SEMARAK ILMU
PUBLISHING

Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

Rate Of Penetration Prediction in Highly Deviated Wells Using Two Machine Learning Techniques; An Application in The Mediterranean Basin

Ahmed S. Ismail[1], Abdulaziz M. Abdulaziz[1*], Heba I. Alsaegh[2*♣], Abdel Sattar A. Dahab[1]

[1] Mining, Petroleum, Metallurgical Engineering Department, Cairo University, Giza, 12316, Egypt
[2] Department of Petroleum Engineering, College of Engineering, AlKitab University, Kirkuk, 36015, Iraq

ARTICLE INFO

Article history:
Received 3 September 2025
Received in revised form 12 March 2026
Accepted 31 March 2026
Available online 11 May 2026

ABSTRACT

Highly deviated and extended reach wells became a popular practice to achieve maximum contact to oil and gas reservoirs and solve numerous production challenges. This study presents an efficient and accurate ROP prediction method that can be applied in highly deviated wells. To accomplish this, artificial neural network (ANN) and random forest (RF) are utilized to predict the rate of penetration in highly deviated wells in Egyptian fields located in the Mediterranean Sea. A data set with 6000 points of common drilling parameters are subdivided into training, validation and testing data. These data include true vertical depth (TVD), flow in, mud weight, hole size, weight on bit (WOB), hook load (HL), standpipe pressure (SPP), torque, rotational speed (RPM), inclination angle and lithology. Using these data, several ANNs (using MemBrain software and Python) and RFs are tested to help predicting ROP efficiently. Comparing the results of different trials to the real reported ROP using the mean square error, the RF models showed two folds improvement in errors (RF= 0.003112; Python ANN= 0.0076287; MemBrain ANN= 0.0080767). Furthermore, RF enables combining numerous parameters in a simple structure and coding compared to ANN technique. To present a real well plan, a random set of hypothetical data is generated, and RF is applied for ROP prediction and alternatively a sensitivity analysis is performed. Results indicated that the developed RF model provided an efficient tool for ROP prediction at low cost that enables adjusting drilling parameters for optimum drilling performance.

Keywords:

Artificial Neural Network; Random Forest; ROP Prediction; Mediterranean Basin; Machine Learning

1. Introduction

Drilling optimization is accomplished using various techniques through different phases including planning of optimized drilling practices [1,2], well control [3], data analysis [4], predicting ROP [5,6], and other drilling issues [7-9] to reduce borehole problems and non-productive time (NPT). Some of these methods are qualitative while others are quantitative. Thus, optimization is based on specific conditions determining the drilling cost per foot that is highly affected by the achieved ROP [10].

---

[*] *Corresponding author.*
E-mail address:
amabdul@miners.utep.edu
♣ heba.i.addulmohsen@uoalkitab.edu.iq

https://doi.org/10.37934/araset.14.1.125138

Consequently, accurate forecasting of ROP controls optimized drilling. Numerous models have been proposed to relate ROP to different drilling parameters such as [11, 12] (d-exponent), Bourgoyne and Young Drilling Rate Model [13], Warren's model [14] and its modifications.

Maurer [11] developed an equation for cutter roller bits that relates ROP to bit size, rock strength, RPM, and WOB. The equation was based on single tooth impact with the formula presented in Eq. (1).

$$\mathrm{ROP} = \frac{\mathrm{K\,N(W-W_0)}}{\mathrm{{d_b}^2 S^2}} \tag{1}$$

Where:

K is constant of proportionality; N is the revolutionary speed (rpm); W is weight on bit ($\mathrm{lb_f}$); W_0 is the threshold weight on bit ($\mathrm{lb_f}$) representing the minimum weight on bit required to start drilling; d_b is the bit diameter (inch); S is rock strength (psi); ROP is rate of penetration (ft/h).

Jorden and Shirely [12] published an analytical model for predicting ROP that involves the well-known d-exponent shown in Eq. (2)

$$ROP = kW^d\,\mathrm{N} \tag{2}$$

Where: k is the drill-ability of the formation (unitless), d is the exponent for weight on bit (unitless).

Bourgoyne and Young [13] developed a model that relate the ROP to numerous drilling parameters including formation strength, compaction, overbalance, mud type, solid content, rotary speed, bit type, bit tooth wear and bit hydraulics (Equation 3). This model is proposed for predicting ROP for a drill string with roller cone bits.

$$ROP = f_1 * f_2 * f_3 * f_4 * f_5 * f_6 * f_7 * f_8 \tag{3}$$

Where:

$f_1 = e^{2.303a_1}$ involves all parameters not included the model function.

$f_2 = e^{2.303a_2(10000-D)}$ is the formation normal compaction function.

$f_3 = e^{2.303a_3D^{0.69}(g_p-9)}$ is the formation compaction function.

$f_4 = e^{2.303a_4D(g_p-p_c)}$ is the differential pressure of bottom hole function.

$f_5 = \left[\frac{\left(\frac{W}{d_b}\right)-\left(\frac{W}{d_b}\right)_t}{4-\left(\frac{W}{d_b}\right)_t}\right]$ is the bit weight and diameter function.

$f_6 = \left(\frac{N}{60}\right)^{a_6}$ is the rotary speed function.

$f_7 = e^{a_7h}$ is the tooth wear function.

$f_8 = \left(\frac{f_j}{1000}\right)^{a_8}$ is the hydraulic function.

a_1 to a_8 are constants based on local drilling conditions; D is the TVD (ft); g_p is the pore pressure gradient ($\mathrm{lb_m}$/gal); $\left(\frac{W}{d_b}\right)_t$ is the threshold weight on bit per inch of bit diameter ($\mathrm{lb_f}$/in); h is the bit tooth dullness.

As an alternative and approximate solution, machine learning techniques are used to capture the well scenario. Neural network has been successfully used in various fields due to their ability to identify complex relationships when sufficient data is available [15, 16]. Artificial neural network (ANN) can combine different parameters to predict various situations [17, 18]. ANN can significantly enhance the ability to predict ROP accurately even with the changes in lithology, bit type, hole size, drilling parameters and mud properties. The flexibility of this method permits analysing an extensive range of information and carry superior ROP prediction [19, 20]. Dashevskiy *et al.* [21] presented an approach that uses neural network for modelling the dynamic behaviour of a non-linear input/output drilling system. They suggested adding an optimizing controller to the model to provide recommended quantified corrective actions required for optimal drilling conditions. The parameters that affect performance are classified into three categories including control parameters, plant characteristics and media parameters. Bataee *et al.* [22] studied optimizing the drilling parameters, predicting ROP, estimating the drilling time and reducing the cost of the drilling process using ANN. Their study was based on data from the Shadegan oil field in Iran. The developed model included bit size, mud weight, WOB, depth and RPM as inputs while the output was set as the ROP. The study resulted in practical ranges and guidelines for different parameters. Monazami *et al.* [23] applied ANN method using MATLAB for accurate ROP estimation for an Iranian oilfield. The input parameters involve mud properties, drill string & BHA size, operation parameters, borehole geometry information and formation characteristics. Similarly, Jahanbakhshi *et al.* [19] used ANN model to investigate and predict ROP in an oilfield in South Iran. Offset wells drilling parameters (operation parameters, mud properties, bit data, formation characteristics, hydraulics, and others like hole size) are used to predict the new wells ROP and the model has proven its efficiency as a real-time tool. Several attempts are made to determine the number of hidden layers and number of hidden neurons in each layer (trial and error) to optimize ANN structure and alternatively achieve minimum errors. Duan *et al.* [24] presented an ROP optimization method based on back propagation (BP) neural network and particle swarm optimization (PSO) algorithm. This involves constructing a prediction model for a target well in Yuanba, China from well logs using BP neural network, and delineating the optimized well operating parameters through PSO algorithm. Manshad *et al.* [25] used multi–layer perception (MLP) neural networks for optimizing the drilling operation with two models to determine the type of bit and ROP. Alternatively, the inputs of the second model were optimized by a genetic algorithm (GA) for achieving maximum ROP. Testing these models showed has high prediction accuracy. Elkatatny [26] evaluated the effect of the drilling parameters, drilling fluid properties and formation strength on ROP and developed a new ANN model to estimate the ROP as a function of drilling parameters and fluid properties using 3333 data points. The developed model was able to predict ROP with high accuracy (R of 0.99 and AAPE of 5.6%). The high accuracy of the developed correlation (AAPE of 4%) confirmed the importance of compiling the drilling parameters and the drilling fluid properties.

Hegde *et al.* [27] used statistical learning (machine learning) methods including trees, bagging, and random forest (RF) to predict ROP. Trees enabled a great visualization of the data but with limited accuracy and can result in considerable overfitting. However, the results were generally adequate for real time prediction. The limitation can be treated using bagging or RF to increase accuracy. Using trees and RF showed advantages over ANN in computational efficiency accuracy and evaluation of the input relative importance. Among the three techniques including trees, bagging and RF, RF showed the least RMSE for all formations. All these works demonstrated the feasibility of using artificial intelligence to control drilling dynamics and predict the optimum ROP. In the present study ANN and RF are tested to develop a real time prediction model for ROP in highly deviated wells in the Mediterranean basin using simple drilling parameters available on rig. These two methods provide nonlinear approximation techniques that utilize various function to best fit the input data to the target output over training and optimization runs. A comparison of the resulting models verifies consistency and ensures noise reduction and optimum model development.

2. Methodology and Data Set

The data set used in the present study is mostly raw data available at mud logging unit in addition to lithology and inclination. These data include TVD, flow in (volumetric rate in), mud weight, hole size (bit size), weight on bit (WOB), hook load (HL), standpipe pressure (SPP), torque, rotational speed (RPM), inclination and lithology. The available data is approximately 6000 data points reported in highly deviated sections only. These data are subdivided randomly into three categories with 70% for training, 20 % for validation and 10 % for testing. First, a simple graphical analysis is used to identify the dominant operating range for different controllable parameters in various borehole sizes. Then, machine learning is applied to generate an accurate model to predict optimum ROP.

Machine learning tools, ANN and RF of various structures, are used for optimizing ROP prediction in highly deviated wells (inclination ≥ 60°) by comparing their results to the real reported ROP. The numbers of hidden layers, neurons, and estimators are changed for ANN and RF, respectively. In ANN structure, the hyper parameters to change involve the number of hidden layers, number of hidden neurons in each layer and the teaching algorithm [28,29]. On the other hand, the number of estimators is the only hyper parameter to change for the RFs. Numerous structures are developed for both ANNs and RFs and meanwhile the results are compared to real data to select the optimum prediction model. This optimum model is selected based on the calculated mean square error (MSE) values. Figure 1 presents a sketch of the methodology applied in the present study. Two different tools, MemBrain software [30] and Python programming language [31] are applied to construct the ANN models while RF models are constructed using Python programming. After identifying the optimum machine learning model and evaluating its performance using the test data, the model would be ready for predicting ROP in highly deviated wells to help planning new wells. A random hypothetical data set is generated using practical criteria of various alternatives for drilling parameters available for choice. The purpose of this step is to simulate a real case where a well designer gets an estimate to ROP for specific drilling parameters using the developed model. In addition, a sensitivity study to several important drilling parameters is accomplished to determine their influence on the optimum ROP.

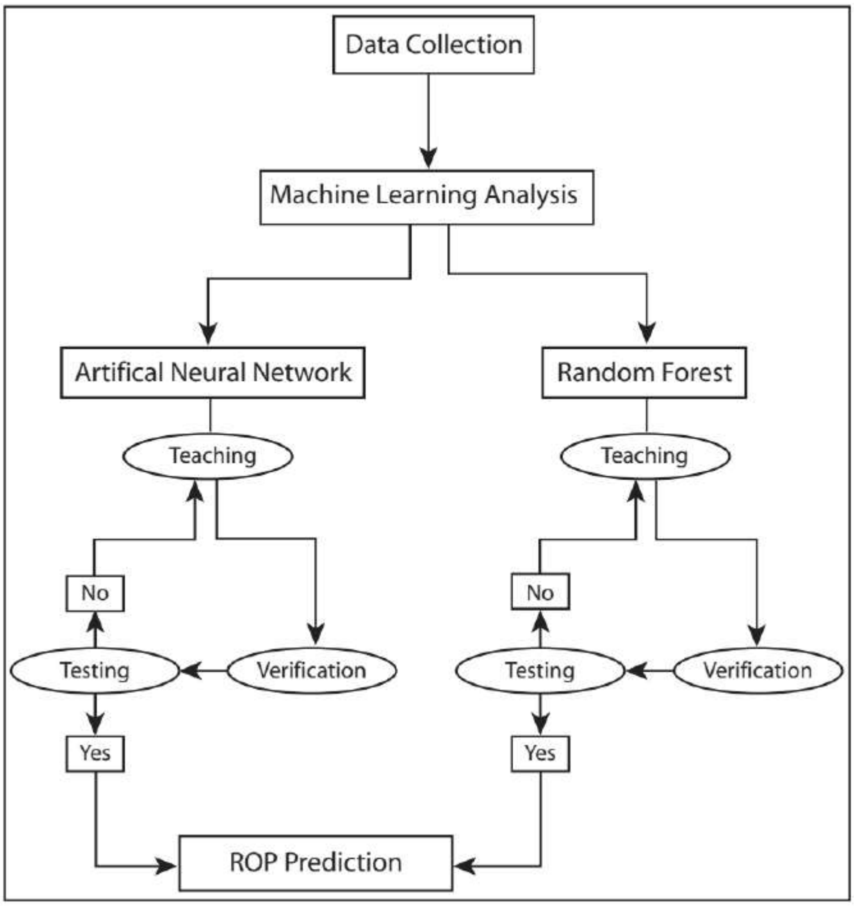


Fig. 1. A flow chart for the applied Methods

3. Results and discussion

The available drilling data of three wells with several side tracks have been analyzed graphically and the data range of dominant operating parameters together with the corresponding ROP are presented for 8.5-inch and 12.25-inch holes in Tables 1 and 2, respectively. Thus, such dataset could present a representative set of values that involve the drilling parameters for optimum ROP in deviated wells. Several ANN and RF models of different structures are developed using these drilling parameters. For the ANN constructed by MemBrain, several back-propagation teachers are tested, and the results of these trials are shown in Table 3. The effect of lithology is evaluated using several trials (Table 3). Lithology is fed to ANN either through three neurons using one binary value (i.e. 0s and 1s) for each neuron or only a single neuron with five binary digits.

Table 1
The dominant operating parameters ranges for highly deviated section in 8.5 " holes.

Parameter	Optimum range
Flow in (gpm)	425-440
WOB (mton)	NA
HL (mton)	95-105
SPP (psi)	2250-2500
Torque (lb.ft)	7600-10150
RPM	115-120

Table 2
The dominant operating ranges of drilling parameters in highly deviated 12.25" holes.

Parameter	Optimum range
Flow in (gpm)	750-830
WOB (mton)	NA
HL (mton)	85-115
SPP (psi)	3100-4250
Torque (lb.ft)	6000-10500
RPM	120-145

Table 3
The MSE values for ANNs developed using MemBrain.

Lithology	Teacher	Number of hidden layers	No. of neurons in 1st layer	No. of neurons in 2nd layer	MSE
Not applied	RPROP	1	3	☒	0.0107244
Not applied	BP-momentum	1	3	☒	0.012903
Not applied	RPROP	1	4	☒	0.01041395
Not applied	RPROP	1	5	☒	0.01538221
Start using the lithology as input parameters					
3 neurons	RPROP	1	3	☒	0.01017448
3 neurons	BP-momentum	1	3	☒	0.0117187
3 neurons	RPROP	1	4	☒	0.0098912
3 neurons	RPROP	1	5	☒	0.00956825
3 neurons	RPROP	1	6	☒	0.01798139
3 neurons	RPROP	1	7	☒	0.00943913

3 neurons	RPROP	1	8	☒	0.0080767
3 neurons	RPROP	1	9	☒	0.00897663
3 neurons	RPROP	1	10	☒	0.00921613
3 neurons	RPROP	1	11	☒	0.00813411
3 neurons	RPROP	1	13	☒	0.008471643
3 neurons	RPROP	2	3	2	0.016676194
3 neurons	RPROP	2	5	2	0.00876026
3 neurons	RPROP	2	10	2	0.00888365
3 neurons	Cascade-RPROP				0.05095855
3 neurons	Cascade-BP				0.575363
1 neuron, 5 binary digits	RPROP	1	3	☒	0.00955239
1 neuron, 5 binary digits	RPROP	1	4	☒	0.0248158

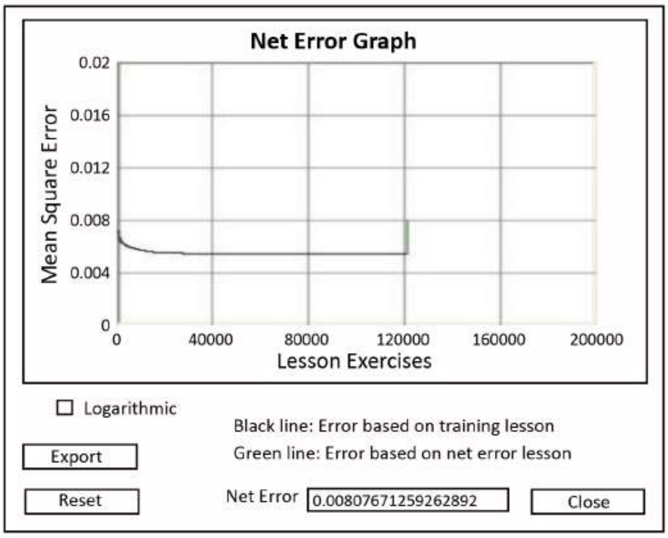
Table 4 presents the lithology encoding terms used in three neurons or single neuron techniques. The three neurons method appeared practical and showed better results compared to the single neuron technique as indicated by the lower MSE (0.0098912 for the three neurons technique and 0.0248158 for the single neuron technique). Thus, the three neurons technique is adopted to complete the ANN structure. The different ANNs adopted several teachers including BP with momentum, RPROP, Cascade correlation RPROP and Cascade correlation BP (Table 3). Table 3 also shows the number of hidden layers together with the number of neurons in the first and second hidden layers. For the same network structure, the performance of RPROP teacher has given slightly better results compared to BP with momentum as indicated by MSE; 0.0107244 for RPROP teacher versus 0.012903 for BP (Table 3). Cascade correlation is used twice, one was BP and the other was RPROP, and the results are not generally convenient compared to other teachers (Table 3).

Table 1
The Lithology encoding techniques applied in the present study

Lithology	Company's index	1 neuron 5 binary code	3 neurons, single digit code
Sand/ Sandstone	20	00010	0-0-0
Clay	31	00100	0-1-0
Shale	28	00001	1-0-0
Siltstone	30	10000	1-1-0

Of the various ANN architectures and the different teachers, the best ANN structure with the lowest MSE (0.00550776) comprised one hidden layer with 8 neurons trained with RPROP teacher and the lithology is considered as an input parameter using 3 neurons. Figure 2 shows the net MSE throughout learning and testing processes for the best reported ANN structure. The testing data is further investigated through plotting the predicted ROP versus the corresponding actual value to evaluate the performance of ROP prediction (Fig. 3). The optimum prediction occurs where both predicted and actual ROP values are identical and such a case is presented with the dotted line in Fig. 3 and the deviation of this trend indicates the miss-prediction.

To validate the performance of the techniques applied for constructing ANN models and at the same time search for a better ANN model to predict ROP in highly deviated wells, new ANN models are constructed using Python programming language using the same data set. Table 5 presents the resulting MSE obtained from Python for each ANN structure using training and validation data for some of the network structures. Since back-propagation algorithm showed the best performance in MemBrain trials, it has been used in training the networks developed by Python. In addition, the lithology is fed into the neural networks through three neurons as previously described in MemBrain models. Of these numerous trials, the best achieved MSE (0.004856) is reported for the network made of a single hidden layer with 18 neurons. The other ANN structures developed using Python varied in the calculated MSE between 0.004976 and 0.01162. Once the optimum structure ANN is determined, the eighteen-neuron single layer ANN is tested using 600 data points. Generally, the MSE for this network for the test data reported 0.0076287 that indicated a slight improvement compared to the optimum model developed by MemBrain, 0.0080767. This slight enhancement in MSE of the test data is probably because of cross validation and/or the capability of the model used to represent the available data set.



Black line: Error based on training lesson
Green line: Error based on net error lesson

Net Error 0.00807671259262892

Fig.2. The Net Error calculated for the training and validation data

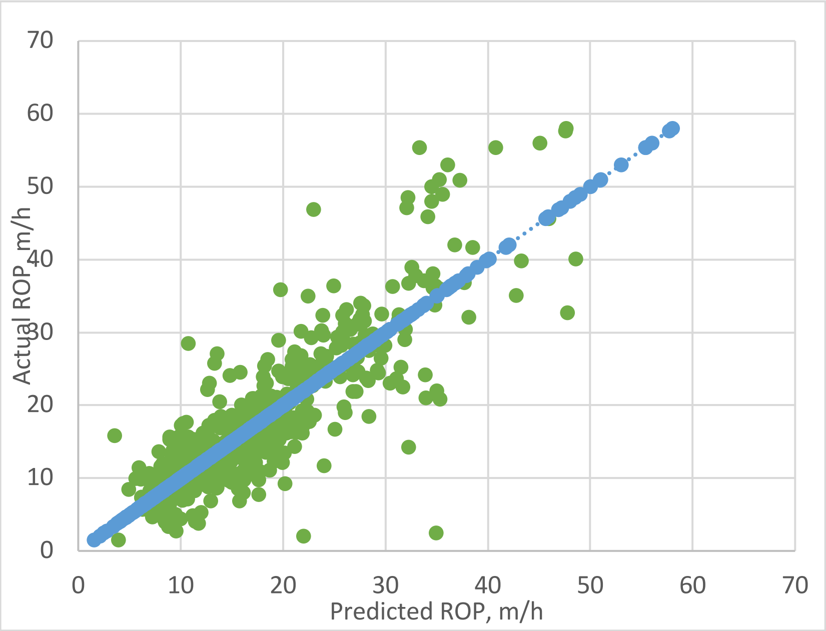


Fig. 3. A cross plot of the predicted ROP using MemBrain versus the corresponding actual value

Table 5
The MSE values for a sample ANNs developed using Python.

Number of hidden layers	No of neurons in the 1st layer	No of neurons in the 2nd layer	MSE
1	1	☒	0.006897
1	2	☒	0.01126
1	3	☒	0.009227
1	4	☒	0.007336
1	5	☒	0.010402
1	6	☒	0.008521
1	7	☒	0.00968
1	8	☒	0.009326
1	9	☒	0.007068
1	10	☒	0.006892
1	11	☒	0.00732
1	12	☒	0.005533
1	13	☒	0.005887
1	14	☒	0.007103
1	15	☒	0.005902
1	16	☒	0.00568
1	17	☒	0.005325
1	18	☒	0.004856
1	19	☒	0.005568
1	20	☒	0.004921
2	1	1	0.007395

Random forest is applied to the same data sets to predict ROP in deviated wells to compare its results to the results obtained by ANN techniques and confirm the validity of prediction. As random forest learns the available data set, prediction performance is evaluated using MSE. These calculated MSEs are plotted against the corresponding value of estimators (trees) (Fig. 4). The MSE starts to stabilize around 0.003 after using 10 estimators that appears match-able to the reported average value (0.00316). In addition, the minimum MSE is calculated for 30 estimators with a calculated value of 0.002772708 reported for training and validation data. When the testing data set is used, it results in a MSE of 0.003112. Figure 5 shows the predicted ROP using the selected RF from python vs the actual ROP for the test data. As shown in Fig. 5, the predicted ROP is windowed around that line which shows small differences between the predicted and the actual values than ANN (Fig. 3) indicating better data match and successful predictions. On comparing the performance of the different prediction methods using MSE as an indicator (RF= 0.003112; Python ANN= 0.0076287; MemBrain ANN= 0.0080767); RF shows two folds improvement in the calculated MSE for the difference between predicted and actual ROP data. In addition, RF has the advantage of being simple in structure and coding that enables incorporating several variables without the complications associating ANN design. The code of the optimum RF and ANNs models that achieved the best RME using Python is available in Appendix 1.

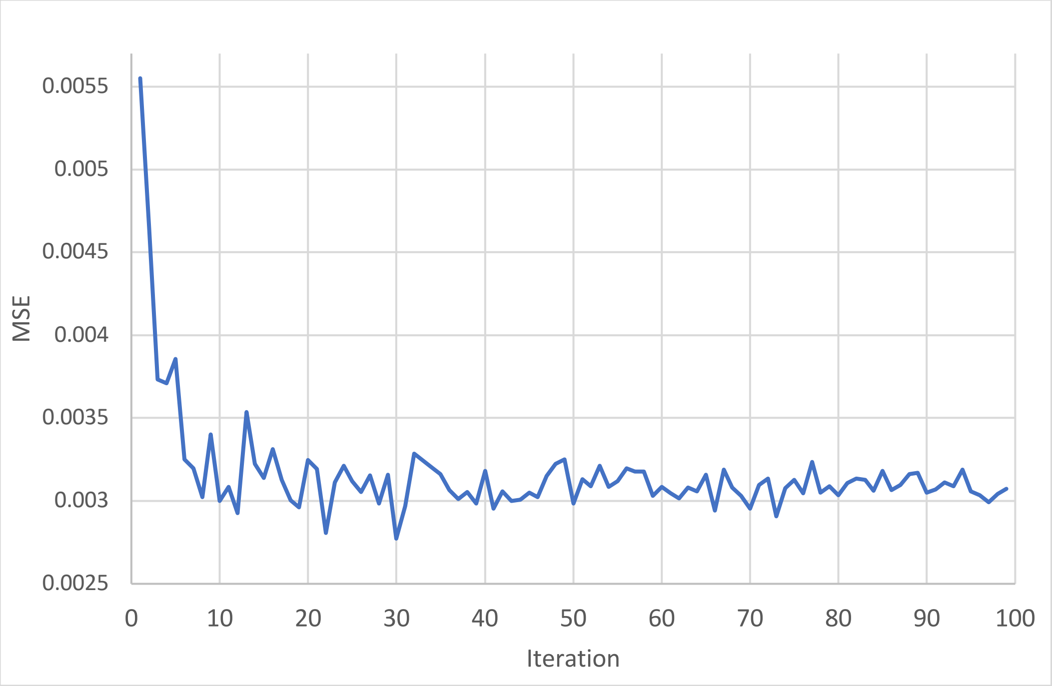


Fig. 4. The MSE for various RFs developed using Python

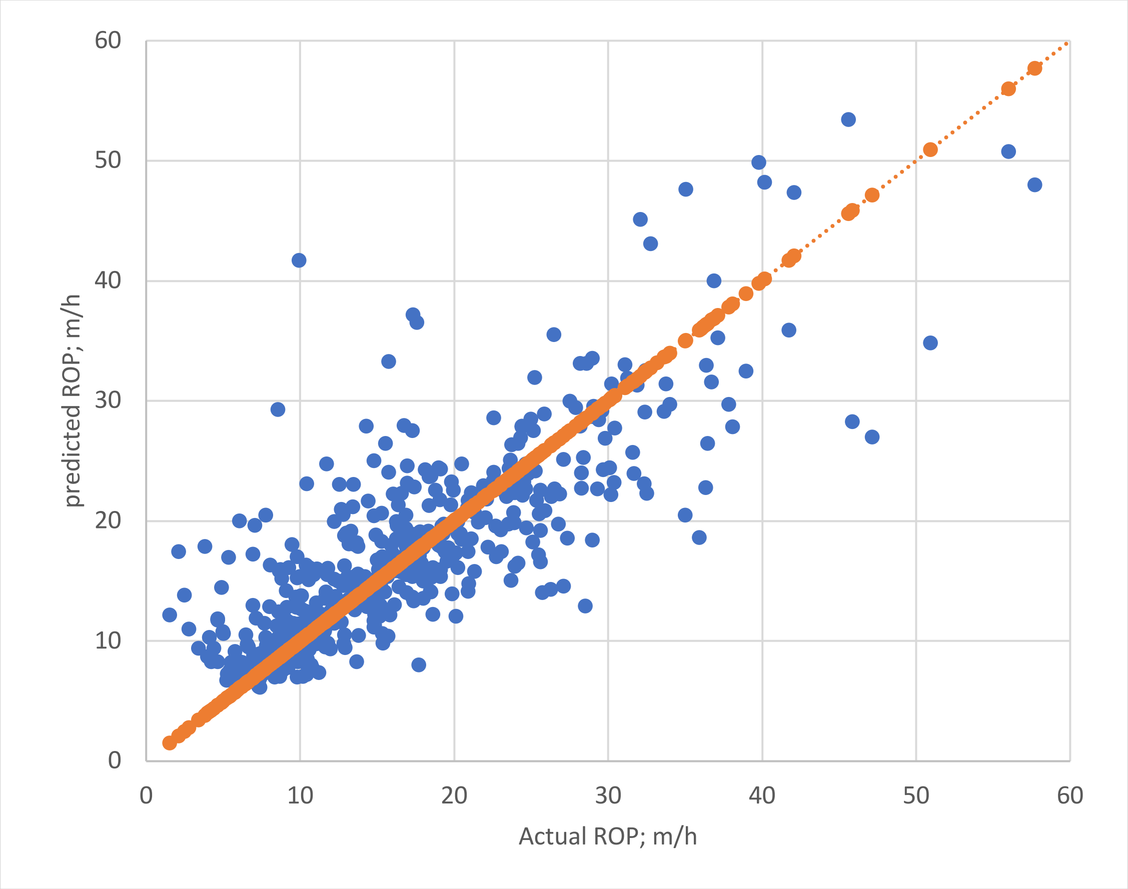


Fig. 5. A cross plot to the predicted ROP from optimum RF model versus actual ROP

Generally, the data points of ROP below 30 m/h typically deviates ± 3 m/h away of the true value but as the ROP increases the prediction error increases to report up to ±10 m/h for ROP of 50 m/h. Such an increase in prediction errors with the ROP increase is typically related to the nature of training data as most ROP values are usually below the 25 m/h. The typical ROP rarely increases above 40 m/h. Such results indicate that the developed models can be used to predict ROP in highly deviated wells with an acceptable accuracy that in most cases fall close to 90%.

For designing a well program, there are unlimited number of drilling parameters to be assigned. Therefore, simplified constraints have been assigned to enable ROP calculations swiftly within practical and achievable ranges of drilling parameters. Using these synthetic data as input parameters to the verified and tested RF model and based on the different combinations, various ROP is calculated. Table 6 lists sample values for both input variable drilling parameters and the resulting ROP for each drilling parameters combination in sandstone and shale, considering layer-cake formation architecture. Generally, the comparison between sand and shale ROPs shows strong correlation and such correlation is best achieved at lower and middle ROP values (up to 20 m/h) with minimal difference that is usually less than 1 m/h. However, as the ROP goes above 30 m/h, a marked deviation between ROP shale and ROP sandstone is depicted and the difference is usually over 5 m/h with ROP in shale dominantly higher than that ROP calculated for sandstone. In all cases the correlation between ROP shale and ROP sandstone is within Pearson correlation coefficient of almost 0.9. Considering the high ROP category with the dominant depth around 2700 m, the shift in this category can be interpreted by presence of shale in more data points of greater depth in the training data. The sensitivity analyses for both depth and lithology showed that both depth and lithology has minor effect on ROP in the available data ranges. However, WOB and SPP showed significant effect on ROP prediction while torque and RPM had moderate effect. Sandstone reported higher ROP compared shale under the same drilling conditions that match real drilling conditions indicating successful model performance. Generally, the data points follow the normal trend of inverse relationship between ROP and depth. The shift from this general trend shown at shallow depth (Table 6) is probably attributed to lack of training data at these depths. Despite the successful performance using the present input data, ROP prediction can be enhanced by including other parameters such as bit type, BHA type and size, bit wear, and hydraulics

Table 6
The predicted ROP values with the corresponding variable parameters values for 8.5-inch hole.

TVD	WOB	SPP	Torque	RPM	Inclination	ROPss	ROPsh
2700	25	3510	8487	140	69	54.25212	46.50094
2700	19	2300	10724	59	73	32.68512	42.84704
2700	21	2560	15160	130	62	25.1173	26.9659
1700	8	1557	14414	142	68	20.19622	19.88627
1700	5	2812	12915	105	71	14.5773	16.66881
2000	21	2560	15160	130	62	13.57457	14.43699
2000	25	3510	8487	140	69	6.629988	6.312067
1700	1	1450	6135	50	58	5.945333	5.456757

4. Conclusions

In the present study different artificial intelligence tools and trials have been applied to predict the rate of penetration in highly deviated wells in the Mediterranean basin. Both Artificial Neural Network and Random Forest are proved to be efficient tools for ROP predication from surface available data at high accuracy and low cost. In a general sense, a simple model is always recommended for ANN and RF structures before considering complicated ones. Random Forest provided the best result for the available data set and proved capable of overcoming some of the limitations in ANN techniques such as complexity of the training process and the volume of the data needed. Generally, the thirty-tree RF model shows high accuracy when 5 ≤ ROP ≤ 30 m/h due to the availability of numerous data points in this range. The applicability of ML models proved to be sensitive to the dynamic range of the teaching data compared to the teaching technique. In addition, ROP prediction can be enhanced by including other parameters such as bit type, BHA type and size, bit wear, and hydraulics. Furthermore, extending the data base (intensity & range) from additional wells at the same basin would significantly improve the prediction quality.

Acknowledgement

The authors of this work express a great gratitude and appreciation to the Egyptian Petroleum Corporation for providing the dataset that made this work possible. A special thanks goes to Mr. Mohammed El-Razzaz for his cooperation and Mr. Thomas Jetter for the technical support.

References

[1] Dahab, A. S., A. M. Abdulaziz, A. Abbas, A. K. Al-Husseini, and N. Manhalawi. 2020. “Managing Wellbore Instability through Geomechanical Modelling and Wellbore Stability Analysis.” Paper presented at the 54th U.S. Rock Mechanics/Geomechanics Symposium.

[2] Abdulaziz, A. M., H. L. Abdulridha, A. S. A. Dahab, S. Alhussainy, and A. K. Abbas. 2021. “3D Mechanical Earth Model for Optimized Wellbore Stability: A Case Study from South of Iraq.” *Journal of Petroleum Exploration and Production Technology* 11 (9): 3409–3420. https://doi.org/10.1007/s13202-021-01255-6.

[3] Manhalawi, A., A. S. Dahab, A. M. Abdulaziz, A. K. Abbas, and N. Al-Husseini. 2020. “Wellbore Stability Evaluation for Depleted Reservoir.” In *54th U.S. Rock Mechanics/Geomechanics Symposium*, 5:3828–3834. Golden, CO, USA.

[4] Abdulridha, H. L., A. M. Abdulaziz, A. A. Khalil, S. Alhussainy, A. S. A. Askar, A. S. A. Dahab, and D. Alfarge. 2022. “Study on Uncertainty Analysis for Drilling Engineering Applications: Wellbore Stability Assessments.” *Arabian Journal for Science and Engineering* 47 (9): 11687–11698. https://doi.org/10.1007/s13369-021-06389-7.

[5] Bataee, M., and S. Mohseni. 2011. “Application of Artificial Intelligent Systems in ROP Optimization: A Case Study.” Paper presented at the SPE Middle East Unconventional Gas Conference and Exhibition, Muscat, Oman.

[6] Abdelhak, A. I., A. F. Ghanem, A. M. Abdulaziz, K. I. Kabel, A. Z. Noah, A. M. Youssef, and A. S. A. Dahab. 2022. *Egyptian Journal of Chemistry* 65 (9): 349–360. https://doi.org/10.21608/EJCHEM.2022.111764.5075.

[7] El Neiri, M. H., A. S. A. Dahab, and A. M. Abdulaziz. 2013. “The Dynamic Underbalanced Drilling: A New Drilling Technique.” Paper presented at SPE/IADC Middle East Drilling Technology Conference and Exhibition. https://doi.org/10.2118/178153-MS.

[8] Abdulaziz, A. M., H. A. Mahdi, and M. H. Sayyouh. 2019. “Prediction of Reservoir Quality Using Well Logs and Seismic Attributes Analysis with an Artificial Neural Network: A Case Study from Farrud Reservoir, Al-Ghani Field, Libya.” *Journal of Applied Geophysics* 161: 239–254. https://doi.org/10.1016/j.jappgeo.2018.09.013.

[9] Gamal, H., S. Elkatatny, and A. M. Abdulaziz. 2024. “Intelligent Solution for Auto-Detecting Lithology Scheme While Drilling by Machine Learning.” Paper presented at IPTC 2024. https://doi.org/10.2523/IPTC-24535-MS.

[10] Abdulridha, H. L., A. M. Abdulaziz, A. S. A. Dahab, S. O. Redywi, and A. S. A. Askar. 2022. “3D Geomechanical Modelling through Geostatistics, and Wellbore Stability Analysis: A Case Study from South of Iraq.” *AIP Conference Proceedings* 2443: 30018. https://doi.org/10.1063/5.0091853.

[11] Maurer, W. C. 1962. “The ‘Perfect-Cleaning’ Theory of Rotary Drilling.” *Journal of Petroleum Technology* 14 (11): 1270–1274. https://doi.org/10.2118/408-PA.

[12] Jorden, J. R., and O. J. Shirley. 1966. “Application of Drilling Performance Data to Overpressure Detection.” *Journal of Petroleum Technology* 18 (11): 1387–1394. https://doi.org/10.2118/1407-PA.

[13] Bourgoyne, A. T., K. K. Millhem, M. E. Chenevert, and F. S. Young. 1991. *Applied Drilling Engineering*. 9th ed. SPE Textbook Series 2. Richardson, TX.

[14] Warren, T. M. 1987. “Penetration Rate Performance of Roller Cone Bits.” *SPE Drilling Engineering* 2 (1): 9–18. https://doi.org/10.2118/13259-PA.

[15] Habib, M. A., A. M. Abdulaziz, and A. A. Dahab. 2016. “Improving History Match Using Artificial Neural Networks.” *International Journal of Petroleum Engineering* 2 (4): 302–319. https://doi.org/10.1504/IJPE.2016.084121.

[16] Abdulaziz, A. M., and S. S. Hawary. 2020. "Prediction of Carbonate Diagenesis from Well Logs Using Artificial Neural Network." *Ain Shams Engineering Journal* 11 (4): 1387–1401. https://doi.org/10.1016/j.asej.2020.01.010.

[17] Salem, K. G., A. M. Abdulaziz, and A. A. Dahab. 2019. "Prediction of Hydraulic Properties in Carbonate Reservoirs Using Artificial Neural Network." Paper presented at ADIPEC. https://doi.org/10.2118/193007-MS.

[18] Hassaan, S., A. M. Abdulaziz, A. F. Ibrahim, and S. Elkatatny. 2024. "Real-Time Prediction of Petrophysical Properties Using Machine Learning Based on Drilling Parameters." *ACS Omega* 9 (15): 17066–17075. https://doi.org/10.1021/acsomega.3c08795.

[19] Jahanbakhshi, R., R. Keshavarzi, and A. Jafarnezhad. 2012. "Real-Time Prediction of Rate of Penetration during Drilling Operation in Oil and Gas Wells." Paper presented at the 6th U.S. Rock Mechanics/Geomechanics Symposium.

[20] Abdulaziz, A. M., and M. J. William. 2023. "Hydraulic Flow Units (HFUs) Prediction Using Integrated Probabilistic Neural Network and Seismic Attributes." Paper presented at ICEG 2023. https://doi.org/10.1190/iceg2023-084.1.

[21] Dashevskiy, D., V. Dubinsky, and J. D. Macpherson. 1999. "Application of Neural Networks for Predictive Control in Drilling Dynamics." Paper presented at SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/56442-MS.

[22] Bataee, M., S. Edalatkhah, and R. Ashna. 2010. "Comparison between Bit Optimization Using Artificial Neural Network and Other Methods." Paper presented at International Oil and Gas Conference and Exhibition in China. https://doi.org/10.2118/132222-MS.

[23] Monazami, M., A. Hasemi, and M. Shahbazian. 2012. "Drilling Rate of Penetration Prediction Using Artificial Neural Network." *Oil and Gas Business*.

[24] Duan, J., J. Zhao, L. Xiao, C. Yang, and C. Li. 2015. "A ROP Optimization Approach Based on Improved BP Neural Network PSO." In *Advances in Swarm and Computational Intelligence*, edited by Y. Tan et al., 9142:xxx–xxx. Cham: Springer.

[25] Manshad, A., H. Rostami, H. Toreifi, and A. Mohammadi. 2017. "Optimization of Drilling Penetration Rate in Oil Fields Using Artificial Intelligence Technique." In *Heavy Oil: Characteristics, Production and Emerging Technologies*, 255–269. Nova Science Publishers.

[26] Elkatatny, S. 2018. "New Approach to Optimize the Rate of Penetration Using Artificial Neural Network." *Arabian Journal for Science and Engineering* 43: 6297–6304. https://doi.org/10.1007/s13369-017-3022-0.

[27] Hegde, C., S. Wallace, and K. Gray. 2015. "Using Trees, Bagging, and Random Forests to Predict Rate of Penetration during Drilling." Paper presented at SPE Middle East Intelligent Oil and Gas Conference and Exhibition. https://doi.org/10.2118/176792-MS.

[28] Abdulaziz, A. M., and H. I. Alsaegh. 2022. "Integration of Seismic Attributes with Well Logs Using Artificial Neural Network." *Arabian Journal of Geosciences* 15: 1407. https://doi.org/10.1007/s12517-022-10440-8.

[29] Abdulaziz, A. M. 2020. "The Effective Seismic Attributes in Porosity Prediction for Different Rock Types." *Egyptian Journal of Petroleum* 29 (1): 95–104. https://doi.org/10.1016/j.ejpe.2019.12.001.

[30] Jetter, T. 2014. *MemBrain Neural Network Simulator Software Manual*. Version 08.00.00.00.

[31] "Python." n.d. Accessed June 21, 2023. https://www.python.org/doc/essays/blurb/.

Appendix 1:

The code developed for testing the optimum ANN and RF models using python.

```
from sklearn.model_selection import cross_val_predict
from sklearn.tree import export_graphviz
from sklearn.model_selection import cross_val_score
import csv
from sklearn.externals import joblib
from time import time
from  sklearn import metrics
from sklearn import preprocessing
from sklearn.neural_network import MLPRegressor
from  sklearn.ensemble import RandomForestRegressor

def ReadTData(dir):
    with open(dir) as tsvfile:
        tsvreader = csv.reader(tsvfile, delimiter="\t")
        allData=[]
        counter=0
        for line in tsvreader:
            allData.append([float(i) for i in line[0:len(line)]])

    Inputs=[]
    Outputs=[]
    for row in allData:
        x=row[0:len(row)-1]
        yy=row[len(row)-1]
        Inputs.append(x)
        Outputs.append(yy)
    return Inputs,Outputs



TrainingInputs,TrainingLables=ReadTData("D:\\masters\\Thesis\\Data\\\work\Python\\data\\train.txt")
TestInputs,TestLables=ReadTData("D:\\masters\\Thesis\\Data\\work\\Python\\data\\validation.txt")


min_max_scaler = preprocessing.MinMaxScaler()
minmaxTrainingInputs = min_max_scaler.fit_transform(TrainingInputs)
minmaxTestInputs = min_max_scaler.transform(TestInputs)


min_max_scaler = preprocessing.MinMaxScaler()
traininglablesBeforeNorm= TrainingLables
TrainingLables = min_max_scaler.fit_transform(TrainingLables)
TestLables = min_max_scaler.transform(TestLables)

#clf = MLPRegressor(hidden_layer_sizes=(18),activation='relu')
clf =RandomForestRegressor(n_estimators=30)
scores = cross_val_score(clf, minmaxTrainingInputs, TrainingLables, scoring='mean_squared_error',cv=10)
print("Accuracy: %0.5f (+/- %0.5f)" % (scores.mean(), scores.std() * 2))
print scores

pred = cross_val_predict(clf, minmaxTrainingInputs, TrainingLables, scoring='mean_squared_error',cv=10)
```

```
testbeforeNorm=TestLables
predBeforeNorm=pred
pred = min_max_scaler.transform(pred)


f=open('D:\\masters\\Thesis\\Data\\work\\Python\\data\\Results.txt',"w")

i=0
for prediction in predBeforeNorm:
    f.write(str(traininglablesBeforeNorm [i])+'\t'+str(prediction))
    f.write('\n')
    i+=1
    f.flush()
f.close()
```