Applying Artificial Neural Networks with Bourgoyne and Young Model to Predict Rate of Penetration in Al-Garraf Oil Field

Ebrahim A. AL-Assad [iD] [✉]*, Sameera M. Hamd-Allah [iD] [✉]

Department of Petroleum Engineering, College of Engineering, University of Baghdad, Baghdad, Iraq

ABSTRACT

Estimating the rate of penetration (ROP) for oil well drilling is essential for cost-effective and safe drilling operations; drilling companies have been aiming for ROP estimation since the industry's first decade. To achieve this goal, among numerous models, the Burgoyne and Young model (BYM), an equation-based approach, was developed and widely used to predict the ROP based on multiple linear regression (MLR). Artificial Neural Network (ANN) is a machine learning technique that analyzes drilling data and makes ROP predictions. Many studies have been conducted worldwide to employ BYM, and others have aimed to improve it for different circumstances. ANN has also shown effectiveness in these fields. This study proposes an approach that combines the benefits of Feedforward Neural Networks (FNN) from the ANN model and BYM equations to enhance ROP prediction. Integrating BYM with FNN leverages the equation-based model and harnesses the power and efficiency of machine learning. These results significantly improved accuracy and efficiency in predicting ROP in oil wells. ROP modeling input parameters include total measured and true vertical depth, Weight on Bit, Rotation Per Minute, standpipe pressure, pump flow, equivalent mud weight, bit size, nozzle size, formation pressure, and bit jet impact force, which are recalculated by BYM equations and fed to both MLR and FNN. When tested on real-time data from Al-Garraf oil field, the outcomes demonstrate higher R2, lower residuals, and zero P-value compared to MLR, which validate the approach accuracy and provide precise ROP prediction in future drilling plans.

Keywords: Artificial neural networks, Bourgoyne and young model, Multiple linear regression, Rate of penetration, Al-Garraf oil field.

1. INTRODUCTION

As oil drilling techniques developed and the reservoir depth target increased, the drilling operation became more expensive, with more drilling risk phases for the companies. ROP estimation has become essential in handling costs and controlling risks; it is a good tool that

146

gives an overall vision of time and actual circumstances that helps designers and workers make accurate decisions.

Numerous studies on predicting ROP were conducted earlier in the 20th century. Researchers depend on charts derived from particular equations to explain the relationship between ROP and different drilling parameters, as demonstrated by the work of **(Maurer, 1962; Bingham, 1964; Warren, 1987)**; other mathematical models proposed to optimize WOB, RPM, and bit hydraulics include the works of **(Speer, 1959; Graham and Muench, 1959; Eckel, 1967)**. The Bourgoyne and Young model, conducted in 1974, introduced a novel mathematical model used to optimize eight controllable and uncontrollable drilling parameters in a single equation for optimal ROP achieved through multiple linear regression (MLR) as demonstrated in papers such as **(Bourgoyne and Young,1974; Bourgoyne et al.,1986)**, BYM is considered among other models as the confident drilling ROP optimization model, and numerous studies focused on employing it to predict ROP on different oil fields in the world, such as Iraqi fields, as demonstrated in papers **(Hamad-Allah and Ismael, 2008; Majid and Ayad, 2019; Darwish et al., 2020)**. Similar in Iranian fields, as shown in paper studies such as **(Irawan et al., 2012)** and also in kenia fields as in **(Miyora, 2014).** The BYM has also been applied to optimize ROP for polycrystalline diamond (PDC) bits, as demonstrated by **(Eren and Ozbayoglu, 2010; Darwish et al., 2020)** and Bits Evaluating and Selection **(Ayad et al., 2015; Amel, 2017).**

Various other new techniques and computerized methods have been employed to predict ROP. These include mathematical optimization techniques **(Bahari et al., 2008; Moradi et al., 2010)**, mechanical-specific energy **(Alsubaih et al., 2018)**, support vector machines (SVM) **(Abdulmalek et al., 2018)**, artificial neural networks (ANN) **(Amar and Ibrahim, 2012; Elkatatny et al., 2017)**, pattern remote system methods **(Liu et al., 2018)**, neuro-fuzzy inference system **(Yasser et al., 2020)**, and automated systems **(Zha et al., 2018)**. Many recent studies were focused on Artificial Neural Networks (ANN) **(AL-Zirej and Hassan, 2019; Huihui et al.,2022; Teeba et al., 2022)**, Long Short-Term Memory (LSTM) from ANN **(Huihui et al.,2022; Hongtao et al., 2022)**, machine learning methods **(Asad et al.,2022; Fei et al.,2023; Chengxi et al.,2023; Chris, 2023)**, and Deep learning methods **(Tong et al., 2023)** to predict the rate of penetration (ROP) in oil wells. Different approaches have been explored, including integrating ANN with optimization algorithms such as Particle Swarm Optimization (PSO) and comparing different ANN models **(Ololade et al., 2021)**.

The mentioned research did not extensively investigate the potential of using ANN integrated with BYM equations and depended only on the primary data as a dataset; BYM equations were made based on confident hypotheses taken from comprehensive studies that established a clear relationship between drilling parameters and the ROP **(Bourgoyne and Young, 1974)**. This study proposes integrating FNN with BYM equations to predict ROP in Al-Garraf oil wells in southern Iraq. It utilizes the benefits of merging FNN with the BYM equations to predict ROP and estimate drilling performance using Al-Garraf field real-time drilling data from three wells. The main principle involves calculating each point from the thousands of data points according to the BYM equations and creating a new dataset as a training, validating, and testing dataset to feed into FNN rather than relying solely on Basic drilling variables as a dataset. The limitation arises from the use of many parameters and complex and lengthy calculations; it involves writing and utilizing hundreds of MATLAB programming codes. The findings showed a significant improvement in the rate of penetration prediction regarding the traditional methods.

2. RESERVOIR AND DATA DESCRIPTION

Al-Garraf oil field in southern Iraq has three main reservoirs, one of which is the al-Mishrif formation. Al-Mishrif is a carbonate rock reservoir from the Cretaceous period that contributes significant amounts of oil to the field's overall production. Development of the al-Garraf field, including the al-Mishrif reservoir, began to increase since the 2010s. Production from the al-Mishrif and other reservoirs has steadily increased, with the field now producing around 100,000 barrels per day. However, extracting oil from the complex al-Mishrif carbonate reservoir presents technical challenges that the field's operators continue to address. **Fig. 1** shows Al-Garraf's location, and **Fig. 2** shows its Lithological prognosis.

The data were collected and analyzed from the drilling mud loggers of three offset wells within a 0.25m depth interval, with over 10,000 data points for each well. Parameters included measured depth, vertical depth, weight on bit, pipe rotation, standpipe pressure, pump flow, equivalent mud weight, bit size, bit nozzle size, formation pressure, and bit jet impact force. Bit tooth wear is assumed to be zero because there is no record **(Bahari et al., 2008)**. **Table 1** shows the statistics of the drilling variables after data normalization and cleaning for the three wells
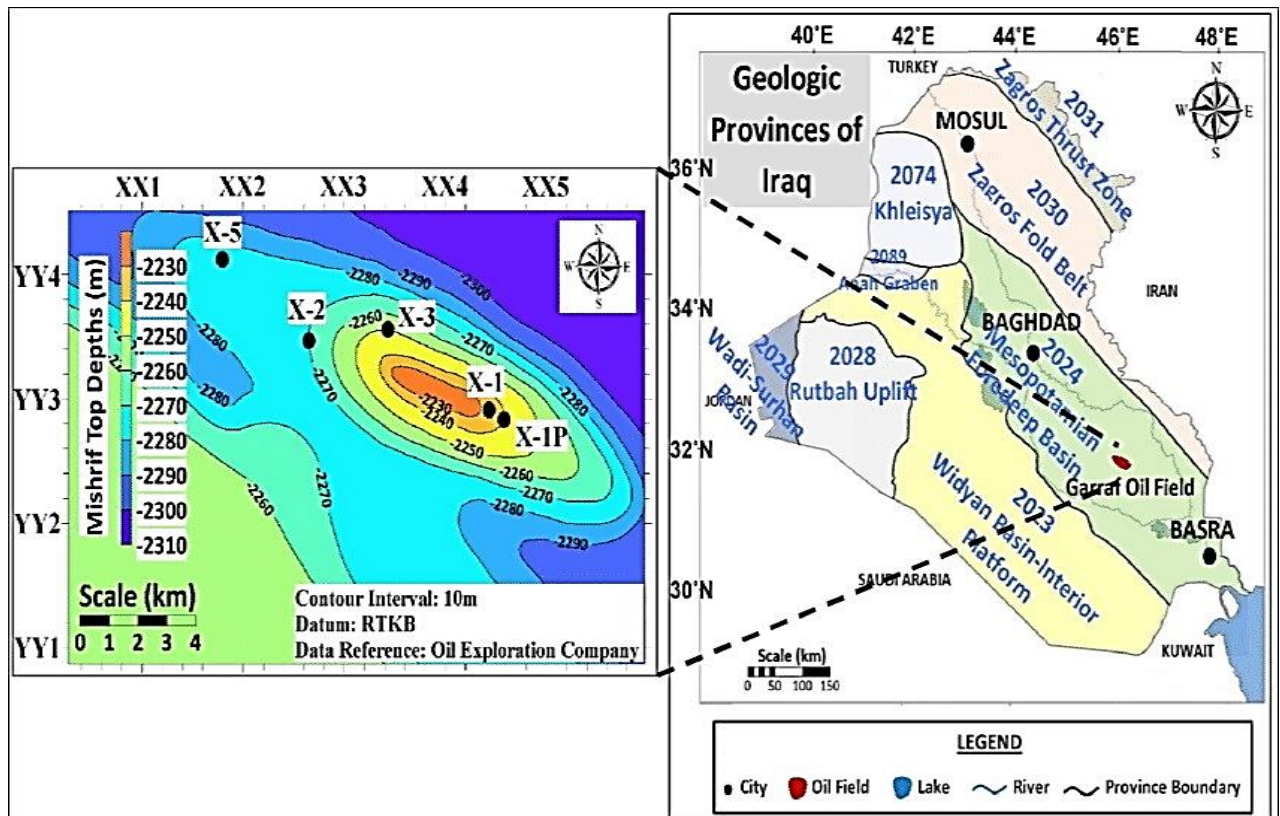

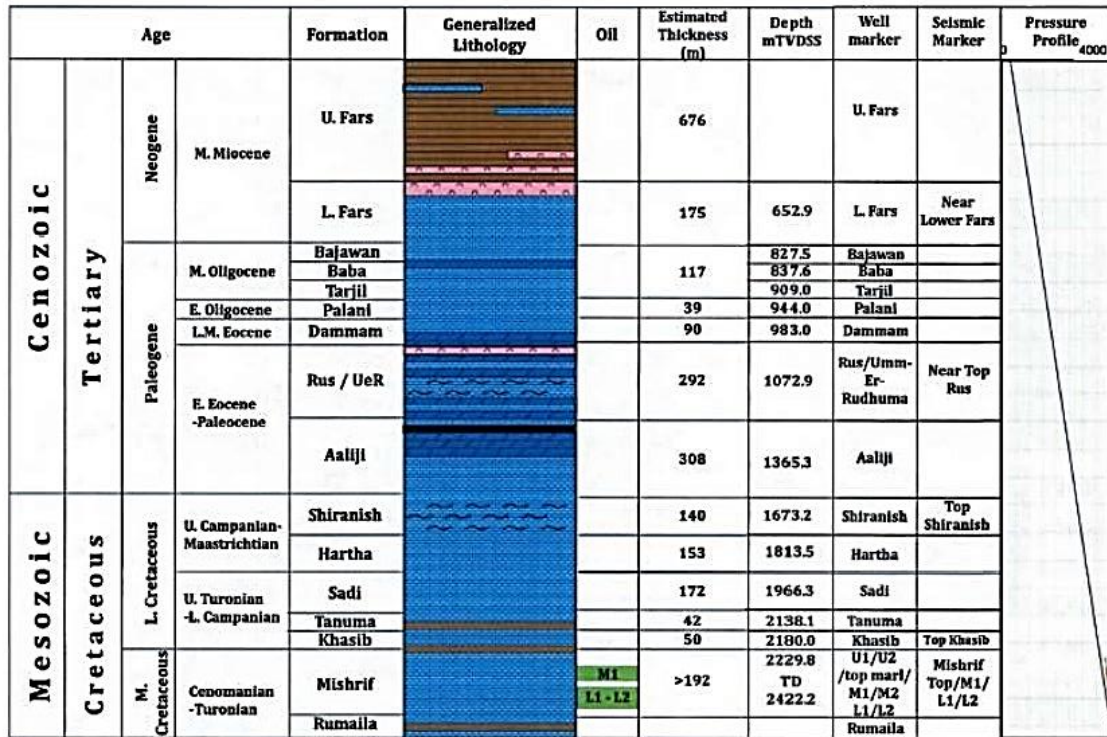**Figure 1**. Al-Garraf oil field location. (**Kareem et al., 2022**)

Figure 2. Lithological prognosis for Al-Garraf oil field **(Petronas operation, 2016)**

Table 1. Al-Garraf three wells drilling variables statistic.

| ………………………| Well Name	GA-38P				GA-52P				GA-88P			
Drilling Variables | statistic	Max	Min	Mean	Std	Max	Min	Mean	Std	Max	Min	Mean	Std
Measured Depth m	2641.75	33.50	1372.65	742.71	2946.00	142.00	1662.03	768.74	3025.00	335.50	1811.48	729.79
True vertical Depth m	2442.53	33.50	1307.67	679.48	2442.00	141.82	1441.12	609.67	2440.52	335.48	1541.21	546.58
Rate of penetration m\hour	19.68	1.01	7.38	4.14	22.32	1.02	8.45	4.60	37.72	1.19	13.79	7.34
Weight On Bit klbm	23.70	1.10	9.49	4.48	24.20	1.10	9.89	4.75	25.80	1.10	11.79	4.52
Rotation Per Minute 1\minute	175.00	46.00	132.49	33.85	163.00	87.00	129.58	16.73	217.00	129.00	177.51	17.62
Stand Pipe Pressure psi	3541.61	258.78	2100.09	844.32	3830.87	964.32	2582.07	649.33	3703.76	1228.30	2814.31	568.86
Pump Flow gpm	1035.56	693.17	909.11	64.51	1001.93	680.73	839.07	52.28	979.62	652.22	848.78	72.35
ECD, density PPG	10.96	8.40	9.68	0.64	10.77	8.50	9.85	0.68	10.92	8.06	10.00	0.76
Bit Nozzle Area in	1.35	1.18	1.32	0.06	1.35	1.19	1.27	0.08	1.35	1.33	1.35	0.01
Bit Pressure loss psi	665.44	264.00	427.65	65.67	564.56	242.52	395.12	65.11	456.36	235.41	364.58	37.54
Bit Jet Impact Force	1414.57	622.34	1010.95	120.28	1215.71	592.07	900.84	103.85	1114.14	566.91	887.18	93.96
Bit Size in	26.00	12.25	17.57	5.15	26.00	12.25	15.07	3.06	17.50	12.25	14.59	2.61
Formation Pressure PPG	10.10	8.20	8.66	0.60	10.20	7.90	8.83	0.81	10.30	7.50	9.32	0.70

3. METHODOLOGY

The methodology of using FNN from ANN integrated with BYM in MATLAB to predict ROP involves multiple steps. First, a drilling dataset with all drilling parameters involved in the eight BYM equations is collected. For data accuracy, the data are collected from real-time drilling sensors on mud loggers instead of depending on daily or final drilling reports. The data are preprocessed with normalization and cleaning and then split into three groups: training (80%), validating (10%), and testing (10%) datasets. The number of hidden layers and neurons per layer is determined, and the FNN model is created using a suitable MATLAB function. The FNN model is trained using an appropriate function on the training dataset. The validation datasets are evaluated, and the model is tested on the testing dataset to assess

its generalization ability. Different visual plots are generated to compare and analyze the predicted versus measured ROP values. The model is fine-tuned if needed by adjusting and training it with additional data. Once satisfied, the model predicts ROP values for new real-time future drilling operations. For successful ROP prediction using FNN, factors like dataset quality, appropriate network architecture, training parameters, and domain knowledge in drilling operations are essential to consider.

Three figures, including measured-predicted ROP, residuals, and ROP-Depth figures, are plotted for each case to demonstrate and compare findings. Four statistical variables, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R2), and adjusted-R^2, are shown in the tables to measure the goodness of fit between predicted and measured ROP values. The P-value is zero for all cases. **Fig.** 3 shows the research flow chart's main lines.
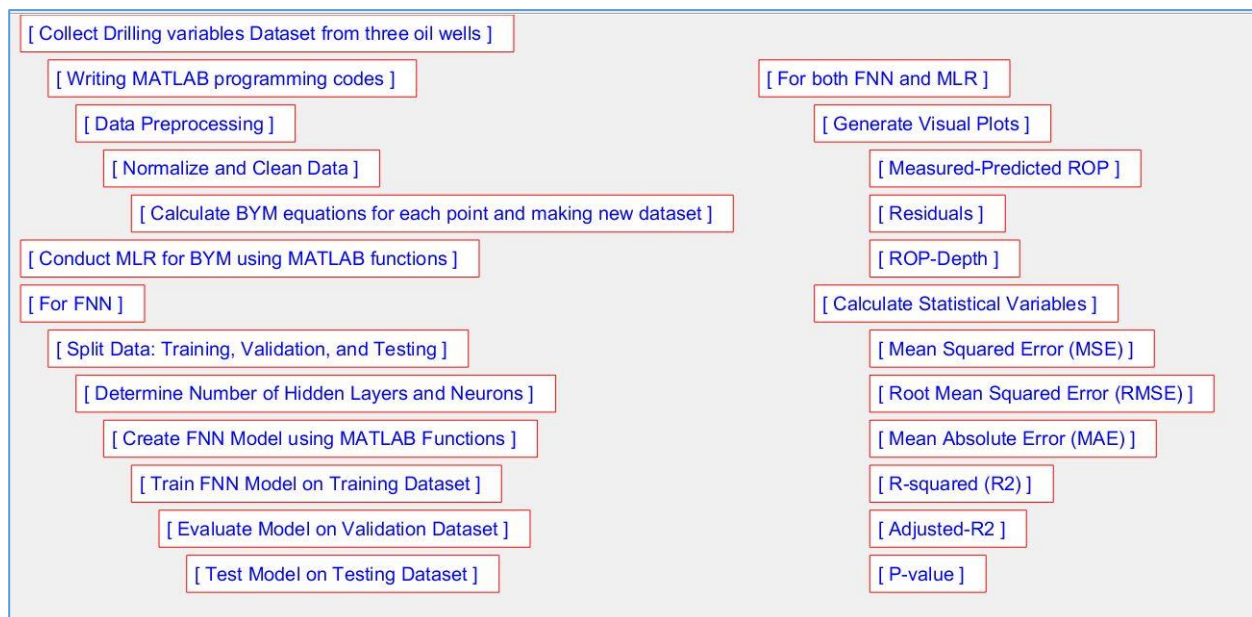
Figure 3. The research flow chart's main lines.

3.1 Feedforward Neural Networks and Burgoyne and Young Model

Feedforward Neural Networks (FNN) is a significant artificial neural network (ANN) system delineated by its one-way data flow from input through hidden layers to output, free of loops or cycles. FNN is extensively utilized for regression tasks, where neurons compute weighted sums of inputs, apply activation functions, and transmit results to subsequent layers, with network parameters learned through training to model intricate input-output relationships. FNN is a machine-learning model that takes stimulation from the networks of neurons in the human brain. Think of FNN as connected nodes, like brain cells, organized in input, hidden, and output layers. FNN is trained using labeled data, where the weights and biases of the nodes are modified during the learning process. which allows FNN to become practiced at predicting complex patterns and relationships, whether for tasks like computing numerical values (regression) or classifying data **(Muhammad et al., 2023).** FNN is used in drilling operations to predict the ROP. It creates a model that captures the connection between various drilling parameters, such as weight on bit, rotary speed, mud flow rate, and so on, in addition to the ROP values. Training an FNN with historical drilling data can help them learn

from the patterns in the data and make accurate ROP predictions based on the given drilling parameters **(Ololade et al., 2021)**. The **(Bourgoyne and Young, 1974)** is a linear relationship between the ROP and eight effective parameters affecting it. This model is based on MLR, the statistical synthesis of the collected data from the offset drilled wells. The MLR equation is:

$$Y = \beta 0 + \beta 1 x 1 + \beta 2 x 2 + \ldots + \beta k x k + \epsilon, \tag{1}$$

Where βj represents the regression coefficient, and $x j$ represents BYM eight terms, the MLR's objective is to identify a vector of least squares estimators that minimizes the error concerning the regression coefficients.

BYM considers the effects of the drilling parameters, such as formation depth, formation compaction, the pressure differential across the bottom of the hole, the bit diameter and weight on the bit, the drill pipe rotary speed, the bit wear, and the bit hydraulics.

$$ROP = exp\left(a1 + \sum_{2}^{8} a_j x_j\right) \tag{2}$$

Constants *a1* to *a8* are computed in MLR for local drilling conditions, utilizing offset well data. The functions are built around Bourgoyne's assumptions regarding variables that impact the drilling rate. These assumptions include x1, which represents un-modeled variables like formation strength; x2, which models an increase in rock strength with depth and assumes an exponential decrease in ROP; x3, which models under-compaction in abnormal pressure formation and assumes an exponential rise in ROP with a pore pressure gradient; x4, which models the differential pressure between hydrostatic and formation pressure on ROP and assumes an exponential decrease in ROP with excess bottom-hole pressure; x5, which models bit weight on ROP and assumes direct relation between ROP and WOB; x6, which models rotary speed on ROP and assumes direct relation between ROP and RPM; x7, which models bit tooth wear on ROP and assumes an exponential decrease in ROP with increased tooth wear; and x8, which models bit hydraulics on ROP and assumes direct relation between ROP and a Reynolds number or bit jet impact force. These assumptions and followed equations are grounded on comprehensive prior studies mentioned in the BYM paper **(Bourgoyne and Young, 1974)**.

$$x1 = 1 \tag{3}$$

$$x2 = 10000 - D \tag{4}$$

$$x3 = D^{0.69}(gp - 9) \tag{5}$$

$$x4 = D\ (gp\ -\ \rho c\ ) \tag{6}$$

$$x5 = ln\left[\frac{\frac{W}{d_b} - [\frac{W}{d_b}]_t}{4 - [\frac{W}{d_b}]_t}\right] \tag{7}$$

$$x6 = ln\left(\frac{N}{60}\right) \tag{8}$$

$$x7 = -h \tag{9}$$

$$x8 = ln\left(\frac{F_j}{1000}\right) \tag{10}$$

Where, D: Oil well, true vertical depth (ft); Fj: Bit Jet impact force (in pounds of force); gp: Pore pressure (ppg); h: Fractional tooth wear; N: Rotary speed (revolutions per minute, RPM); ROP: BYM dependent variable, rate of penetration; Wb/db: (Weight on bit) per (bit diameter) (pounds per inch); $x1$ to $x8$: BYM independent variables; $[\frac{W}{d_b}]_t$: The drilling begins at Threshold WOB 1000lbf/inch; ρc: Equivalent circulation density (ppg).

3.2 Plots and Statistical Data

Related to the methodology and the outcome of the cases, three figures are plotted for each case:

1- Scatter plot: This plot compares the measured ROP on x-axes versus the predicted ROP observed from MLR for BYM, besides the FNN model, with different clear signs and colors to all data points and the least square line.

2- Residual plot: This plot helps to evaluate the performance of the FNN model and MLR by showing the differences between measured and predicted ROP on the y-axis against the FNN and MLR predicted ROP on the x-axis. Residuals are distributed around zero in negative and positive values.

3- Line plot: This plot shows, by different colored lines, the comparison between measured and MLR predicted ROP, in addition to FNN model predicted ROP, to analyze and evaluate their ROP trends and patterns compared to measured ROP data.

Evaluation of the FNN model and MLR for BYM performance and accuracy using various statistical measurements and presenting related results in tables:

1- Mean Squared Error (MSE) measures the average of squared differences between measured and predicted ROP values by the FNN.

$$MSE \ = \ mean((Predicted \ ROP \ - \ Measured \ ROP)^2) \tag{11}$$

2- Root Mean Squared Error (RMSE) is the square root of MSE.

$$RMSE \ = \ sqrt\big(mean((Predicted \ ROP \ - \ Measured \ ROP)^2)\big) \tag{12}$$

3- Mean Absolute Error (MAE) calculates the average absolute differences between predicted and measured ROP values.

$$MAE \ = \ mean\big(abs(Predicted \ ROP \ - \ Measured \ ROP)\big) \tag{13}$$

4- The R-squared (R^2) Coefficient of Determination measures the goodness of fit between predicted and measured ROP values.

$$R^2 \ = \ 1 \ - \ sum((Measured \ ROP \ – \ Predicted \ ROP)^2) \ / \ sum((Measured \ ROP \ - \ mean(Measured \ ROP))^2) \tag{14}$$

5- The adjusted-R^2 adjusts the R-squared value by considering the number of predictors in the model and the sample size. It penalizes adding additional predictors that do not significantly improve the model's fit, helping to prevent overfitting.

$$adjusted \ R^2 \ = \ 1 \ - \ (1 \ - \ R^2) \ * \ (n \ - \ 1) \ / \ (n \ - \ p \ - \ 1) \tag{15}$$

R^2 is the ordinary R-squared value, n is the number of samples or observations, and p is the number of predictors or independent variables in the model.

6- The p-value is a crucial statistical measure that calculates the probability of achieving outcomes as extreme as the recorded data, assuming the null hypothesis is correct. It is widely utilized to ascertain the significance of a statistical test or the efficacy of evidence against the null hypothesis. A smaller p-value suggests more substantial evidence against the

null hypothesis, implying that the observed results are improbable to occur randomly; a p-value less than 0.05 indicates statistical significance at the 5% level **(Padhma, 2023)**.

4. RESULTS AND DISCUSSION

Four cases are thoroughly examined, each belonging to one oil well in the Al-Garraf oil field: case 1 for Ga-J38P, case 2 for Ga-J52P, case 3 for Ga-J88P, and case 4 combines the three wells. All four cases are utilized on an extended depth interval to confirm the results between each other, starting from 250m and 750m and ending with 1500m. Because the depth interval for data points is 0.25m for each point, recording the data points will result in about 6000 dataset points for 1500m. This is an extensive dataset, and the FNN and MLR models successfully manage it and provide accurate results. MATLAB programming has proven a powerful tool for researchers; Appendix A shows a small part of the programming codes.

By analyzing the findings of case 1, related to the well Ga-J38P, the statistical parameters in **Table 2** shows the improvement of the FNN model that incorporates the BYM equations compared with the MLR. The R^2 increases from 0.478 to 0.804, indicating a significant enhancement in the FNN model compared to the MLR. For MSE, RMSE, and MAE, which are different types of residual statistical measurements, there is a substantial decrease in their values in the FNN model compared to the MLR model.

These differences in statistical parameters indicate a significant improvement in the new approach that combines FNN and BYM equations over the MLR approach for BYM. Analyzing **Fig. 4** shows that the FNN red least square line is aligned 45 degrees more than the MLR least squares line, 45-degree mean ($R^2=1$). Due to the 0.25m dataset increment, **Fig. 4** shows many data points for the two models, even in the 250m depth interval.

Fig. 5 represents the residual plot for the FNN and MLR models, showing that the ANN signs are less spread on the Y-axis than the MLR signs. The Y-axis represents the residual between the Measured and predicted ROP, either in a positive or negative value, indicating that the FNN has less error than the MLR. This fact is further confirmed by other visual evidence in the same plot: the FNN signs are more distributed on the X-axis, representing the predicted ROP, than the MLR signs. This means the FNN-ROP prediction is more aligned with the Measured ROP; However, some data points were inaccurate, this did not affect the overall result due to the large number of points showing excellent accuracy.

Fig. 6 represents the ROP versus oil well depth and compares the FNN and MLR with the Measured ROP. The FNN curve is more aligned with the Measured ROP curve than the MLR curve, confirming the findings of the other figures. This indicates the improvement of the FNN-ROP prediction compared to the MLR prediction. MLR always tends to go with the optimal values collected among all the dataset points and represents the overall effect on one MLR equation.

Table 2. Case 1 details and statistics.

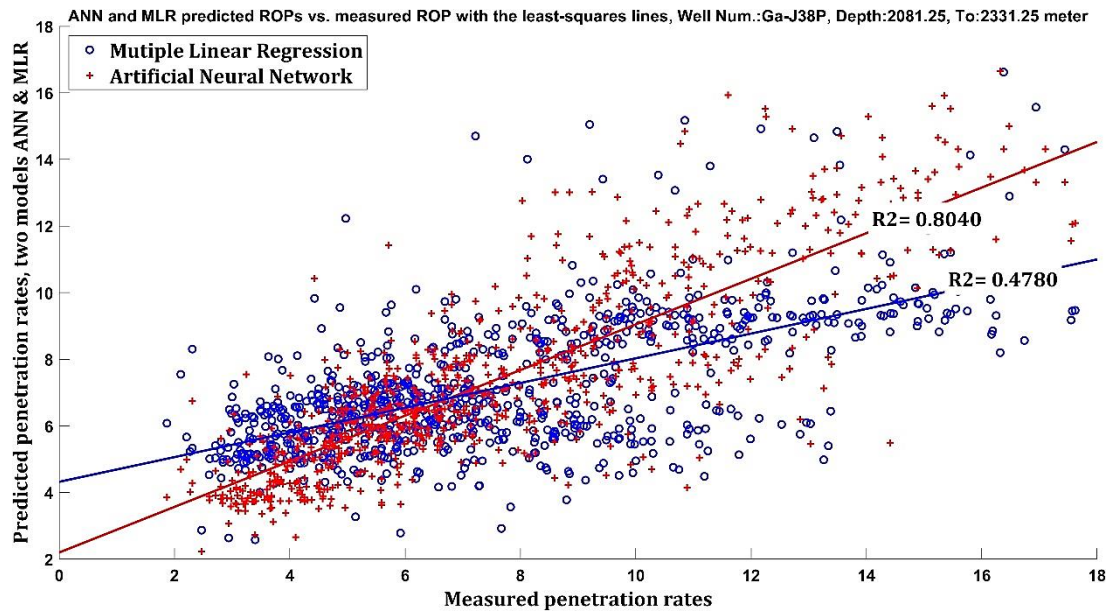| Case	Well number	From depth m	To depth m	Data points	P-value
1	Ga-J38P	2081.25	2331.25	1000	0
Models || Statistic	MSE	RMSE	MAE	R^2	Adjusted-R^2
MLR	7.049	2.655	2.0706	0.4780	0.44385
FNN	3.354	1.8313	1.3387	0.8040	0.78227

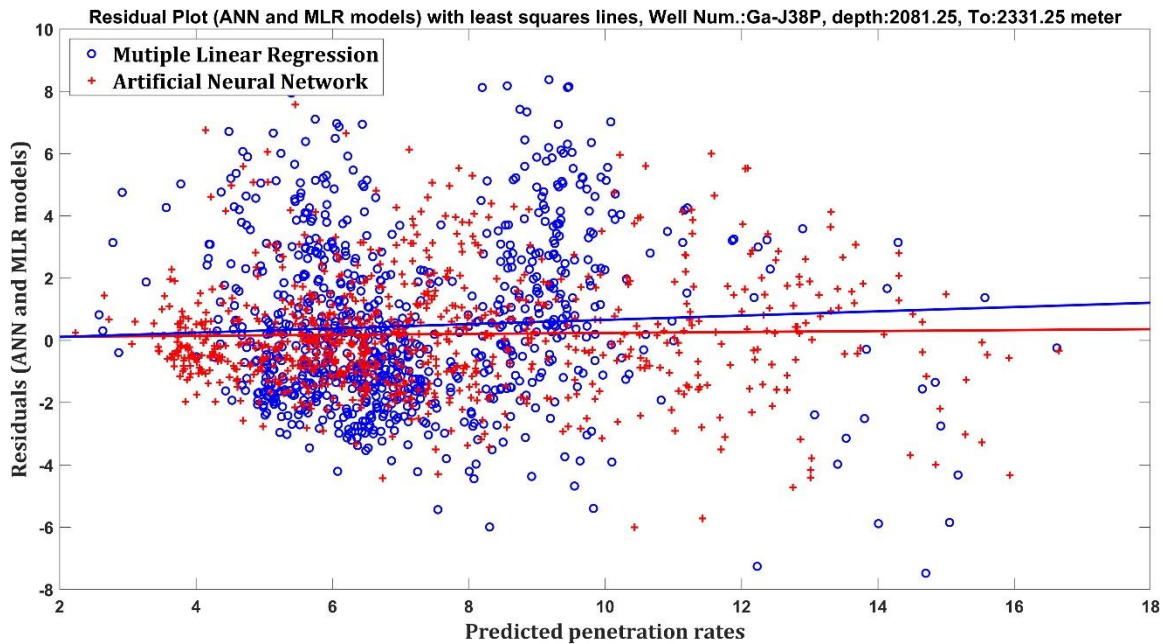**Figure 4**: Measured ROP vs. predicted ROP of the FNN & MLR, case 1.

Figure 5. Residuals vs. predicted ROP of the FNN & MLR, case 1.

Analyzing case 2 for oil well Ga-J52P with the same depth interval as case 1, which is 250m but for a different starting depth, with data points of about 1000, indicates the improvement of the FNN approach in various ways in **Table 3**. For example, the R^2 increases from 0.475 for MLR to 0.771 for FNN. The residuals decrease regarding MSE, RMSE, and MAE statistic variables. **Fig. 7** shows that the FNN least-square line is more aligned to 45 degrees than the MLR least-square line. **Fig. 8** indicates that the FNN signs are less distributed on the Y-axis and more distributed on the X-axis than the MLR signs, indicating the improvement of the FNN over the MLR.
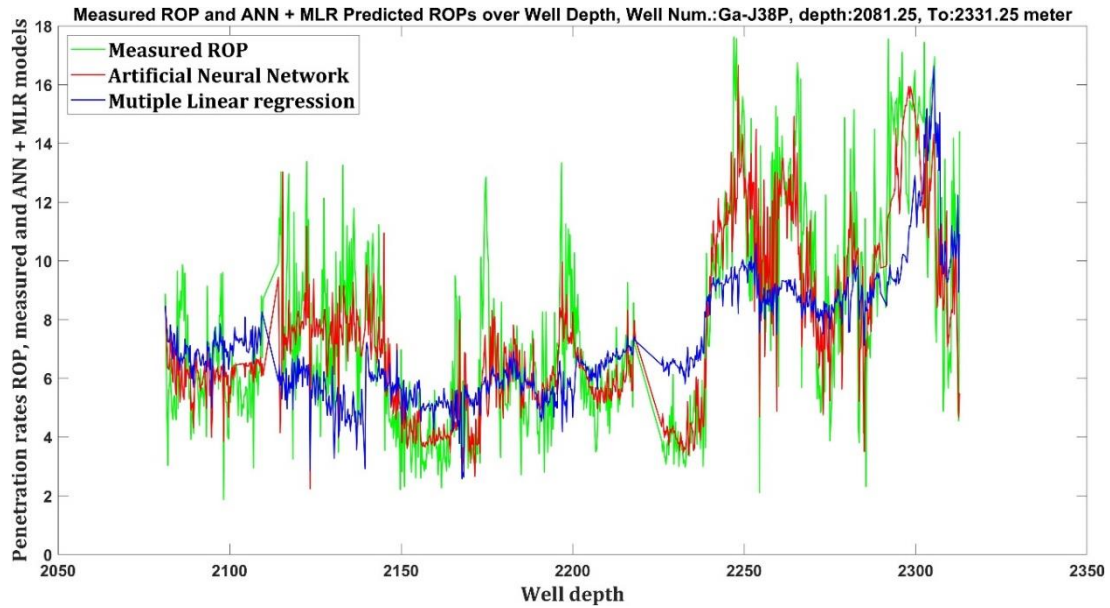
154

Figure 6. Well Depth vs. Measured & Predicted ROP of FNN & MLR.

Fig. 9 shows that the FNN-ROP predicted curve is more aligned with the Measured ROP curve than the MLR-ROP predicted curve, and the FNN curve tries to reach the Measured curve more closely than the MLR curve.

Table 3. Case 2 details and statistic

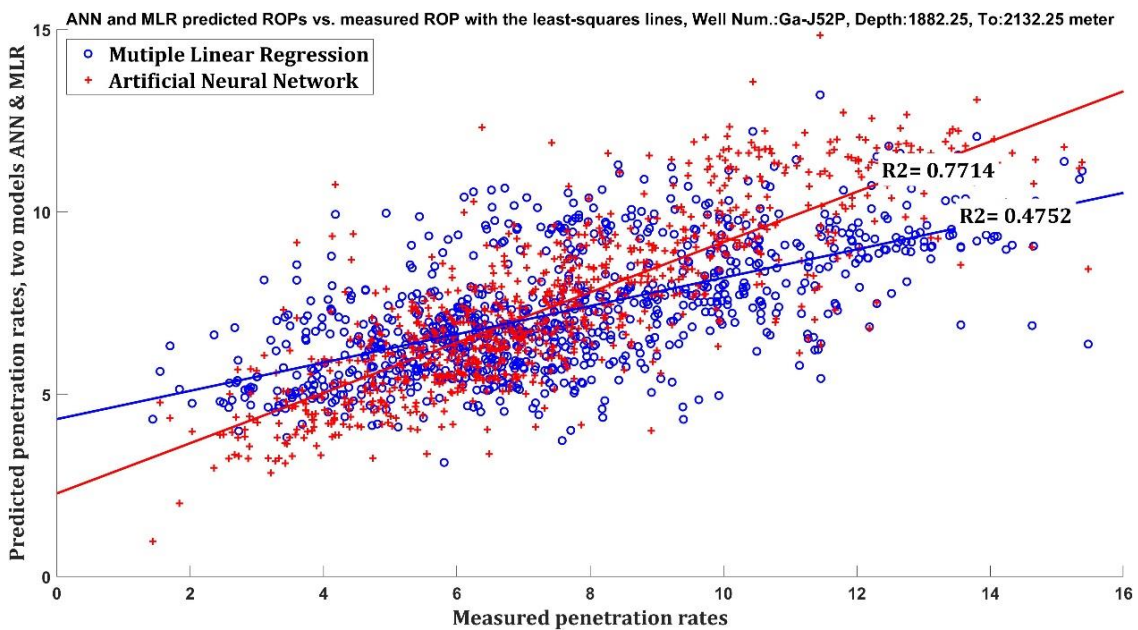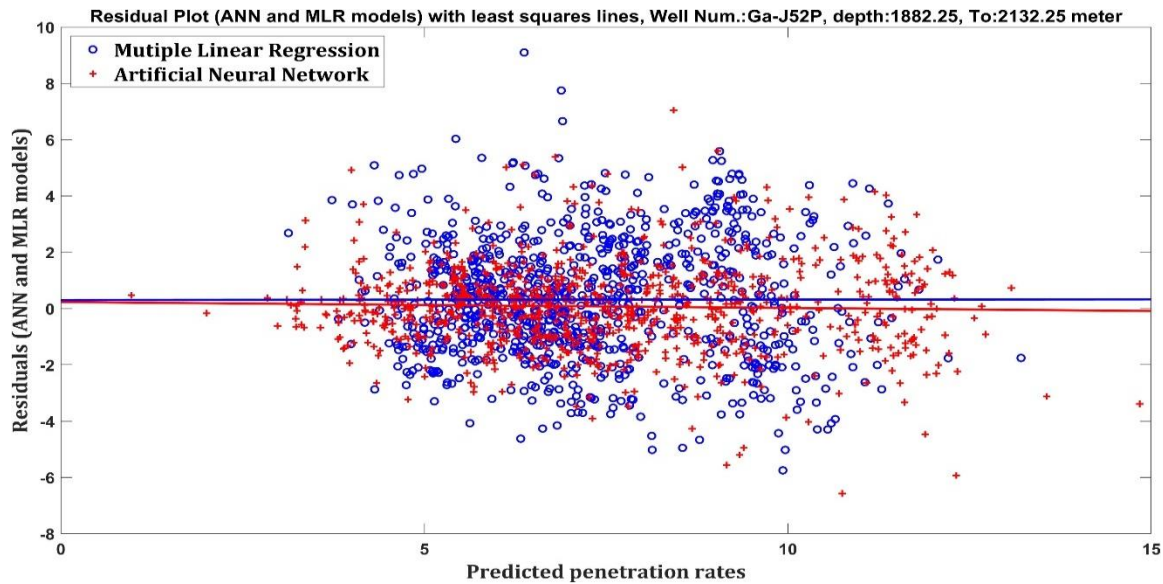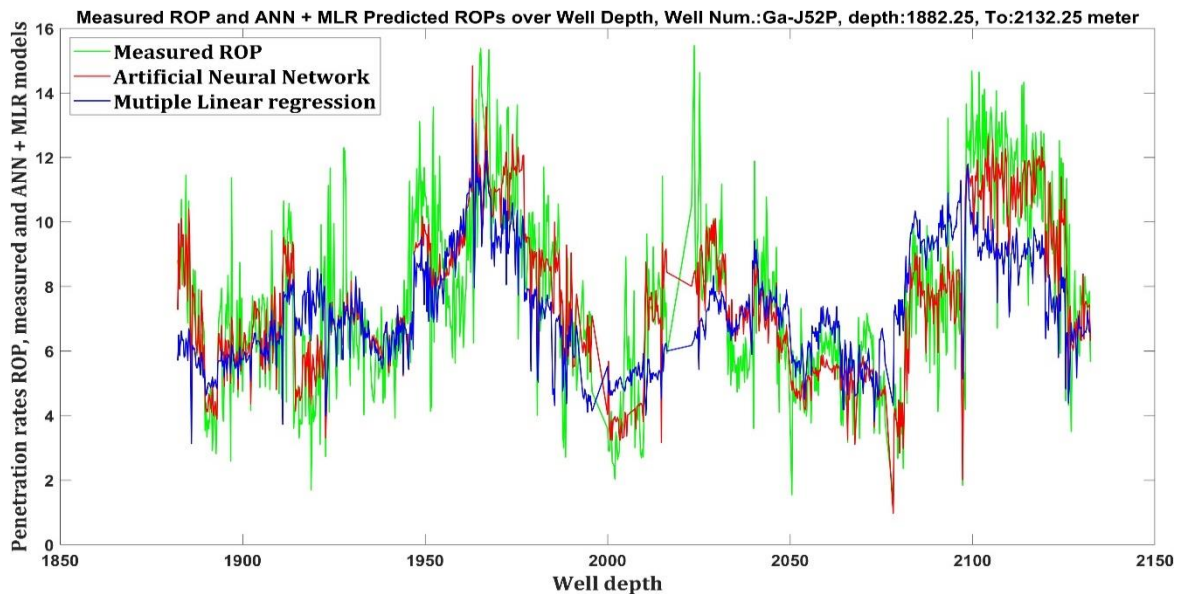| Case	Well number	From depth m	To depth m	Data points	P-value
2	Ga-J52P	1882.25	2132.25	1000	0
Models || Statistic	MSE	RMSE	MAE	R²	Adjusted-R²
MLR	4.6356	2.153	1.7248	0.4752	0.46342
FNN	2.43771	1.5613	1.1663	0.7714	0.74216

Figure 7. Measured ROP vs Predicted ROP of the FNN + MLR, case 2.

155

Figure 8: Residuals vs Predicted ROP of the FNN + MLR, case 2.

Figure 9. Well Depth vs Measured + Predicted ROP of FNN + MLR, case 2.

Analyzing case 3 for oil well Ga-J88P, which has a more extended interval depth of 1500m and approximately 6000 data points, **Table 4** shows the improvement of the FNN approach in various aspects. This includes an increase in R^2 from 0.4494 for MLR to 0.6615 for FNN. The residuals also decrease in MSE, RMSE, and MAE statistic variables in FNN compared to MLR, indicating improved performance. **Fig. 10** depicts the superior performance of the FNN least-square line compared to the MLR least-square line, and it is clear that the vast dataset does not affect the findings. **Fig. 11** shows that the FNN has less residual than MLR and is more aligned with the measured ROP. Finally, **Fig. 12** illustrates that the FNN-ROP predicted curve is more closely aligned with the measured ROP than the MLR-ROP predicted curve. The FNN curve displays a more concerted effort to reach the measured curve than the MLR curve despite the crowded dataset in **Fig. 12**. The MLR curve considers the optimal values collected from all the data points and represents them in one MLR equation, while the

behavior of the FNN curve is evident as it tries to reach the measured ROP curve by a good percentage. The decrease in R^2 in FNN and MLR compared to other cases is due to the vast number of 6000 data points. Both FNN and MLR attempt to handle the effect of the entire dataset and demonstrate optimal behavior for ROP within an overall point acceptable limit.

Table 4. Case 3 details and statistic

Case	Well number	From depth m	To depth m	Data points	P-value
3	Ga-J88P	233	1733	6000	0
Models || Statistic	**MSE**	**RMSE**	**MAE**	**R²**	**Adjusted-R²**
MLR	72.534	8.5167	6.3125	0.4494	0.46192
FNN	48.887	6.9919	4.9884	0.6615	0.69432
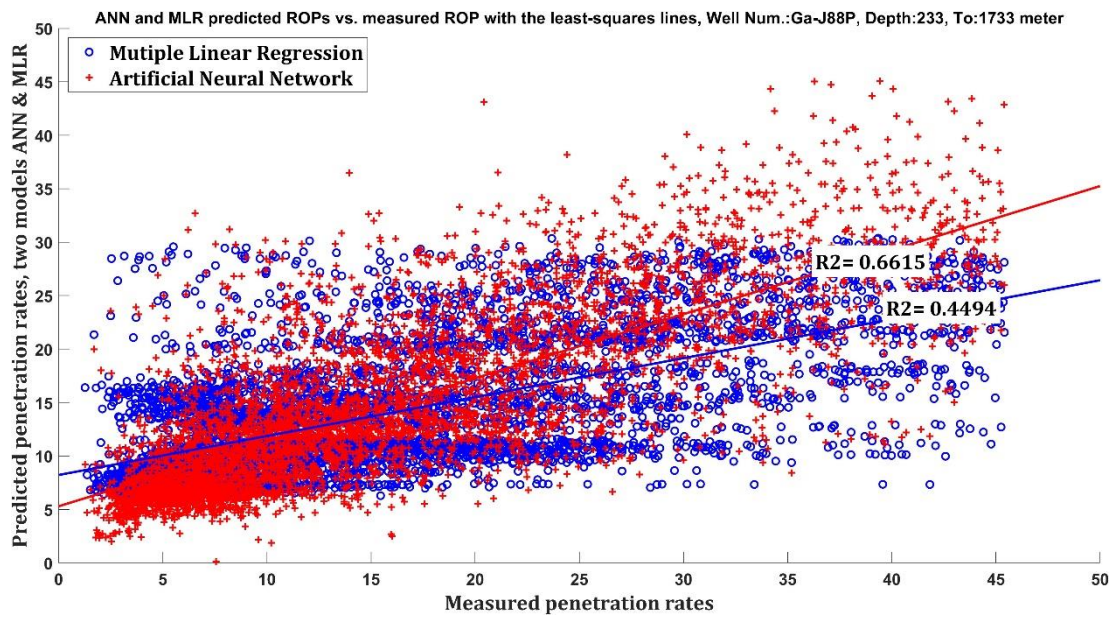
Figure 10. Measured ROP vs Predicted ROP of the ANN + MLR, case 3.
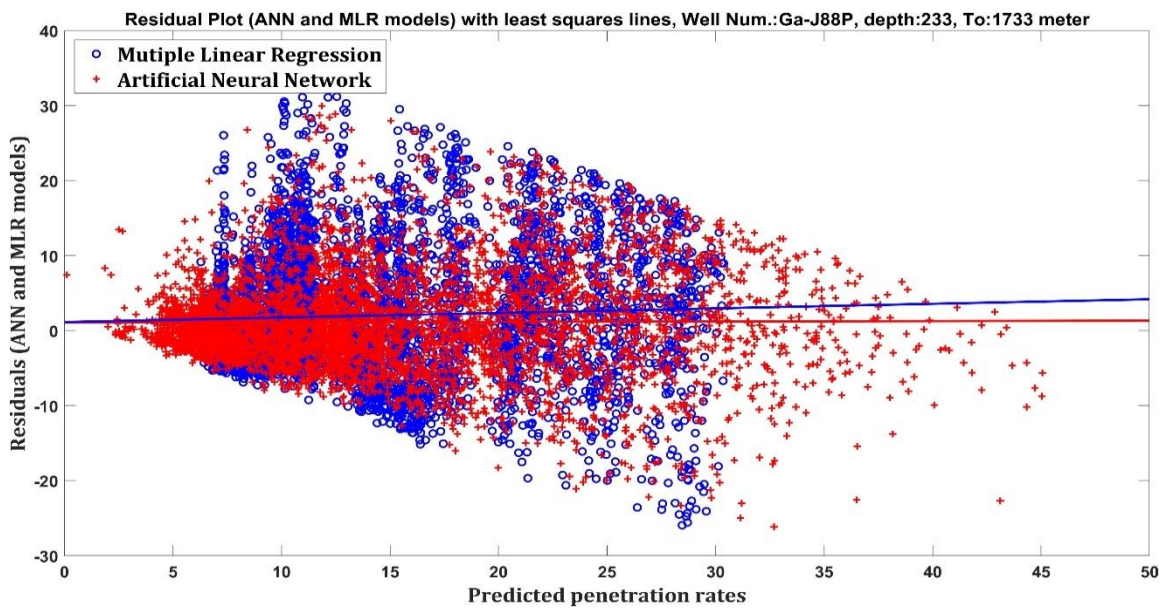
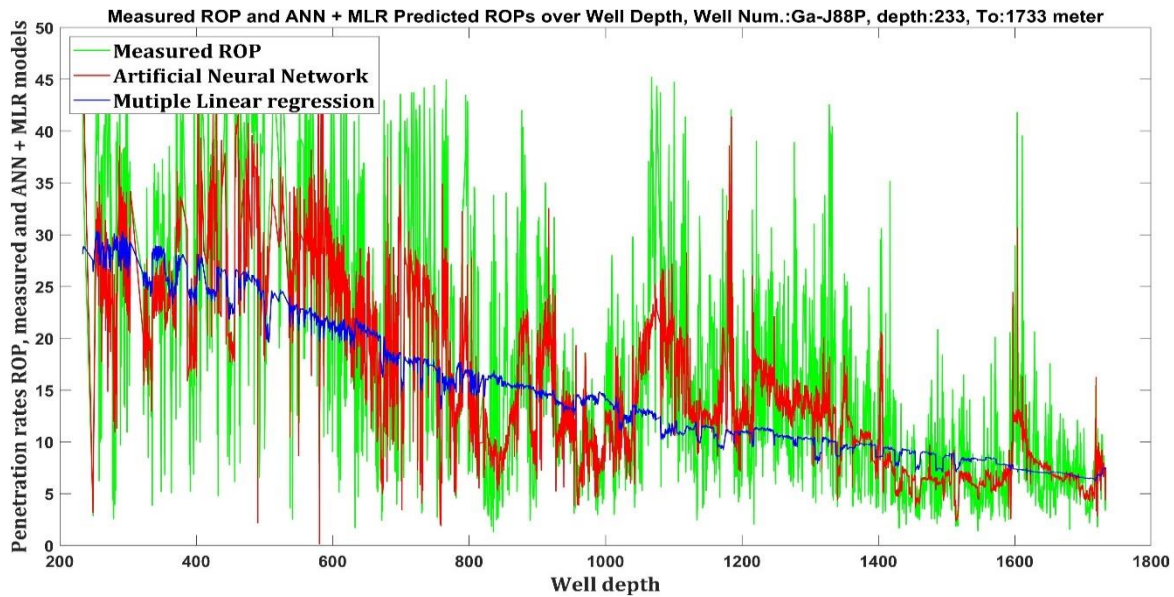**Figure 11.** Residuals vs Predicted ROP of the ANN + MLR, case 3.

Figure 12. Well Depth vs Measured + Predicted ROP of ANN + MLR, case 3.

Case 4, which represents the combination of three oil wells, has an extended interval depth of 750m and approximately 3000 data points extracted from the three wells. The first data point is from Ga-J38P, the second from Ga-J52P, the third from Ga-J88P, and so on for the rest of the dataset points, alternating between the three wells.

The statistical findings in **Table 5** and **Figs. 13** to **15** indicate that FNN outperforms MLR.

Table 5. Case 4 details and statistic

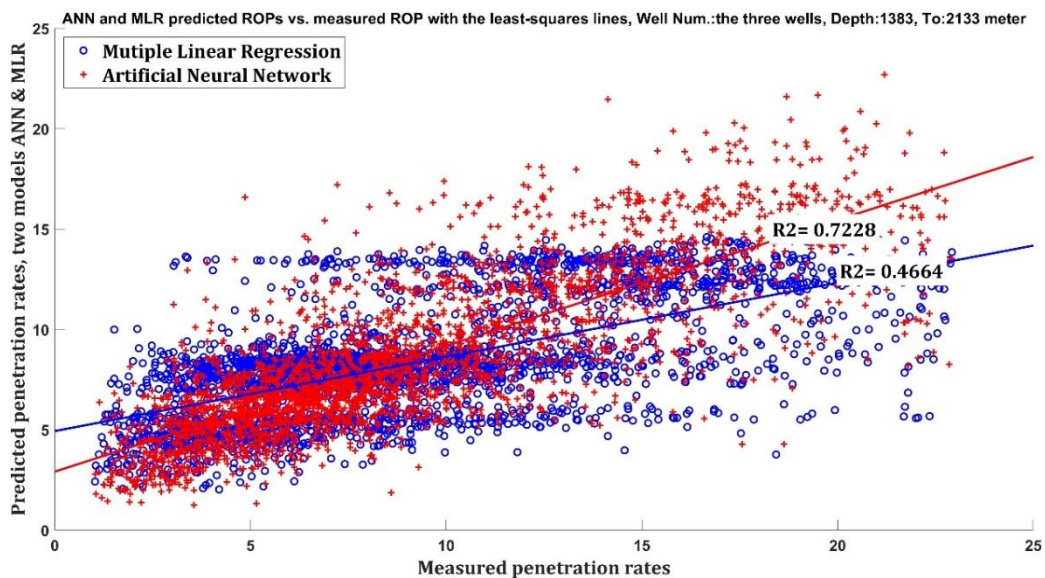| Case	Well number	From depth m	To depth m	Data points	P-value
4	Ga-J(38+52+88)P	1383	2133	3000	0
Models || Statistic	MSE	RMSE	MAE	R²	Adjusted-R²
MLR	15.187	3.8971	2.9127	0.4664	0.48932
FNN	9.0411	3.0068	2.1922	0.7228	0.75632

Figure 13. Measured ROP vs Predicted ROP of the FNN + MLR, case 4.

Fig. 15 appears cluttered due to the vast dataset collected, as the data points have different values and consistency from point to point. Each data point differs from the ones before and after because they belong to different oil wells. However, despite this variation, the overall result is the same in the other cases, and the FNN curve is more inclined than the MLR curve to integrate with the measured ROP curve.
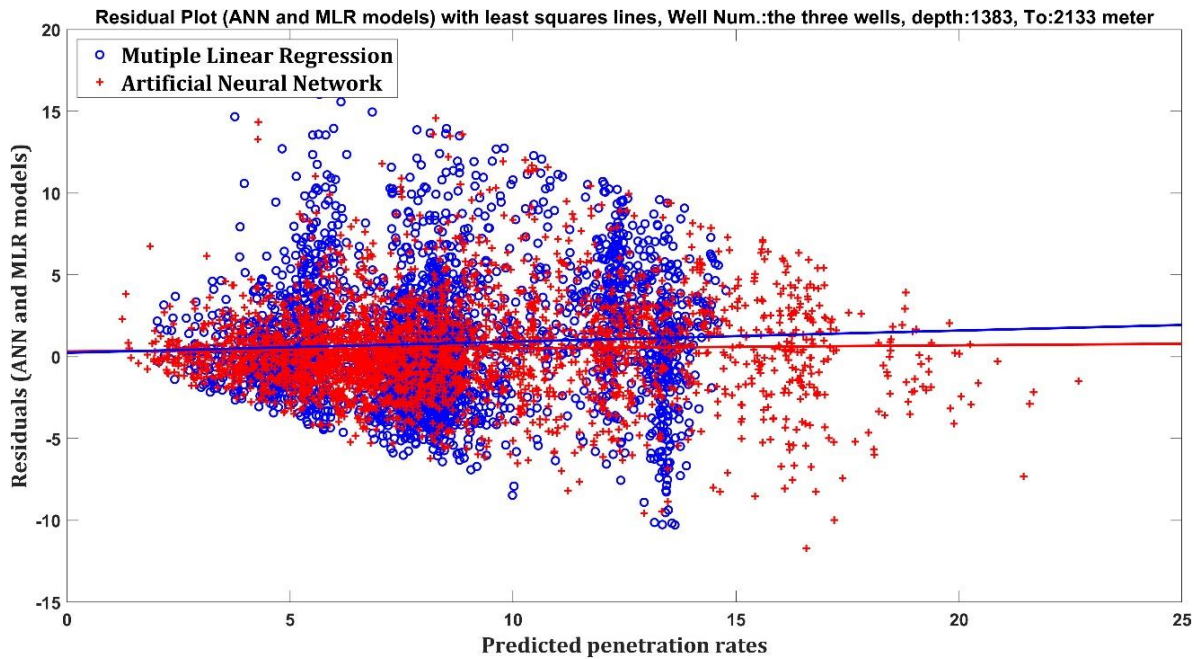

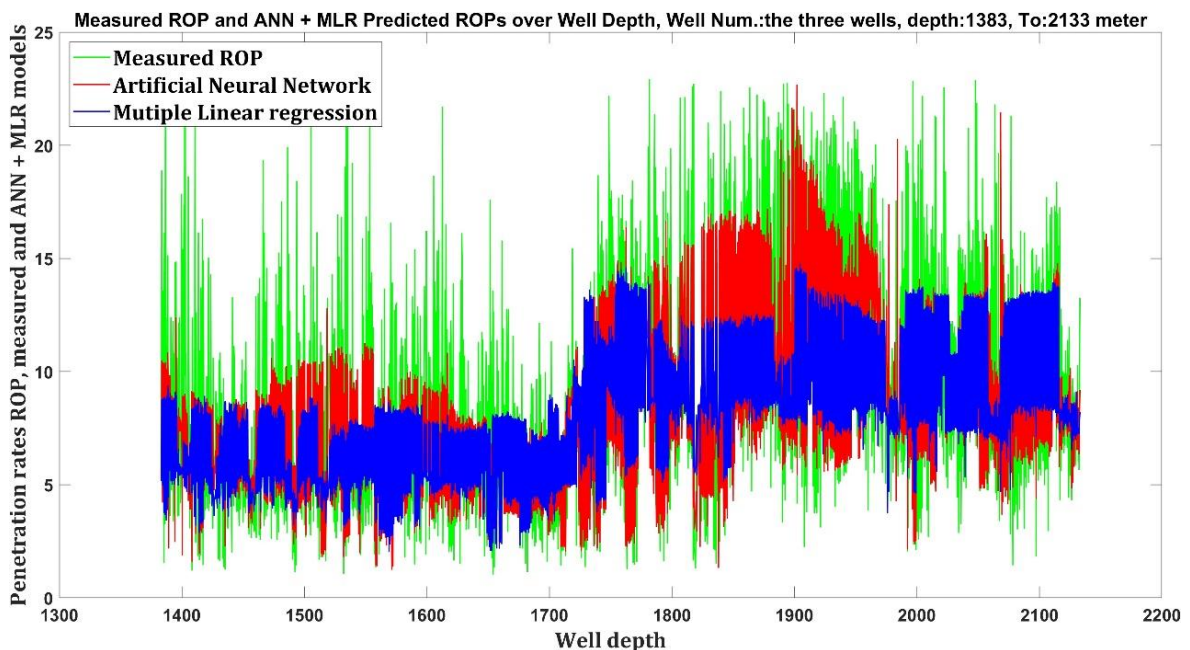**Figure 14.** Residuals vs Predicted ROP of the FNN + MLR, case 4.

Figure 15. Well Depth vs Measured + Predicted ROP of FNN + MLR, case 4.

5. CONCLUSIONS

Getting all the variables for BYM equations can be challenging because they need many controllable and uncontrollable variables, and BYM's linear assumption may not effectively capture complex interconnections. On the other hand, FNN requires high-quality datasets to train accurately. It can suffer from over-fitting or under-fitting for rubbish and outlier data points, making it challenging to strike the right balance. However, FNN can capture complex patterns that MLR may miss.

1. By integrating FNN with BYM equations, which depend on solid hypotheses, the model becomes better equipped to handle the complexities of the drilling process, including nonlinearity and interactions between various drilling parameters.
2. The experiment showed that using fewer data points as a dataset made the results more accurate, but using more data points as much as possible reflected the actual drilling circumstances in the well.
3. The present work results indicate improvement in the approach to real-time data testing. This is clear from the tables and graphs that compare everything involved, like R2, residuals, and ROP values. This finding demonstrates that combining BYM equations as a dataset for FNN enhances penetration prediction performance regardless of the number of data points available.
4. These study outcomes were obtained by using FNN as a kind of ANN and particular cross-validation. Better results may be obtained by using different kinds of ANN and other kinds of cross-validation so that more research can be developed.

NOMENCLATURES

Symbol	Description	Symbol	Description
a1 to *a8*	The regression coefficients.	*x1*	The effect of variables not considered in the BYM on ROP.
BYM	Bourgoyne and young model.	*x2*	The effect of increased rock strength due to normal compaction with depth on ROP.
D	Oil well, true vertical depth (ft).	*x3*	The impact of under-compaction experienced in abnormally pressured.
Fj	Bit Jet impact force (in pounds of force)	*x4*	Related to the effect of the hydrostatic and formation pressure differential on ROP.
gp	Pore pressure (ppg).	*x5*	The impact of bit weight on ROP.
h	Fractional tooth wear.	*x6*	Related to the drilling pipe rotary speed on ROP
MLR	Multiple linear regression.	*x7*	Models the impact of tooth wear on ROP.
N	Rotary speed (revolutions per minute, RPM).	*x8*	Models the effect of bit hydraulics on ROP.
ROP	BYM dependent variable, rate of penetration.	$[\frac{W}{d_b}]_t$	The drilling begins at Threshold WOB 1000lbf/inch.
Wb/db	(Weight on bit) per (bit diameter) (pounds per inch).	ρ	mud weight (ppg).
x1 to x8	BYM independent variables.	ρc	Equivalent circulation density (ppg).

endeavor. The support received from the Dean's Office of the College of Engineering and the head of the Petroleum Engineering Department is also greatly appreciated.

Credit Authorship Contribution Statement

Ebrahim Al-Assad: Writing the original draft, reviewing and editing, collecting the data, writing the MATLAB programming code, and working on it, including creating all the result tables and figures; Sameera M. Hamd-Allah: advisor, reviewing and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abdulmalek, A.S., Elkatatny, S., Abdulraheem, A., Mahmoud, M., Abdulwahab, Z.A., and Mohamed, I.M., 2018. Prediction of rate of penetration of deep and tight formation using support vector machine. *Paper presented at Conference, SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, April 23-26, SPE-192316-MS. https://doi.org/10.2118/192316-MS

Alsubaih, A., Albadran, F., and Alkanaani, N., 2018. Mechanical specific energy and statistical techniques to maximizing the drilling rates for production section of Mishrif wells in southern Iraq fields. *Paper presented in SPE/IADC Middle East Drilling Technology Conference and Exhibition,* Abu Dhabi, UAE, January 29-31, SPE-189354-MS. https://doi.org/10.2118/189354-MS

AL-Zirej, Z.J., and Hassan, A.H., 2020. An artificial neural network for predicting rate of penetration in AL-Khasib formation – Ahdeb Oil Field. *Iraqi Journal of Science*, 61(05), pp. 1051-1062. https://doi.org/10.24996/ijs.2020.61.5.14

Amar, K., and Ibrahim, A., 2012. Rate of penetration prediction and optimization using advances in artificial neural networks, a comparative study. *Paper presented in Proceedings of the 4th International Joint Conference on Computational Intelligence* - Volume 1: NCTA 647-652, Barcelona, Spain. https://doi.org/10.5220/0004172506470652

Amel, H.A., 2017. Bit records analysis for bits evaluating and selection. *Journal of Engineering,* 23(10), pp. 97–113. https://doi.org/10.31026/j.eng.2017.10.08.

Asad, S., Vusal, I., and Solomonov, D.X., 2022. Application of machine learning techniques for rate of penetration prediction. *Paper presented at the SPE Annual Caspian Technical Conference*, Nur-Sultan, Kazakhstan, November 15–17, SPE-212088-MS. https://doi.org/10.2118/212088-MS

Ayad, A.A., Safaa, H.S., and Amel, H.A., 2015. Bit performance in directional oil wells. Journal of Engineering, 21(11), pp. 80–93. https://doi.org/10.31026/j.eng.2015.11.05.

Bahari, M.H., Bahari, A., Nejati, M.F., and Naghibi, S.M.B., 2008. Determining bourgoyne and young model coefficients using a genetic algorithm to predict drilling rate. *Journal of Applied Sciences*, 8(17), pp. 3050-3054. https://doi.org/10.3923/jas.2008.3050.3054

Bingham, M.G., 1964. A new approach to interpreting rock drillability. *Book published by Petroleum Publishing Company,* Houston, Texas, USA, April.

Bourgoyne, J.A.T., Millheim, K.K., Chenevert, M.E., and Young J.F.S., 1986. Applied drilling engineering. *Book Published by SPE the Society of Petroleum Engineers,* EISBN: 978-1-61399-919-6. https://doi.org/10.2118/9781555630010

Bourgoyne. J.A.T., and Young J.F.S., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. *SPE, Society of Petroleum Engineers Journal,* 14 (4), pp. 371–384, SPE-4238-PA. https://doi.org/10.2118/4238-PA

Chengxi, L., Peng, C., and Chris, C., 2023. A comparison of machine learning algorithms for rate of penetration prediction for directional wells. *Paper presented at the SPE Middle East Oil, Gas and Geosciences Show,* Manama, Bahrain, February 19-21, SPE-213321-MS. https://doi.org/10.2118/213321-MS

Chris, C., 2023. Machine-learning model developed for rate-of-penetration optimization. *Journal of Petroleum Technology JPT*, 75(02), pp. 59–61, SPE-0223-0059-JPT. https://doi.org/10.2118/0223-0059-jpt

Darwish, A.K., Rasmussen, T.M., and Al-Ansari, N., 2020. Controllable drilling parameter optimization for roller cone and polycrystalline diamond bits. *Journal of Petroleum Exploration and Production Technology*, 10(04), pp. 1657–1674. https://doi.org/10.1007/s13202-019-00823-1

Eckel, J.R. 1967. Microbit studies of the effect of fluid properties and hydraulics on drilling rate. *J Pet Technol* Apr;19(4):541-546. *Trans. AIME*, Vol. 240. https://doi.org/10.2118/1520-PA

Elkatatny, S.M., Tariq, Z., Mahmoud, M.A., and Al-AbdulJabbar, A., 2017. Optimization of the rate of penetration using artificial intelligence techniques. *Paper presented at the 51st U.S. Rock Mechanics/Geomechanics Symposium,* San Francisco, California, June 25-28, ARMA-17-429.

Eren, T., and Ozbayoglu, M.E., 2010. Real-time optimization of drilling parameters during drilling operations. *Paper presented at the SPE Oil and Gas India Conference and Exhibition*, Mumbai, India, 20-22 January, SPE-129126-MS. https://doi.org/10.2118/129126-MS

Fei, Z., Honghai, F., Yuhan, L., Hongbao, Z., and Rongyi, J., 2023. Hybrid model of machine learning method and empirical method for rate of penetration prediction based on data similarity. *Applied Sciences*, 13(10), pp. 5870. https://doi.org/10.3390/app13105870.

Galle, E.M., and Woods, H.B., 1963. Best constant weight and rotary speed for rotary rock bits. *Book published by Hughes Tool Company*.

Graham, J.W., and N.L. Muench. 1959. Analytical determination of optimum bit weight and rotary speed combinations. *Paper presented at the Fall Meeting of the Society of Petroleum Engineers of AIME*, Dallas, Texas, October 4-7. https://doi.org/10.2118/1349-G

Hamad-Allah, S.M., and Ismael, A.A., 2008. Application of mathematical drilling model on southern Iraq oil fields. *Journal of Engineering*, 14(03), pp. 2763–2784. https://doi.org/10.31026/j.eng.2008.03.17.

Hassan. A.H., 2015. Correlation of penetration rate with drilling parameters for an Iraqi field using mud logging data. *Iraqi Journal of Chemical and Petroleum Engineering IJCPE*, 16(03), pp. 35–44. https://doi.org/10.31699/IJCPE.2015.3.4.

Hongtao, L., Yan, J., Xianzhi, S., and Zhijun, P., 2022. Rate of penetration prediction method for ultra-deep wells based on LSTM–FNN. *Applied Sciences*, 12(15), pp. 7731. https://doi.org/10.3390/app12157731.

Hou, B., Chen, M., and Yuan, J., 2014. Optimization and application of bit selection technology for improving the penetration rate. *Research Journal of Applied Sciences Engineering and Technology*, 8(2), pp. 179-187, https://doi.org/10.19026/rjaset.8.957.

Huihui, J.i., Yi, S.L., Shuting, C., Zelong, X., and Liang, Z., 2022. An advanced long short-term memory (LSTM) neural network method for predicting rate of penetration (ROP). *ACS Omega*, 8(1), pp. 934–945. https://doi.org/10.1021/acsomega.2c06308

Irawan, S., Abd-Rahman, A.M., and Tunio, S.Q., 2012. Optimization of weight on bit during drilling operation based on rate of penetration model. *Research Journal of Applied Sciences, Engineering and Technology*, 4(12), pp. 1690–1695.

Kareem, N., Al-Khafaji, A., Sadooni, F. 2022. Evaluation of petrophysical characteristics of Mishrif and Yamama reservoirs, in Garraf Oil Field, Southern Iraq, Based on Well-Logging Interpretation. *Iraqi Journal of Science*. 63(3), pp. 115-1128. https://doi.org/10.24996/ijs.2022.63.3.19

Liu, Y., Kibbey, J., Bai, Y., and Wu, X., 2018. Real-time bit wear monitoring and prediction using surface mechanics data analytics: a step toward digitization through agile development. *Paper presented at the IADC/SPE Drilling Conference and Exhibition,* Fort Worth, Texas, USA, March 6-8, SPE-189602-MS. https://doi.org/10.2118/189602-MS.

Majid, M., and Ayad, A.A., 2019. Selection of suitable drilling parameters for obtaining high rate of penetration in Majnoon oilfield. *Iraqi Journal of Chemical and Petroleum Engineering IJCPE,* 20(1), pp. 65-68. https://doi.org/10.31699/IJCPE.2019.1.9.

Maurer W.C., 1962. The 'perfect cleaning' theory of rotary drilling. *Journal of Petroleum Technology* JPT, 14 (11), pp. 1270–1274. https://doi.org/10.2118/408-PA.

Miyora, T.O., 2014. Modeling and optimization of geothermal drilling parameters: a case study of well MW-17 in Menengai, Kenya. *Master thesis, United Nations University, Tokyo, Japan, ISBN 9979683449*.

Mochizuki, S., Saputelli, L.A., Kabir, C.D., Cramer, R., and Lochmann, M.J., 2004. Real-time optimization: classification and assessment. *SPE Production & Operation*, 21 (04), pp. 455–466, SPE-90213-PA. https://doi.org/10.2118/90213-PA.

Moradi, H., Bahari, M.H., Sistani, M.B.N., and Bahari, A., 2010. Drilling rate prediction using an innovative soft computing approach. *Scientific Research and Essays*, 5(13), pp. 1583-1588.

Muhammad, T.F., Irawan, S.B., Taufan, M., Pri, A.R., Marmora, T.M., Asri, N., and Onnie, R., 2023. Application of artificial neural network to estimate rate of penetration for geothermal well drilling in south Sumatera. *International journal Emerging technology and advanced engineering*, 13, pp. 135-140. https://doi.org/10.46338/ijetae0323_14

Ololade, A., Ibiye, I., and Kingsley, A., 2021. Comparative evaluation of artificial intelligence models for drilling rate of penetration prediction. *Paper presented at the SPE Nigeria Annual International Conference and Exhibition,* Lagos, Nigeria, August 2-4, SPE-208451-MS. https://doi.org/10.2118/208451-MS.

Padhma, M., 2023. A comprehensive introduction to evaluating regression models. *Data Science Blogathon updated On October 31st*.

Petronas operation, 2016. Notice of operation for Al-Garraf, Ga-J38P oil well, on-shore Iraq. *Dhi-qar oil campany library,* land drilling, IDC: 55.

Speer, J.W.A., 1959. A method for determining optimum drilling techniques. *Paper presented at the Gulf Coast Drilling and Production Meeting,* USA, LA, April 24, SPE-1242-G. https://doi.org/10.2118/1242-G.

Teeba, J.S., and Farqad, H., 2022. Development of artificial intelligence models for estimating rate of penetration in east Baghdad field, middle Iraq. *Iraqi Geological Journal*, 55(1C), pp. 112-124. https://doi.org/10.46717/igj.55.1C.9Ms-2022-03-28.

Tong, J., Jie, C., and Dan, S., 2023. A comparative study of deep learning methods for drilling performance prediction. *Proceedings SPIE 12604, Paper presented in International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID)*, Guangzhou, China, 23 May, 126044A, pp. 1099-1105. https://doi.org/10.1117/12.2674979.

Warren TM. 1987. Penetration-rate performance of roller-cone bits. *SPE Drilling Engineering*; 2:9–18. https://doi.org/10.2118/13259-PA.

Yasser, A.K., Fadhil, S.K., and Yousif, K.Y., 2020. Using adaptive neuro fuzzy inference system to predict rate of penetration from dynamic elastic properties. *Journal of Engineering*, 26(7), pp. 45–61. https://doi.org/10.31026/j.eng.2020.07.04.

Zha, Y., Ramsay, S., and Pham, S., 2018. Real-time surface data-driven wob estimation and control. *Paper presented at the SPE Annual Technical Conference and Exhibition,* Dallas, Texas, USA, September 24-26, SPE-191723-MS. https://doi.org/10.2118/191723-MS

Appendix A

The following is a small part of the big MATLAB programming codes that I wrote to make the calculations, statistics, and plots; this part is related to FNN equations codes in MATLAB.

```
%% Load the ROP dataset (drilling parameters and corresponding ROP values)
%% load('rop_dataset.mat'); %% Replace 'rop_dataset.mat' with your dataset file then
%% calculate each data point depending on B&Y equations and this done by a lot of
%% programming codes
%% Split the dataset into training, validation and testing sets
trainRatio = 0.8;          % 80% for training
validationRatio = 0.1;    % 10% for validation
testRatio = 0.1;           % 10% for testing
 [trainInd, valInd, testInd] = dividerand(size(X, 1), trainRatio, validationRatio, testRatio);
 X_train = X(trainInd, :);
 Y_train = Y(trainInd);
 X_val = X(valInd, :);
 Y_val = Y(valInd);
 X_test = X(testInd, :);
 Y_test = Y(testInd);
%% Create the FNN model
hiddenLayerSize = 50; % Number of neurons in the hidden layer
net = fitnet(hiddenLayerSize);
%% Set the training parameters
net.trainParam.epochs = 100; % Number of training epochs
net.trainParam.lr = 0.01; % Learning rate
net.trainParam.min_grad = 1e-6; % Minimum gradient threshold for stopping training
net.trainParam.showWindow = false; % Disable training window display
%% Train the FNN model
net = train(net, X_train', Y_train');
%% Evaluate the performance of the validation set
 Y_val_pred = net(X_val');
 mse_val = mean((Y_val - Y_val_pred').^2);
 rmse_val = sqrt(mse_val);
 R2_val = 1 - (sum((Y_val - Y_val_pred').^2) / sum((Y_val - mean(Y_val)).^2));
%% Generate predictions for the testing set
Y_test_pred = net(X_test');
mse_test = mean((Y_test - Y_test_pred').^2);
rmse_test = sqrt(mse_test);
R2_test = 1 - (sum((Y_test - Y_test_pred').^2) / sum((Y_test - mean(Y_test)).^2));
%% Then the section on calculations and measuring MLR, by many programming codes
%% Then the section on making different kinds of comparison graphs and statistical
%% results and values by many programming codes
```

تطبيق الشبكات العصبية الاصطناعية مع موديل بورغوين ويونغ للتنبؤ بمعدل الاختراق في حقل الغراف النفطي

إبراهيم العصاد*، سميرة حمد الله

قسم هندسة النفط، كلية الهندسة، جامعة بغداد، بغداد، العراق

الخلاصة

يعتبر توقع معدل الاختراق (ROP) لحفر آبار النفط أمرًا ضـــــروريًا لعمليات الحفر الآمنة والفعالة من حيث التكلفة؛ تسـعى شـــركات الحفر إلى تخمين (ROP) منذ العقود الأولى لهذه الصـــناعة. ولتحقيق هذا الهدف، ومن بين العديد من النماذج، تم تطوير نموذج بورغوين ويونغ (BYM)، وهو نهج قائم على المعادلات، واسـتخدم على نطاق واسـع للتنبؤ بـــ (ROP) على أسـاس الانحدار الخطي المتعدد (MLR). تعد الشـبكة العصبية الاصطناعية (ANN) إحدى تقنيات التعلم الآلي التي تحلل بيانات الحفر وتقوم بتنبؤات (ROP). تم إجراء العديد من الدراسـات في جميع أنحاء العالم لاسـتخدام (BYM)، بينما هدفت دراسات أخرى إلى تحسينها في ظروف مختلفة. وقد أظهرت (ANN) أيضًا فعاليتها في هذه المجالات. تقترح هذه الدراسة نهجا يجمع بين فوائد الشبكات العصبية التغذية الأمامية (FNN) من نموذج (ANN) ومعادلات (BYM) لتعزيز التنبؤ بـ (ROP). يؤدي دمج (BYM) مع (FNN) إلى تعزيز النموذج القائم على المعادلة ويعمل أيضًا على تسـخير قوة وكفاءة التعلم الآلي. ويؤدي هذا إلى تحسن كبير في الدقة والكفاءة في التنبؤ بـــ (ROP) في آبار النفط. تتضمن معلمات إدخال نمذجة (ROP) العمق الرأسي الإجمالي المقاس والحقيقي، والوزن على الحفرة، والدوران في الدقيقة، وضغط الأنبوب، وتدفق المضـخة، ووزن الطين المكافئ، وحجم الحفرة، وحجم الفوهة، وضغط التكوين، وقوة تأثير نفاثة الحفرة، والتي يتم إعادة حسابها بواسطة معادلات (BYM) وإدخالها إلى كل من (MLR) و(FNN).عند اختبارها على بيانات في الوقت الفعلي من حقل الغراف، تظهر النتائج R2 أعلى، وبقايا أقل، وقيمة P صفرية مقارنة بـ (MLR) مما يثبت دقة النهج ويوفر تنبؤًا دقيقًا بمعدل الانحدار الخطي في خطط الحفر المستقبلية.

الكلمات المفتاحية: الشبكات العصبية الاصطناعية، نموذج بورغوين ويونغ، الانحدار الخطي المتعدد، معدل الاختراق، حقل الغراف النفطي.