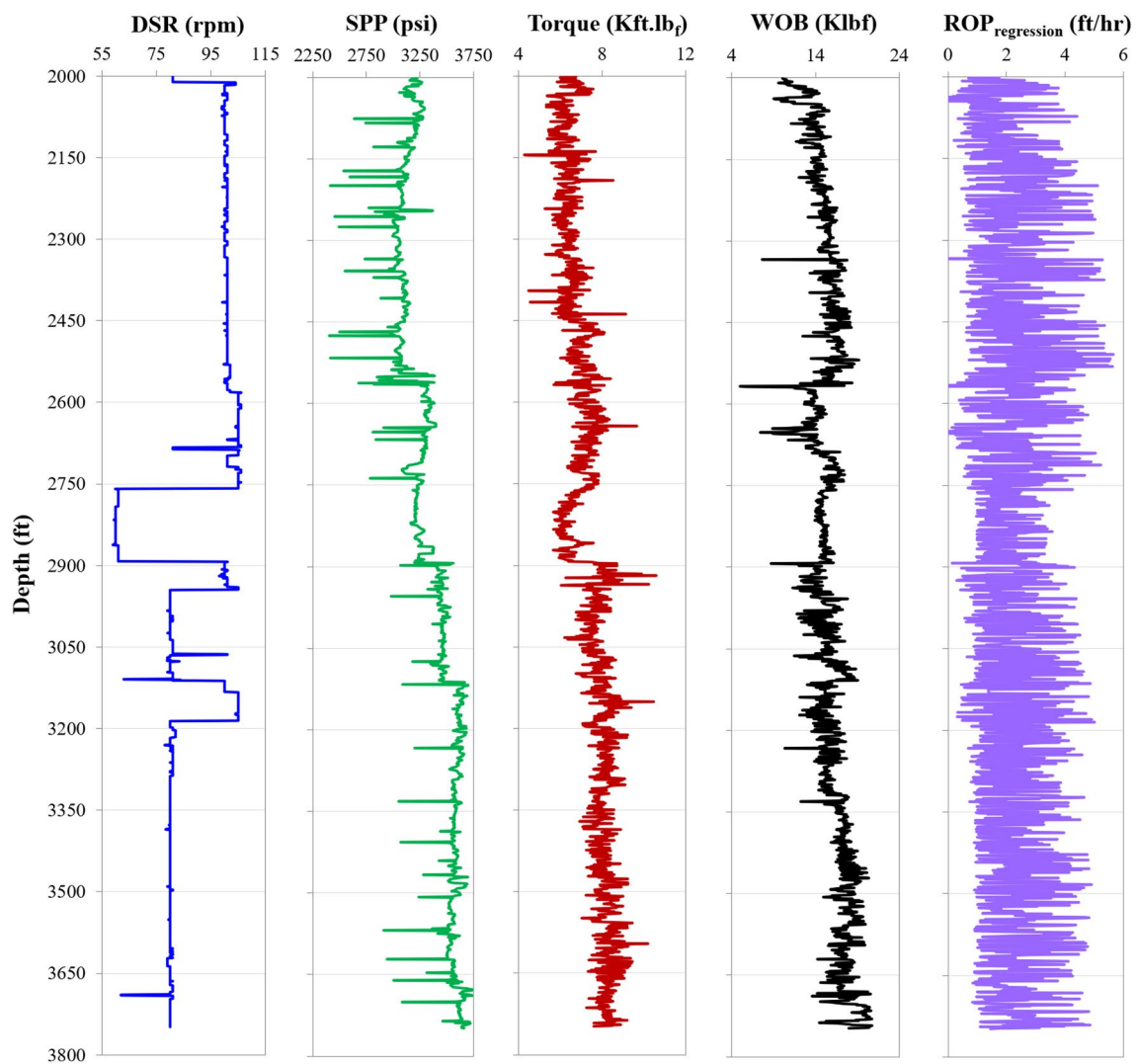


Fig. 6 The training input parameters

Table 3 The statistical characteristic of the training datasets

	DSR (rpm)	SPP (psi)	Torque (kft lb <sub>f</sub> )	WOB (klb <sub>f</sub> )	ROP (ft/hr)
Minimum	59.0	2401	4.32	5.07	1.20
Maximum	106	3746	10.6	20.7	9.85
Range	47.0	1345	6.29	15.7	8.65
Standard Deviation	13.4	229	0.93	1.93	2.09
Sample Variance	179	52,531	0.86	3.73	4.38

in Table 4, the optimized ANN model has one training layer with 15 neurons with Bayesian regularization backpropagation training function and one output layer with the tangential sigmoid transferring function. The expression that represents this model could be presented as in (7).

$$y = \left[ \sum_{j=1}^M w_{j1} \text{tansig} \left( \sum_{i=1}^N w_{ij} x_i + b_j \right) \right] + b_2 \tag{7}$$

where  $y$  represents the output or ROP in this case,  $w$  denotes the different weights,  $x$  denotes the input parameters, and  $b$  represents the biases. Table 5 lists the values of  $w$  and  $b$ .

**Table 4** The optimized parameters of the machine learning techniques

ANN	
Training function	Bayesian regularization back-propagation
Transferring function	Tangential sigmoid
Training layers (neurons)	One layer (15 neurons)
FIS-SC	
Cluster radius	0.5
Ep-size	$2 \times 10^3$
FNN	
Training method	NLFIT
Training function type	FS method
SVR	
Kernel	Gaussian
Kernel options	10
Lambda	$10^{-5}$
Epsilon	$10^{-5}$
C	$10^2$

(7) could be written for the ANN model optimized for ROP estimation which has 15 neurons, 5 inputs, and  $b_2$  of 5.44 as in (8).

$$\text{ROP} = \left[ \sum_{j=1}^{15} w_{j1} \text{tansig} \left( \sum_{i=1}^5 w_{ij} x_i + b_j \right) \right] + 5.44 \quad (8)$$

After expanding (8), the ROP could be expressed by (9).

$$\text{ROP} = \left[ \sum_{j=1}^{15} w_{j1} \left( \frac{1}{1 + e^{-2 * (w_{1,1}(\text{DSR}) + w_{1,2}(\text{SPP}) + w_{1,3}(\text{Torque}) + w_{1,4}(\text{WOB}) + w_{1,5}(\text{ROP}_{\text{regression}}) + b_{1j})}} \right) \right] + 5.44 \quad (9)$$

It is important to mention here that the parameters used to predict the ROP using (9) should be normalized between  $-1$  a  $1$ , and the value of ROP calculated using this equation is in normalized state, which should be denormalized to have the actual ROP value. More information about normalization could be found in our previous publication Al-Abduljabbar et al. (2020).

### Testing and validating the optimized machine learning models and the new ROP correlation

The developed empirical Eq. (9) and the optimized FIS-SC, SVR, and FNN models were tested using 524 data points (Well-X) and then validated using 2213 data points

(Well-Y). The predictability of the developed equation and models Well-Y was compared with that of the available correlations to investigate the enhancement in the assessment of the rate of penetration using (9) and the models developed in this study.

## Results and discussion

### Training the machine learning models

All machine learning models were firstly trained on 1224 datasets of Well-X data (70% of Well-X data). Figure 7 shows the actual versus the predicted ROP from different models for the 1224 training dataset. Comparing the actual ROP (blue diamonds in Fig. 7) and the ROP estimated with the different models (the continuous line in Fig. 7) confirms the perfect matching between these values, which confirms the superior accuracy of the considered machine learning models. As shown in Fig. 7, the ROP for the training data set was assessed accurately with AAPEs of only 0.4%, 2.3%, 2.6%, and 3.6% using the ANN, SVR, FIS-SC, and FNN models, respectively. The ROP predicted with ANN, SVR, FIS-SC, and FNN models have Rs of 0.999, 0.998, 0.998, and 0.997 with the real ROP, respectively, as shown in Fig. 7. The previously discussed results proved the high precision of the ANN, FIS-SC, FNN, and SVR models in assessing the ROP.

### Testing the developed equation and the optimized machine learning models

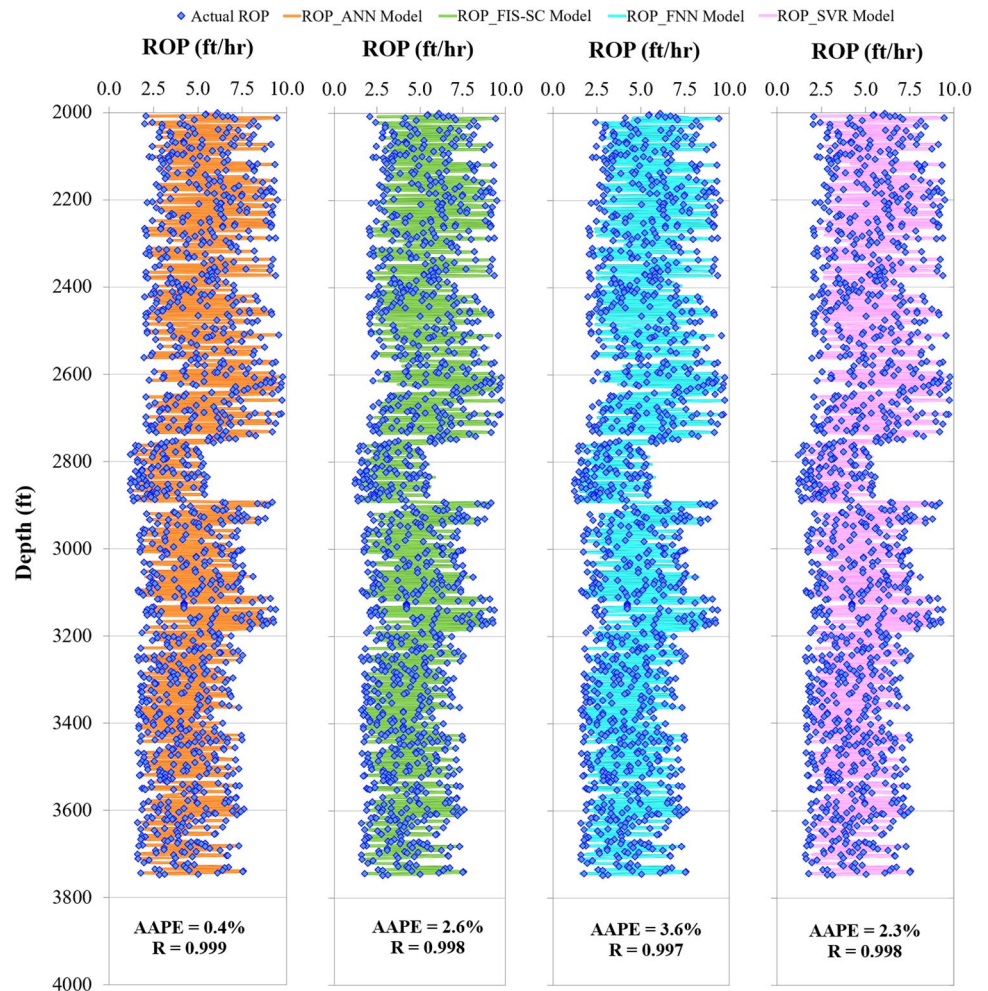
The optimized FIS-SC, SVR, and FNN models and (9), which was based on the ANN model, were tested on 524 unseen datasets from Well-X (30% of Well-X data). Figure 8 displays the actual versus the predicted ROP from different models for the 524 testing datasets. As indicated in Fig. 8, although all models predicted the ROP accurately, (9) was the most precious in estimating the ROP for this dataset of Well-X. From Fig. 8, (9), SVR, FIS-SC, and FNN models assessed the ROP with very low AAPE's of 0.3%, 2.7%, 3.4%, and 3.6% and R's of 0.999, 0.998, 0.997, and 0.992, respectively. Visual comparison of the actual and predicted



**Table 5** The extract weights and biases of the optimized ANN model

No. of neurons	Training layer					Output layer				
	Weight ( $w_1$ )					Biases ( $b_1$ )		Weights ( $w_2$ )		Bias ( $b_2$ )
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$					
$j=1$	0.404	-0.097	0.069	-0.992	2.754	-1.067	-2.110	5.440		
$j=2$	0.223	-0.017	0.015	-0.419	1.493	-1.029	2.747			
$j=3$	3.739	-0.591	0.251	-1.249	-1.979	-6.795	-6.783			
$j=4$	4.616	-2.481	2.058	-11.627	-20.612	-8.933	6.162			
$j=5$	-1.389	-0.047	0.045	5.152	-4.477	5.949	-3.319			
$j=6$	0.137	0.003	-0.002	-0.635	-0.702	-0.915	5.360			
$j=7$	-0.026	-0.010	0.006	0.372	1.420	0.484	3.061			
$j=8$	0.882	0.379	-0.249	-0.481	25.971	4.193	-19.708			
$j=9$	2.481	-0.040	0.005	-4.377	9.721	-10.653	5.680			
$j=10$	5.475	-2.728	2.438	-13.179	-20.507	-9.883	-4.821			
$j=11$	-2.531	-0.243	0.041	-2.214	3.814	-2.243	-0.012			
$j=12$	-1.369	-0.249	0.179	1.723	-28.500	-4.102	-11.859			
$j=13$	0.450	-0.153	0.103	-1.164	3.308	-1.201	1.048			
$j=14$	-0.106	0.750	-0.446	2.241	22.746	4.679	9.207			
$j=15$	0.991	0.186	-0.067	-1.754	13.393	0.733	0.107			

**Fig. 7** The actual and evaluated ROP for the 1224 training datasets



ROP as indicated in Fig. 8 confirms the high reliability of (9) and the optimized FIS-SC, SVR, and FNN models in estimating the ROP.

### Validating the developed equation and the optimized machine learning models

The predictability of (9) and the other optimized machine learning models was also evaluated on the 2213 data points of Well-Y. At this stage, the predictability of (9) and the optimized machine learning models were also compared with four of the available ROP empirical correlations; these are Bingham, Maurer, and Bourgoyne and Young’s correlation. Table 6 lists the calculated constants needed to be used with the different ROP empirical models.

Figure 9 represents the results comparison of the ROP estimation in Well-Y (2213 data points) using Eq. (9) with various available models. As shown in Fig. 9, (9), FIS-SC, and FNN models are most accurate in contrast with the SVR model and the correlations developed earlier for ROP estimation as indicated by the perfect agreement between the real and assessed ROP when (9), FIS-SC, and FNN models are used. All previous empirical correlations assessed the ROP with a very low R of less than 0.24, and the R between the actual ROP and these evaluated with the SVR was 0.50,

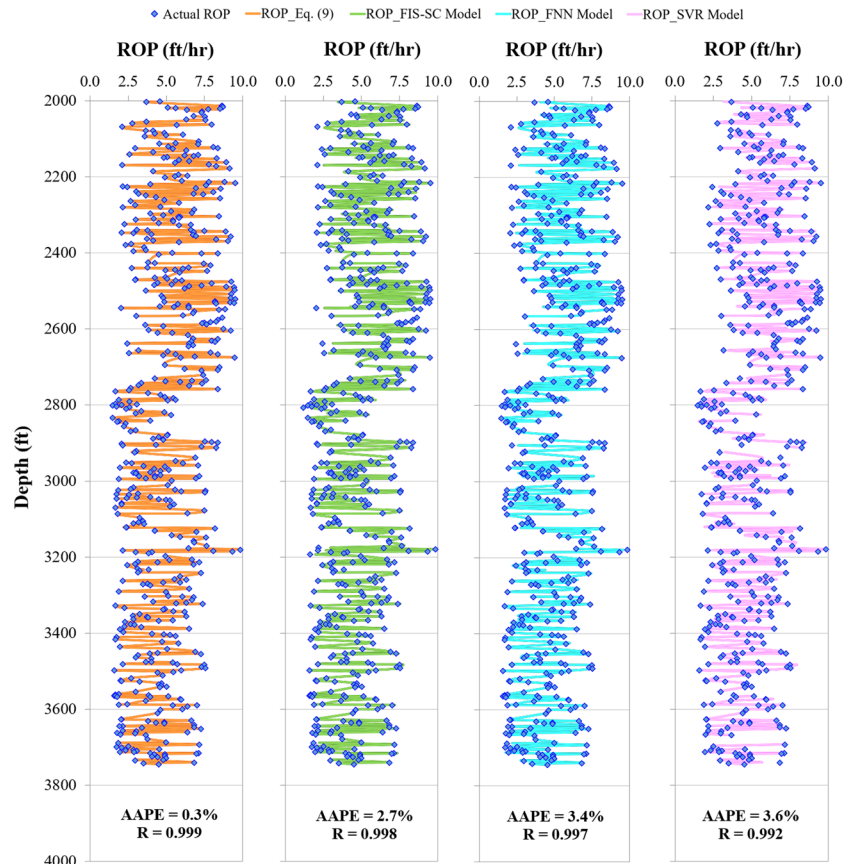
while (9), FIS-SC, and FNN models assessed the ROP with R’s of 0.99, 0.99, and 0.97 as shown in Fig. 9.

Figure 10 demonstrates the ROP accuracy comparison for Well-Y data using with various available models. As indicated in Fig. 10, the SVR model and all previous correlations assessed the ROP with very high AAPE and low root-mean-square error (RMSE). The AAPE and RMSE for the rate of penetration predicted using the SVR model were

**Table 6** The constants associated with the ROP correlations

Correlation	Constants
Maurer’s correlation	$k_M = 10,146,000$
Bingham’s correlation	$k_B = 0.339$ $a_B = -0.269$ $c_B = 0.636$
Bourgoyne and Young’s correlation	$a_1 = 6.534$ $a_2 = 0.0032$ $a_3 = -0.0022$ $a_4 = 0.0002$ $a_5 = -0.4789$ $a_6 = 0.9349$ $a_7 = 0$ $a_8 = 0.3314$

**Fig. 8** The actual and evaluated ROP for the 524 testing datasets



26.5% and 1.1 ft/hr, respectively. While the AAPEs for the ROP predicted with the Bingham, Maurer, and Bourgoyne and Young correlation are 51.0%, 47.6%, and 36.6%, while the RMSEs are 1.7, 2.0, and 1.3 ft/hr, respectively. On the other hand, the rate of penetration was predicted with very low AAPEs of 1.0%, 3.4%, and 8.2% and RMSEs of only 0.1, 0.2, and 0.4 ft/hr using (9), FIS-SC, and FNN models, respectively. These results confirmed the high accuracy of (9) and both FIS-SC and FNN models in evaluating the ROP while horizontally drilling sandstone formations.

### Summary and conclusions

In this study, the predictability of the oil well drillability while horizontally drilling sandstone formations using four artificial intelligence tools was evaluated. The following conclusions can be withdrawn:

1. The machine learning models were learned to assess the rate of penetration from only the surface measurable drilling parameters.

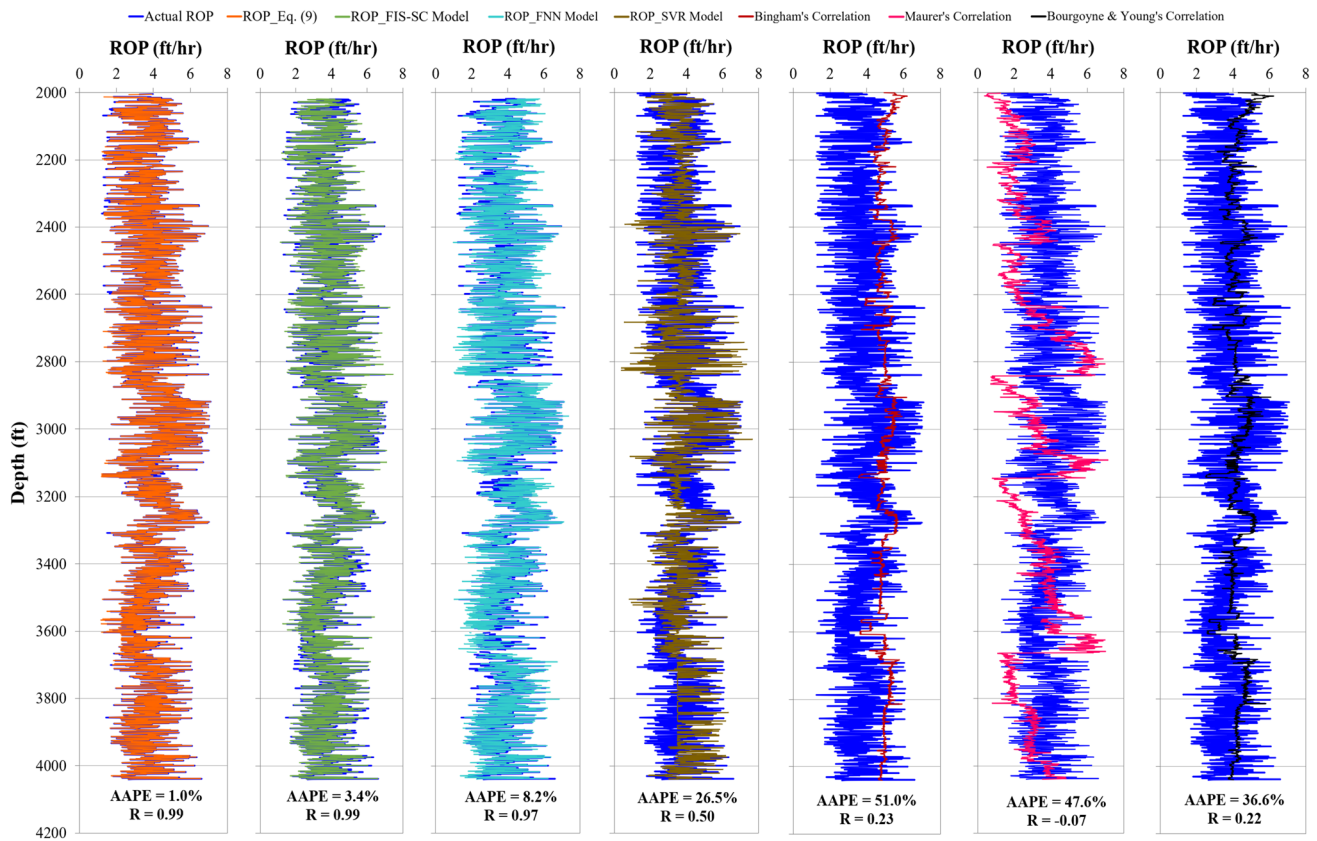
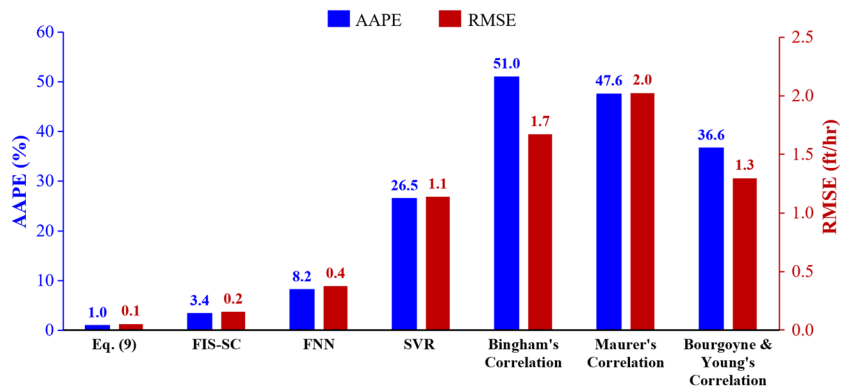


Fig. 9 Comparison of the ROP estimation in Well-Y (2213 data points) with various available models

Fig. 10 ROP accuracy comparison for Well-Y data using with various available models



- All machine learning models were firstly learned using 1224 datasets from Well-X. The learned artificial neural network model was then used to develop a correlation for the rate of penetration assessment.
- The developed empirical correlation and the optimized models' models were tested on 524 datasets from Well-X and validated on 2213 datasets from Well-Y.
- This research presented high accurate new correlation for ROP prediction that is machine learning-based that showed high degree of match with the actual ROP during the model training, testing, and validation phases.
- The new correlation and the optimized fuzzy inference system with subtractive clustering and functional neural network model do a great prediction over the other techniques based on the results obtained in this study.

Based on the research findings, the new correlation will add a great contribution regarding the rock drillability prediction and optimization for drilling oil and gas wells. The machine learning models will help for enhancing the drilling operation automation for cost savings and safe drilling operations.

**Acknowledgements** The authors wish to acknowledge King Fahd University of Petroleum and Minerals for providing the research facilities and permitting the publication of this work.

**Funding** This research received no external funding.

## Declarations

**Conflicts of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed SA, Elkhatny S, Abdurraheem A, Mahmoud M, Ali AZ, Mohamed IM (2018) Prediction of rate of penetration of deep and tight formation using support vector machine. In: the proceeding of the SPE kingdom of Saudi Arabia annual technical symposium and exhibition, Dammam, Saudi Arabia, April. <https://doi.org/10.2118/192316-MS>
- Akgun F (2002) How to estimate the maximum achievable drilling rate without jeopardizing safety. *Soc Pet Eng*. <https://doi.org/10.2118/78567-MS>
- Al-AbdulJabbar AM (2017) 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). <https://eprints.kfupm.edu.sa/id/eprint/140167>
- Al-AbdulJabbar A, Elkhatny 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>
- Al-AbdulJabbar A, Mahmoud AA, Elkhatny S (2021) Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations. *Arab J Geosci* 14:117. <https://doi.org/10.1007/s12517-021-06457-0>
- Al-AbdulJabbar A, Mahmoud AA, Elkhatny S, Abughaban M (2022a) 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. *J Sens* 2022:9444076. <https://doi.org/10.1155/2022/9444076>
- Al-AbdulJabbar A, Mahmoud AA, Elkhatny S, Abughaban M (2022b) Artificial neural networks-based correlation for evaluating the rate of penetration in a vertical carbonate formation for an entire oil field. *J Pet Sci Eng* 208, Part D:109693. <https://doi.org/10.1016/j.petrol.2021.109693>
- Al-AbdulJabbar A, Elkhatny S, Mahmoud M, Abdurraheem A (2018) Predicting rate of penetration using artificial intelligence techniques. In: the proceeding of the SPE Kingdom of Saudi Arabia annual technical symposium and exhibition, Dammam, Saudi Arabia, April. <https://doi.org/10.2118/192343-MS>
- Alsaihati A, Elkhatny S, Mahmoud AA, Abdurraheem 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*. <https://doi.org/10.1115/1.4048070>
- Anemangely M, Ramezanzadeh A, Tokhmechi B, Molaghab A, Mohammadian A (2018) Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *J Geophys Eng* 15(4):1146–1159
- Bahari MH, Bahari A, Moradi H (2011) Intelligent drilling rate predictor. *Int J Innov Comput Inf Control* 7(2):1511–1520
- Barbosa LFF, Nascimento A, Mathias MH, de Carvalho Jr JA (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>
- Bingham MG (1965) A new approach to interpreting rock drillability. Petroleum Publishing Co., Tulsa, Oklahoma
- Bourgoyne AT Jr, Millheim KK, Chenevert ME, Young FS (1991) Applied drilling engineering. *Soc Pet Eng Text Book Series* 2. ISBN: 978-1-55563-001-0
- Bourgoyne A, Young FS (1974) A multiple regression approach to optimal drilling and abnormal pressure detection. *Soc Petrol Eng J* 14:371–384. <https://doi.org/10.2118/4238-PA>
- Dupriest FE, Koederitz WL (2005) Maximizing drill rates with real-time surveillance of mechanical specific energy. In: the proceeding of the SPE/IADC drilling conference, Amsterdam, Netherlands. <https://doi.org/10.2118/92194-MS>
- Elkhatny S (2018) New approach to optimize the rate of penetration using artificial neural network. *Arab J Sci Eng* 43(11):6297–6304. <https://doi.org/10.1007/s13369-017-3022-0>
- Elkhatny 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(06):825–837. <https://doi.org/10.30632/PJV60N6-2019a7>
- Elzain HE, Chung SY, Senapathi V, Sekar S, Park N, Mahmoud AA (2021) Modeling of aquifer vulnerability index using deep learning neural networks coupling with optimization algorithms.

- Environ Sci Pollut Res 28:57030–57045. <https://doi.org/10.1007/s11356-021-14522-0>
- Gamal H, Elkhatatny S, Abdulraheem A (2020 November) Rock drillability intelligent prediction for a complex lithology using artificial neural network. In: paper presented at the Abu Dhabi international petroleum exhibition & conference, Abu Dhabi, UAE. doi: <https://doi.org/10.2118/202767-MS>.
- Hossain ME, Al-Majed AA (2015) Fundamentals of sustainable drilling engineering. Scrivener Publishing LLC, Beverly. <https://doi.org/10.1002/9781119100300.ISBN:9781119100300>
- Lyons WC, Plisga GJ (2004) Standard handbook of petroleum and natural gas Engineering, 2nd edn. Gulf Professional Publishing, Woburn. ISBN 10: 0750677856
- Mahmoud AA, Elkhatatny S, Al-Shehri D (2020b) Application of machine learning in evaluation of the static young's modulus for sandstone formations. Sustainability 12(5):1880. <https://doi.org/10.3390/su12051880>
- Mahmoud AA, Elzenary M, Elkhatatny S (2020c) New hybrid hole cleaning model for vertical and deviated wells. J Energy Resour Technol 142(3):034501. <https://doi.org/10.1115/1.4045169>
- Mahmoud AA, Elkhatatny S, Al-AbdulJabbar A, Moussa T, Gamal H, Shehri DA (2020a, June) Artificial neural networks model for prediction of the rate of penetration while horizontally drilling carbonate formations. In: Paper presented at the 54th U.S. Rock Mechanics/Geomechanics Symposium, June 2020a. Paper number: ARMA-2020a-1694
- Mahmoud AA, Elkhatatny S, Abouelresh M, Abdulraheem A, Ali A (2020d) 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>.
- Maurer W (1962) The “perfect-cleaning” theory of rotary drilling. J Petrol Technol 14(11):1270–1274. <https://doi.org/10.2118/408-pa>
- Mitchell RF, Miska SZ (2011) Fundamentals of drilling engineering. Society of Petroleum Engineers: Richardson, TX, USA. ASIN: B01L008WJA
- Najafzadeh M (2019) Evaluation of conjugate depths of hydraulic jump in circular pipes using evolutionary computing. Soft Comput 23:13375–13391. <https://doi.org/10.1007/s00500-019-03877-9>
- Osgouei RE (2007) Rate of penetration estimation model for directional and horizontal wells. In: Master's thesis, middle east technical university, Turkish
- Osman H, Ali A, Mahmoud AA, Elkhatatny S (2021) Estimation of the rate of penetration while horizontally drilling carbonate formation using random forest. J Energy Res Technol 143(9):093003
- Rabia H (2001) Well engineering and construction. Entrac consulting. ISBN-10 : 0954108701
- Saberi-Movahed F, Najafzadeh M, Mehrpooya A (2020) Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions. Water Resour Manag 34:529–561. <https://doi.org/10.1007/s11269-019-02463-w>
- Siddig O, Mahmoud AA, Elkhatatny S, Soupios P (2021) Utilization of artificial neural network in predicting the total organic carbon in devonian shale using the conventional well logs and the spectral gamma ray. Comput Intell Neurosci 2021:2486046. <https://doi.org/10.1155/2021/2486046>
- Teale R (1965) The concept of specific energy in rock drilling. Int J Rock Mech Min Sci Geomech Abstr 2(1):57–73. [https://doi.org/10.1016/0148-9062\(65\)90022-7](https://doi.org/10.1016/0148-9062(65)90022-7)
- Thanh HV, Lee KK (2022) Application of machine learning to predict CO<sub>2</sub> trapping performance in deep saline aquifers. Energy 239, Part E:122457. <https://doi.org/10.1016/j.energy.2021.122457>
- Thanh HV, Sugai Y, Sasaki K (2020) Application of artificial neural network for predicting the performance of CO<sub>2</sub> enhanced oil recovery and storage in residual oil zones. Sci Rep 10:18204. <https://doi.org/10.1038/s41598-020-73931-2>
- Thanh HV, Amar MN, Lee KK (2022a) Robust machine learning models of carbon dioxide trapping indexes at geological storage sites. Fuel 316:123391. <https://doi.org/10.1016/j.fuel.2022.123391>
- Thanh HV, Yasin Q, Al-Mudhafar WJ, Lee KK (2022b) Knowledge-based machine learning techniques for accurate prediction of CO<sub>2</sub> storage performance in underground saline aquifers. Appl Energy 314:118985. <https://doi.org/10.1016/j.apenergy.2022.118985>
- Warren TM (1987) Penetration rate performance of roller cone bits. Soc Pet Eng. <https://doi.org/10.2118/13259-PA>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.