



OPEN

A hybrid data-driven solution to facilitate safe mud window prediction

Ahmed Gowida, Ahmed Farid Ibrahim & Salaheldin Elkatatny✉

Safe mud window (SMW) defines the allowable limits of the mud weights that can be used while drilling O&G wells. Controlling the mud weight within the SMW limits would help avoid many serious problems such as wellbore instability issues, loss of circulation, etc. SMW can be defined by the minimum mud weight below which shear failure (breakout) may occur (MW_{BO}) and the maximum mud weight above which tensile failure (breakdown) may occur (MW_{BD}). These limits can be determined from the geomechanical analysis of downhole formations. However, such analysis is not always accessible for most drilled wells. Therefore, in this study, a new approach is introduced to develop a new data-driven model to estimate the safe mud weight range in no time and without additional cost. New models were developed using an artificial neural network (ANN) to estimate both MW_{BO} and MW_{BD} directly from the logging data that are usually available for most wells. The ANN-based models were trained using actual data from a Middle Eastern field before being tested by an unseen dataset. The models achieved high accuracy exceeding 92% upon comparing the predicted and observed output values. Additionally, new equations were established based on the optimized ANN models' weights and biases whereby both MW_{BO} and MW_{BD} can be calculated without the need for any complicated codes. Finally, another dataset from the same field was then used to validate the new equations and the results demonstrated the high robustness of the new equations to estimate MW_{BO} and MW_{BD} with a low mean absolute percentage error of 0.60% at maximum. So, unlike the costly conventional approaches, the newly developed equations would facilitate determining the SMW limits in a timely and economically effective way, with high accuracy whenever the logging data are available.

Abbreviations

ANN	Artificial neural network
MW_{BO}	Minimum mud weight below which shear failure may occur
MW_{BD}	Maximum mud weight above which tensile failure may occur
SMW	Safe mud window
P_f	Formation pore pressure
P_w	Hydrostatic mud pressure
GR	Gamma-ray logging data
RHOB	Formation bulk density log data
DTC	Compressional wave transit-time logging data
DTS	Shear wave transit-time logging data
NPHI	Neutron porosity logging data
CALI	Caliper log data
E	Elastic modulus
PR	Poisson's ratio
UCS	Unconfined compressive strength
T_s	Tensile strength
S_{rr}	Radial stress
$S_{\theta\theta}$	Hoop stress
S_{zz}	Axial stress
β	Failure angle

Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia.
✉email: elkatatny@kfupm.edu.sa

trainlm	Levenberg–Marquardt
trainbr	Bayesian regularization backpropagation
trainrp	Resilient Backpropagation
trainscg	Scaled Conjugate Gradient
traincgb	Conjugate Gradient with Powell/Beale Restarts
traincgf	Fletcher–Powell Conjugate Gradient
traincgp	Polak–Ribière Conjugate Gradient
trainoss	One Step Secant
traingdx	Variable Learning Rate Backpropagation
Tansig	Hyperbolic tangent sigmoid transfer function
Logsig	Log-sigmoid transfer function
ElliotSig	Elliot symmetric sigmoid transfer function
Radbas	Radial basis transfer function
Hardlim	Symmetric hard-limit transfer function
Satlin	Symmetric saturating linear transfer function

Geomechanics is among the sciences that represent a cornerstone when it comes to planning and optimization of drilling and development of petroleum fields. It comprises studying the mechanical properties of the rocks in addition to the distribution of the in-situ stresses in the vicinity of the wellbore. Careful estimation of such information is considered a key factor that could help avoid a broad range of costly issues that may be encountered while drilling, completion, and stimulation operations. Geomechanics of the subterranean formations can be studied by building geomechanical models that mimic the in-situ geomechanical state of the formations¹. This could contribute to the technical investigation of different processes and in turn facilitate cost reduction. The geomechanical model can be constructed using conventional logging data, i.e., formation density logs, porosity logs, acoustic logs, etc. in addition to some field tests and core-based lab experiments for calibration purposes. Wellbore stability is considered among the main concerns during the drilling process. Two common issues relate to wellbore stability: tensile failure (fractures) or wellbore breakdown (collapse). As a result, many problems may occur loss of circulation, pipe sticking, washouts, etc. Therefore, care should be taken while well planning for determining the safe mud window to avoid such costly and wasting-time events. The safe mud window is typically determined in light of the in-situ stress state of the downhole formations². The formation stresses are described in terms of both magnitude and direction whereby the probability of wellbore-instability problems can be evaluated. Accordingly, accurate data on the in-situ stress configuration are usually required along the well depth to come up with viable solutions for potential instability issues while drilling. This can be attained by deploying the optimal mud weight that exists within the range of the predefined safe mud window (SMW)³. SMW can be determined by applying geomechanical modeling based on well data including the in-situ stresses, pore pressure, and rock mechanical properties coupled with suitable rock failure criteria⁴. However, such analysis is considered costly and time-consuming which limits its accessibility for most of the drilled wells.

Recently, similar to different industries (Haghighat and Li, 2021), there is a growing need for automated intelligent systems that can make use of the availability of enormous data during different operations in the O&G industry. These systems are based on different machine learning approaches that aim at modeling and estimating different key parameters in a cost-effective and time-efficient way⁵. Many studies in the literature have employed machine learning approaches for estimating different geomechanics-related parameters that would be used for designing SMW. In 2012, Rabbani et al. developed a model using neural networks to predict the unconfined compressive strength (UCS) which is one of the key parameters in earth geomechanical modeling, using the well log data such as density, porosity, etc. from an oilfield in southern Iran. Pereira et al. (2013) introduced a model based on decision trees to predict the appropriate mud weight for safely drilling salty formations. They used drilling records (e.g., date/depth, observed mud weight), lithology types encountered and the records of the incidents, e.g., pipe sticking due to formation swelling. Tabaeh and Mohammad (2016) also developed a neural network-based model to predict the shear wave velocity based on the logging data. Then they incorporated this information into a new workflow to estimate the least principal stresses in a tectonically active region. Zhou et al. (2016) developed a new model to predict the adequate equivalent circulating density (ECD) to control the downhole pressures under HPHT conditions. They used the type of the drilling mud, fractions of oil and water phases in the used mud in addition to the formation conditions of pressure and temperature as input features to their developed neural network. Okpo et al. (2016) investigated wellbore instability in a Nigerian field by deploying a neural network approach. They used the mud properties, formation stresses, bottom-hole pressure, pore pressure, and borehole diameter as inputs to determine the stability of the downhole formations. Zahiri et al. (2019) studied the safe mud window by developing a mechanical earth model of an Iranian field. They then employed the neural network approach to predict the failure criterion based on the logging data in a black-box model. Recent studies on SMW have applied rock wall stability/failure assessment models under tensile, shear and in situ stresses combined with ML models^{6,7}. Though Phan et al.⁷ reported high prediction accuracy, they used several input parameters that are not always accessible for most oilwells such as principal stresses, elastic modulus, Poisson's ratio, cohesion strength, tensile strength. The need for such parameters limits the applicability of these models in the wells where such data are not available. In⁶, Abbas et al., introduced the application of ML to predict SMW from logging data. The study lacks any error metrics that are essential to quantify the prediction error and the model's accuracy. Based on the review of the literature, the existing intelligent models applied to predict SMW leave scope for improvements in accuracy and applicability which should beneficially contribute to safe and optimized drilling operations.

Therefore, in this study, the safe mud window (SMW) has been investigated using a machine learning approach (artificial neural network, ANN). SMW has been determined by studying the minimum and maximum mud weight limits beyond which either shear or tensile failure may be encountered respectively. This study aims at predicting the SMW from the logging data directly using an ANN-based model for a Middle Eastern field. Besides, the developed ANN model has been presented in a white-box mode by establishing new equations that imitate the processing of the developed ANN.

The new approach was established based on the analysis of the geomechanical data to enable the direct determination of the safe range of drilling mud weights whenever the logging data are available. Unlike the conventional high-priced approach, the newly developed equations would facilitate estimating the MW_{BO} and MW_{BD} directly from the logging data in an economic and time-effective way. This, in turn, would help the petroleum engineers design adequate mud properties whereas many wellbore instability problems can be avoided. In addition, drilling planning and optimizations could be improved further.

In the upcoming section, the methodology adopted for this study is illustrated followed by a detailed demonstration of the models' development process and their optimization. Finally, the results of this research are discussed in the discussion section with summarized outcomes in the conclusions section.

Materials and methods

Methodology. Rock mechanics refers to the science of investigating the geomechanical behavior of the rocks either under elastic or failure states. These behaviors are expressed in terms of both elastic parameters, i.e., elastic moduli (E), Poisson's ratio (PR), etc. in addition to failure parameters, i.e., unconfined compressive strength (UCS), friction angle (ϕ), tensile strength (T_s), etc. Such parameters are basically incorporated while developing the mechanical earth model (MEM) whereby the geomechanical behavior of the subterranean formations can be studied. Firstly, data have been collected for constructing MEM from three wells within a Middle Eastern field. Three groups of data were included, these are:

- Petrophysical logging data, e.g., formation bulk density (RHOB), sonic data (DTC and DTS), gamma-ray log (GR), neutron logging (NPHI), and caliper logs (CALI).
- Core data are based on the experimental tests conducted on the retrieved core samples. These data are usually used for model validation and calibration. The calibration means identifying the possible relations among the dynamic elastic properties that are estimated from the well logs and the corresponding static ones that are measured in the lab.
- In-drilling data that are collected during the drilling operation such as mud losses, and wellbore instability issues reporting.

Secondly, the dynamic elastic parameters (E and PR) have been estimated using the petrophysical data and then calibrated into the corresponding static ones using the available core data. This led to having continuous profiles of static E and PR along with the depth of the studied wells. Next, the failure parameters (UCS, T_s , and ϕ) have been correlated to obtained well logs to get them in a continuous-profile form. These steps are essential for developing MEM that effectively represents the geomechanical behavior of the downhole formations.

Then, the stress field (e.g., overburden stress S_v , maximum and minimum horizontal stresses S_{hmin} and S_{hmax}) has been determined using the dual poroelastic models. More information on the determination of the formation stresses for the field under study are available in the published studies^{8–10}.

When a well is drilled into a formation, stressed solid material is removed. The borehole wall is then supported only by the hydrostatic pressure of the drilling mud in the hole¹¹. As this fluid pressure generally does not match the in-situ formation stresses, there will be a stress redistribution around the well, known as the post-drilling stress field or induced stress field. The secondary induced stresses: hoop stress ($S_{\theta\theta}$), radial stress (S_{rr}), and axial stress (S_{zz}) have been calculated based on the S_{hmin} and S_{hmax} using the following equations²:

$$S_{\theta\theta max} = 3S_{Hmax} - S_{hmin} - P_w - P_f \quad (1)$$

$$S_{\theta\theta min} = 3S_{Hmin} - S_{hmax} - P_w - P_f \quad (2)$$

$$S_{zz max} = S_v + 2PR(S_{hmax} - S_{hmin}) - P_f \quad (3)$$

$$S_{zz min} = S_v - 2PR(S_{hmax} - S_{hmin}) - P_f \quad (4)$$

$$S_{rr} = P_w - P_f, \quad (5)$$

* P_w : Hydrostatic mud pressure, * P_f : Formation pore pressure.

There are two main scenarios through which the borehole may fail, depending on the relative magnitudes of these secondary stresses at the borehole wall. First, a shear failure that results from using underweighted drilling mud compared to the formation pressure, leads to breakout failure. Second, a tensile failure that results from using overweighted drilling mud, leads to induced fractures (breakdown failure) and hence partial or complete losses.

At this stage, wellbore stability analysis has been undertaken to determine SWM by estimating three key parameters:

Case	Stress state	Borehole failure occurs when
1	$S_{\theta\theta} \geq S_{zz} \geq S_{rr}$	$P_w \leq \frac{A-C}{1+q}$
2	$S_{zz} \geq S_{\theta\theta} \geq S_{rr}$	$P_w \leq \frac{B-C}{1+q}$
3	$S_{zz} \geq S_{rr} \geq S_{\theta\theta}$	$P_w \leq A - C - qB$

Table 1. Borehole shear failure (Breakout) criterion to determine MW_{BO} for borehole stability. Where; $A = 3S_{Hmax} - S_{hmin}$, $B = S_v + 2PR(S_{hmax} - S_{hmin})$, $C = UCS - P_p(\tan^2(\beta) - 1)$, $q = \tan^2(\beta)$, where, A, B, and C are constants that are calculated based on other parameters, i.e., S_{hmin} , S_{hmax} , S_v and PR. UCS refers to the uniaxial compressive strength and β denotes the failure angle.

Parameter	GR (API unit)	DTC (μ s/ft)	DTS (μ s/ft)	RHOB (g/cm^3)	NPHI	MW_{BO} (lb/ft ³)	MW_{BD} (lb/ft ³)
Minimum	3.38	44.89	81.28	2.32	0.28	92.73	149.82
Maximum	85.79	66.12	132.28	3.04	0.32	98.01	167.71
Mean	29.25	48.55	90.26	2.83	0.30	95.11	156.23
STD	15.12	2.91	7.00	0.11	0.01	0.79	3.11
Skewness	4.12	44.82	81.28	2.38	0.28	92.73	150.15

Table 2. Descriptive statistical analysis of the dataset used in this study.

- Pore pressure (P_f) that are usually collected during well testing.
- Breakout mud weight (MW_{BO})
- Breakout mud weight (MW_{BD}).

For determining the breakout mud weight (MW_{BO}), in practice, different permutations of the stresses $S_{\theta\theta}$, S_{rr} , and S_{zz} should be considered while studying the mechanical stability of the region and then the largest borehole mud pressure at failure would be selected as a breakout failure criterion to determine MW_{BO} . The equations for these permutations are summarized in (Table 1)¹¹.

In addition, the tensile failure may be encountered when a significantly high mud weight is used that may exceed the value given by Eq. (6). This MW_{BD} value has been considered as the breakdown failure criterion^{1,11}. Finally, the MW_{BD} and MW_{BO} profiles were generated and used as the outputs for the proposed models.

$$MW_{BD} = 3S_{Hmax} - S_{hmin} - P_f + T_s \quad (6)$$

Data analysis. *Data description.* In this study, a dataset of 1858 actual observations are gathered from three wells in a Middle Eastern field representing a complex carbonate reservoir. The data included the petrophysical logging data: GR, RHOB, DTC, DTS, and NPHI and the corresponding estimated MW_{BO} and MW_{BD} at the same depth. The safe mud weight limits were determined based on the analysis of the developed MEM for the area under study as described in the “Methodology” section to eventually determine the MW_{BO} and MW_{BD} profiles. The data collected from two wells were used for training and testing the model (divided at a certain ratio) while the rest of the data (the third well) have been used for validating the developed models.

Data pre-processing and filtration. The prediction accuracy of the AI-based models is significantly affected by the quality of the data used while developing the model. Therefore, the data were pre-processed and filtered using statistical analysis and engineering sense based on the literature. A specially designed Python code was used to remove unreasonable values like zeros and negative values in addition to any missing points. Then, outliers were removed using a box and whisker plot, in which the top whisker represents the upper limit of the data and the bottom whisker represents the lower limit of the data¹². Any value beyond these limits was considered an outlier and then was removed. These limits were determined using the statistical parameters listed in Table 2 which lists the descriptive statistical summary of the dataset used in this study to reflect its distribution and covered ranges. Figure 1 shows the histogram distributions of the input and output data.

Dimensionality reduction analysis. The collinearity between each parameter and others is presented in the heat map shown in Fig. 2, in the way that the darker the color is, the higher degree of collinearity between the two parameters. The values presented in the heatmap represent the Pearson correlation coefficient (R-value) between every two parameters. The correlation coefficient is used to identify how strongly two parameters are linearly related to each other. Its value ranges from -1 to $+1$. A strong direct linear relation is indicated with an R-value of $+1$. On the contrary, the R-value of -1 shows a strong inverse linear relationship between these two variables. While an R-value of zero indicates no linear relationship exists between the two studies' variables. Moreover, Fig. 2 displays P-values associated with the correlation coefficient between the input and the output parameters to examine the significance of the correlation between the input and the outputs in the regression process. The

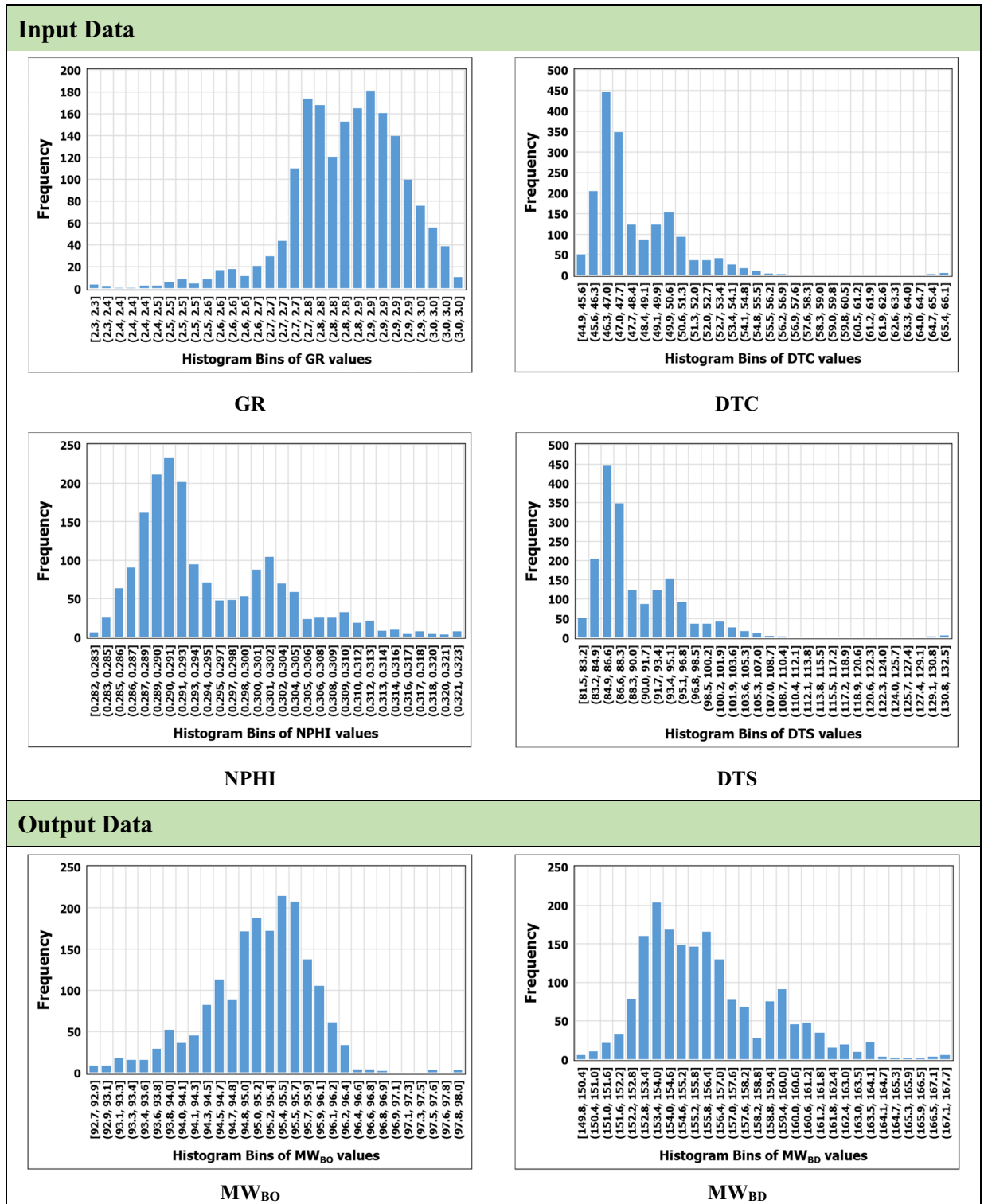


Figure 1. Histogram of the data used in this study.

DTC has shown a high correlation coefficient of 1 and 0.95 with DTS and NPHI respectively. Thus, DTC has been only considered as an input feature for the developed model to avoid the redundancy of the input information to the proposed models. These results were also confirmed by the p-values, where p-values were less than 0.05 (that means it is significant) except for the DTS and NPHI. Accordingly, the final set of the selected input features was GR, DTC, and RHOB.

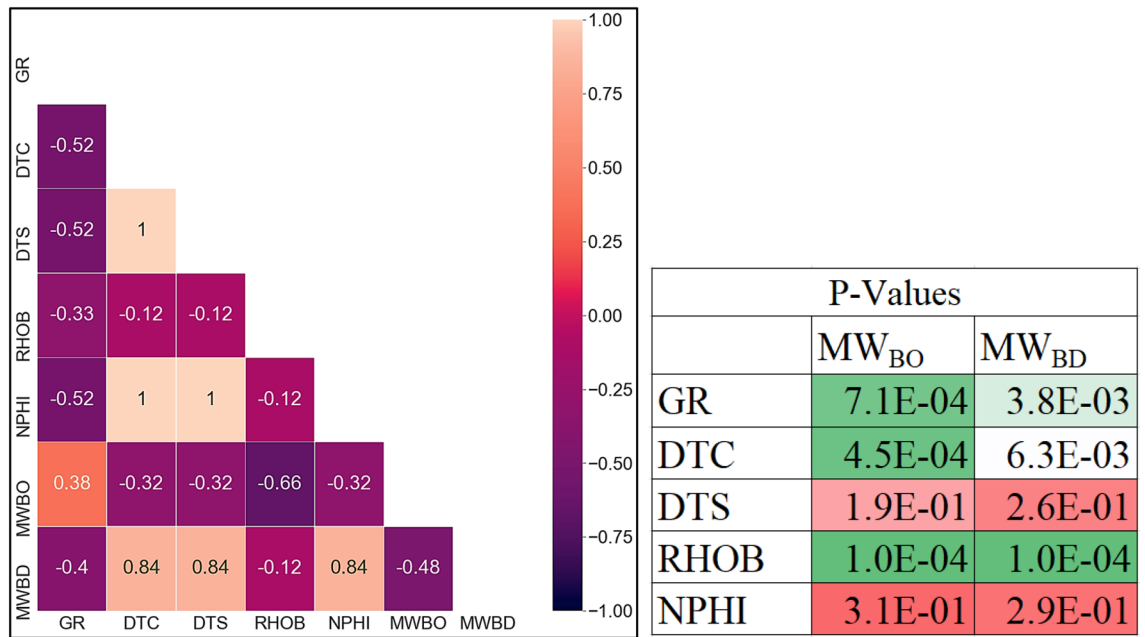


Figure 2. Heat map reflecting the collinearity among the input/output parameter with the P-values associated with the correlation coefficients.

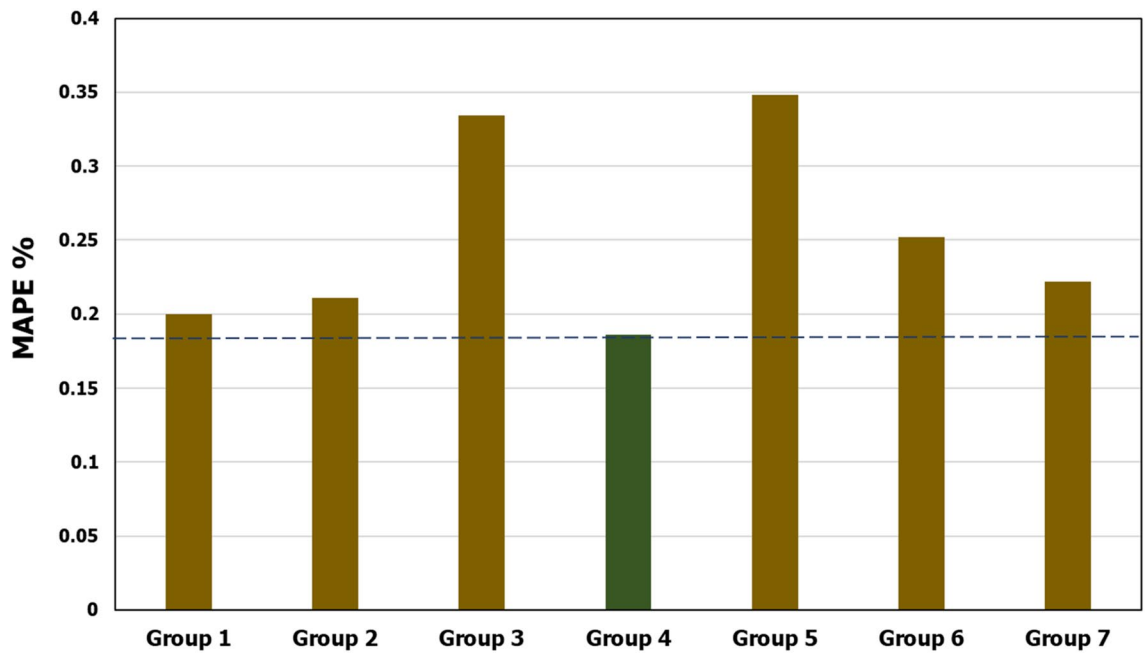


Figure 3. Comparison of the prediction error for the tested input groups.

Sensitivity analysis. In addition to studying the collinearity between the parameters to select the most effective ones, a sensitivity analysis has also been conducted. Different trials have been conducted using different sets of input parameters to predict MW_{BO}. This step was to get initial insights into the effect of each input parameter on the prediction results. Different groups of input parameters have been tested. These groups are Group 1 (GR, DTC, DTS, RHOB, NPHI), Group 2 (GR, DTC, DTS, RHOB), Group 3 (GR, DTC, DTS), Group 4 (GR, DTC, RHOB), Group 5 (GR, DTC), Group 6 (GR, RHOB) and Group 7 (DTC, RHOB). The results have been evaluated in terms of mean absolute percentage error (MAPE) between the actual and estimated output values. Figure 3 shows a comparison of the prediction error for the tested input groups. The results showed that the combination of GR, DTC, and RHOB (Group 4) yielded the lowest MAPE and hence the best prediction accuracy. This observation agrees with the conclusion of the dimensionality reduction analysis that ended with using

Parameter	Tested options/ranges			Optimized parameters	
				MW _{BO} model	MW _{BD} model
Number of hidden layers	1–4			1	
Number of neurons in each layer	5–40			10	7
Split Ratio	70–90% (For training set) The rest was used for testing			70 (train) : 30 (test)	
Training algorithms	trainlm	trainbr	trainrp	trainbr	
	trainscg	traincgb	traincgf		
	traincgp	trainoss	trainidx		
Transfer function	tansig	logsig	elliotsig	tansig	
	radbas	hardlim	satlin		
Learning rate	0.01–0.9			0.12	

Table 3. The tested options for optimizing the developed ANN models.

GR, DTC, and RHOB combination as inputs for developing the proposed models. Accordingly, this set of inputs has been decided for developing the proposed models.

Model development. In this research, artificial neural network (ANN) has been applied to develop a new model that can directly predict both MW_{BO} and MW_{BD} based on well-logging data. ANN has been selected due to its recent successful application in petroleum-related geomechanics. Many studies in the literature have reported the application of ANN to predict several geomechanical parameters to be used in the petroleum field context such as UCS prediction^{13–15}, elastic moduli, and Poisson's ratio^{16–18}, stress field prediction^{8,9,19}, etc. Typically, ANN consists of three basic types of layers: the input layer including GR, DTC, and RHOB, the hidden layer(s), and the output layer comprising MW_{BO} and MW_{BD}. Training the network starts with processing the data from the hidden layer through weighted connections to the neurons in the hidden layer to ultimately estimate the outputs²⁰. For optimizing the network, *GridSearchCV* function was designed using Python code to iteratively test different parameters for optimizing each algorithm within the predefined ranges by the user, to report the best value of each parameter for each algorithm. The obtained data has been divided into two sets: training and testing. The training set was used to train the model and optimize its hyper-parameters. During the optimization process, the results of the models were internally tested to evaluate the selected hyper-parameters. For each trial, the predictions were evaluated using the R-value and the error between the actual and predicted output values for the training, and testing processes. The objective of this step is to identify the hyper-parameters that could achieve the lowest possible prediction error through many iterative trials. Afterwards, the model with the optimized hyper-parameters was evaluated using the testing set to estimate the generalization error of the optimized model²¹.

The prediction results have been evaluated in terms of correlation coefficient (R), MAPE, mean squared error (MSE), and root means squared error (RMSE) between the observed and predicted outputs. The mathematical formulas for the evaluation metrics are stated in Appendix A. Table 3 lists the predefined options of the ANN parameters (e.g., training algorithm, transfer function, number of neurons, number of hidden layers, and learning rate) that have been tested during the tuning process. Figure 4 shows a typical schematic architecture of the developed ANN model. Figure 5 shows the workflow followed in this study to develop the proposed equations.

Results and discussion

Safe mud window prediction. Two models have been developed using ANN to predict both MW_{BO} and MW_{BD}. The developed models have been optimized by selecting the tuning parameters that yielded the lowest MAPE and the highest R between the observed and predicted output sets.

For MW_{BO}, the optimized model comprised one hidden layer with 10 neurons and was trained using 70% of the dataset while the rest of the data was used for testing the prediction performance. While the best results for the MW_{BD} model were attained using 7 neurons in a single hidden layer. The training process of the optimized MW_{BD} model was implemented using the 70:30 splitting ratio of the dataset for training and testing, respectively. Both models have been trained using the Bayesian regularization backpropagation (trainbr) algorithm and tan-sigmoidal (tansig) transfer function with a learning rate of 0.12.

The selected hyper-parameters for both models are summarized in Table 3. The results showed that the models' predictions achieved a high match between the observed and output sets indicated by R of 0.91 and 0.95 for MW_{BO} and MW_{BD} models, respectively. In terms of prediction errors, the developed models resulted in a low MAPE not exceeding 0.53% for both MW_{BO} and MW_{BD} models. The high accuracy of the developed models can be also inferred from the noticeable agreement in the graphical representation of both the actual and predicted values for the testing process as shown in Fig. 6 for MW_{BO} and MW_{BD} models.

Moreover, the close scatter of the observed and predicted output data for both models demonstrated a remarkable match among these values as shown in the crossplots depicted in Fig. 7. This revealed the high accuracy of the developed models' predictions. Accordingly, the appropriate mud window for safe drilling can be defined using the MW_{BO} and MW_{BD} predictions using the developed models.

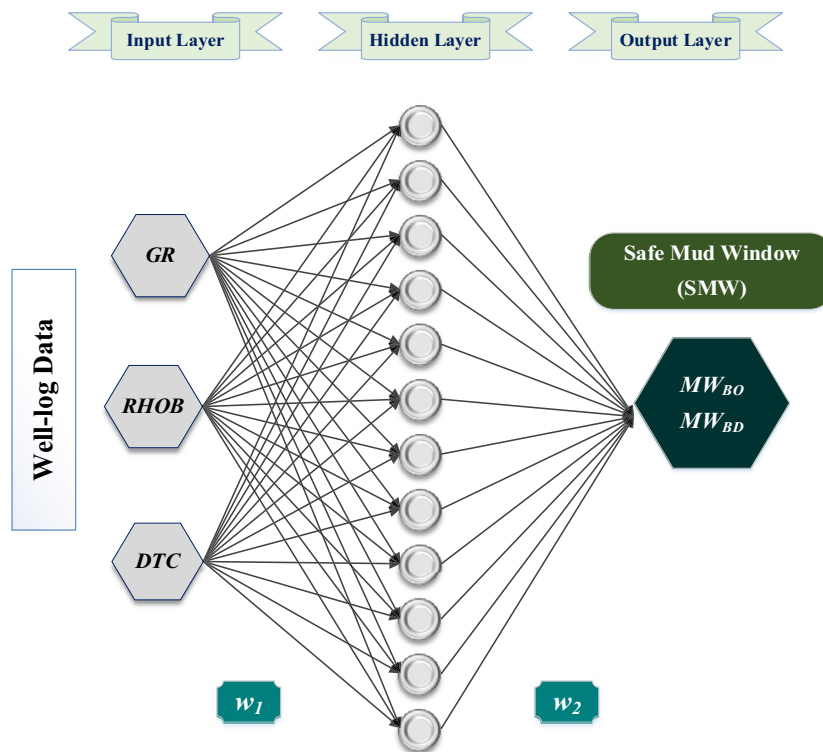


Figure 4. Typical schematic of the developed ANN architectures.

New equations development. One of the main outcomes of this study is to develop new equations whereby the mud window for safe drilling can be determined conveniently. Therefore, new Eqs. (7) and (8) have been established using the same inputs (GR, RHOB, and DTC) to estimate both MW_{BO} and MW_{BD} . The established correlations are based on the tuned weights and biases of the developed models listed in Tables 4 and 5.

$$MW_{BO} = 2.29(MW_{BO})_{normalized} + 95.02 \quad (7)$$

$$MW_{BD} = 8.732(MW_{BD})_{normalized} + 158.887 \quad (8)$$

where, $(MW_{BO})_{normalized}$ and $(MW_{BD})_{normalized}$ are the normalized forms of the MW_{BO} and MW_{BD} values. The procedure required to use Eqs. (7) and (8) are described in detail in Appendix B.

Model validation. In this section, the developed ANN-based equations have been validated using an unseen dataset that has not been used in the learning process. A total of 637 data points has been fed to the developed Eqs. (7) and (8) to estimate the safe mud window in terms of MW_{BO} and MW_{BD} . The results have been then evaluated using R-value, MSE, RMSE and MAPE between the actual and estimated output sets. The estimated MW_{BO} and MW_{BD} showed a high agreement with the actual values as depicted in Fig. 8 with R exceeding 92% and MAPE of 0.60% at maximum. Table 6 summarizes the prediction accuracy of the developed equations in terms of MAPE, MSE and RMSE.

It should be highlighted that the developed correlations are more recommended for carbonate formations from which most of the data used in developing the models, were collected. This can be explained as different formation types may have different log responses and geomechanical properties. These properties control the downhole stress distributions and hence the breakout and breakdown limits. Therefore, some errors might be expected upon the application of different formation lithologies. Moreover, it is recommended to deploy the developed equations using inputs within the range and the same units listed in Table 2 to ensure reliable results.

Conclusions

An artificial neural network (ANN) has been successfully applied to define the safe mud window limits in terms of MW_{BO} and MW_{BD} . The developed models use the logging data namely GR, RHOB, and DTC as input features. The findings of this research can be summarized as follows:

- The developed ANN-based models resulted in a considerable match with the observed MW_{BO} and MW_{BD} values with accuracy exceeding 92% and a maximum mean absolute percentage error (MAPE) of 0.53%.

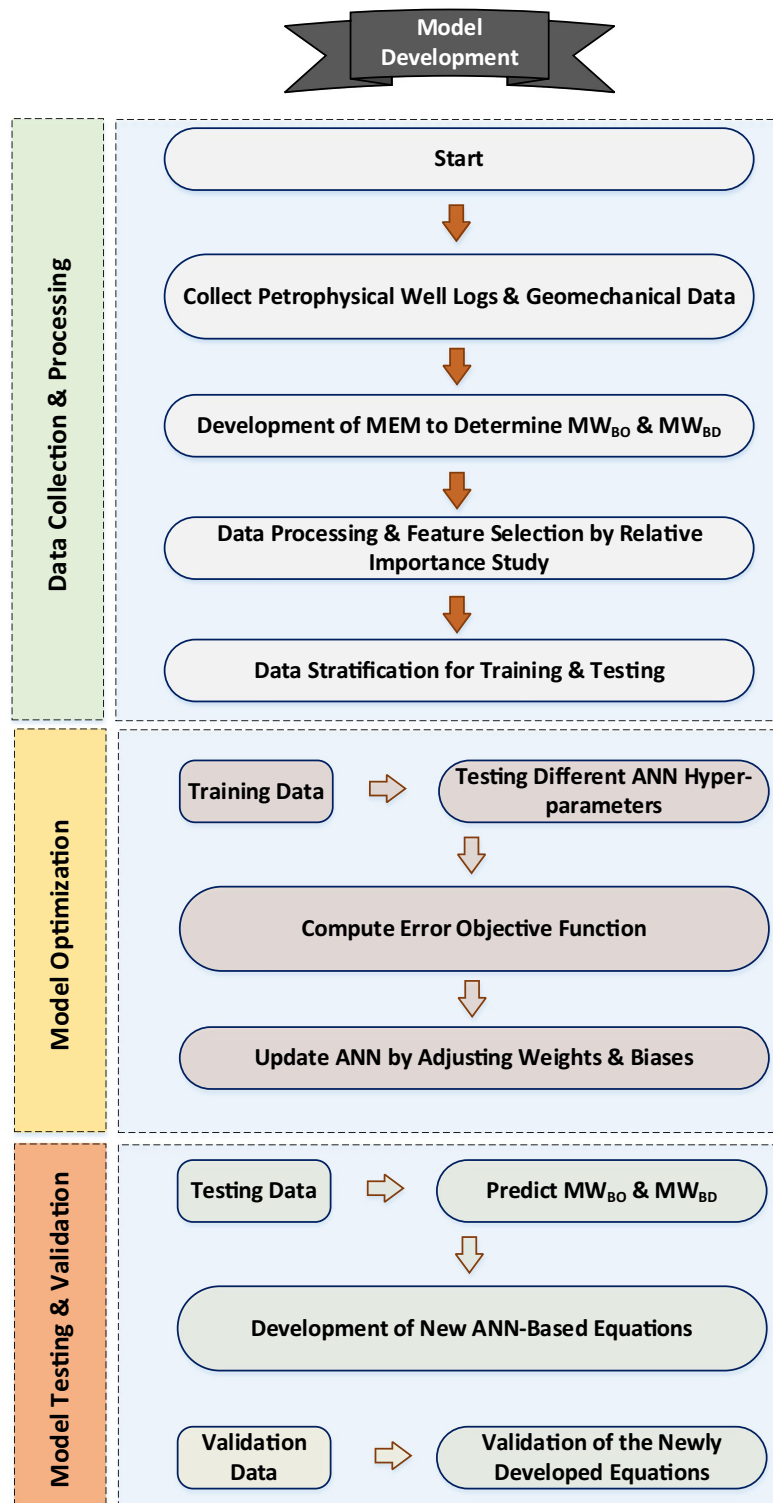


Figure 5. Flowchart for developing the proposed new ANN-Based equations.

- New equations were established using the developed models to estimate MW_{BO} and MW_{BD} from the logging data directly without the need of running special codes.
- The developed equations were validated using unseen data from the same field. The results demonstrated the robustness of the developed equations to estimate the MW_{BO} and MW_{BD} directly from the logging data with MAPE not exceeding 0.60%.

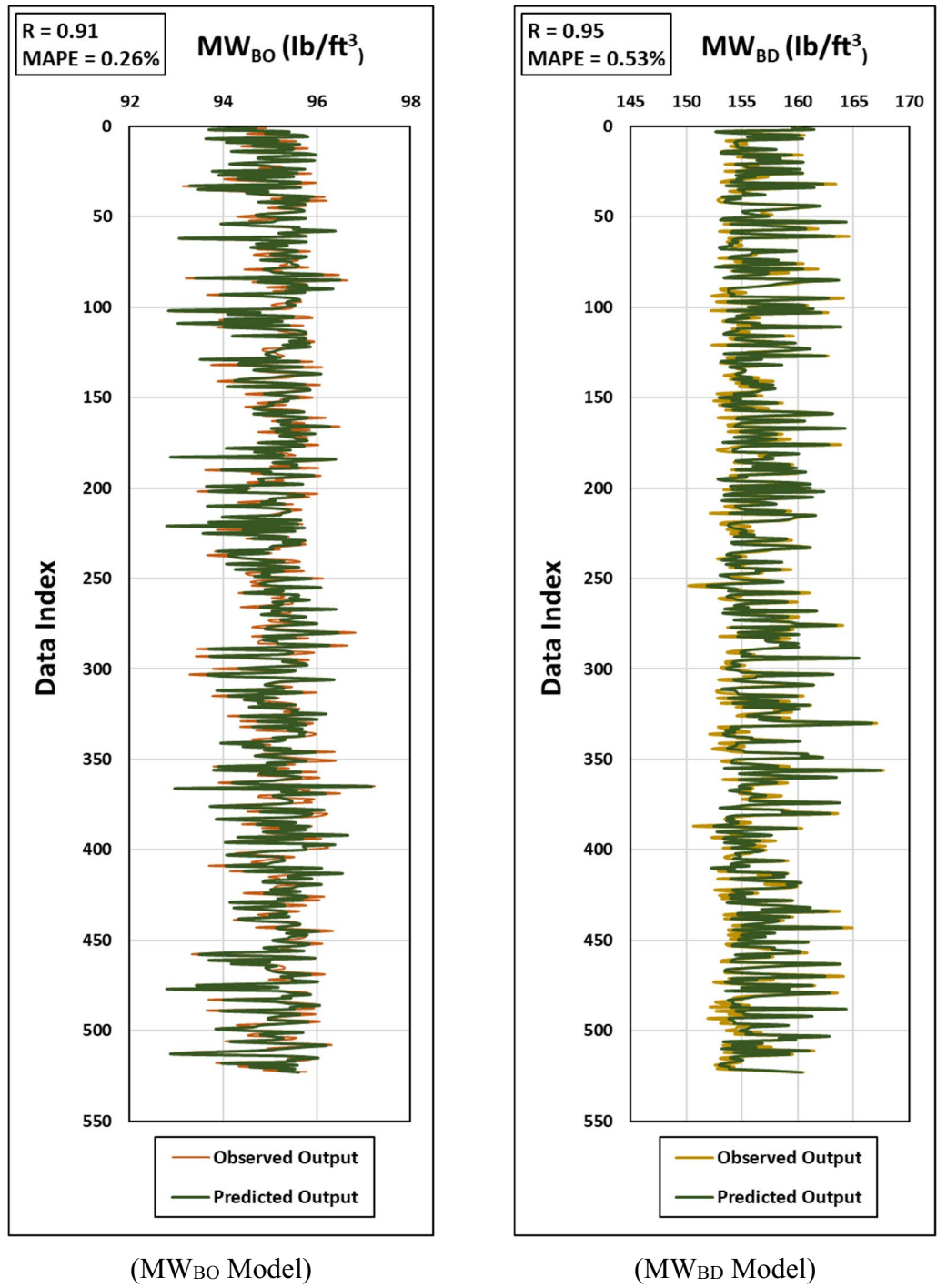


Figure 6. Graphical representations for the Observed vs. Predicted outputs for both MWBO (left) and MWBD (right) models for the testing process.

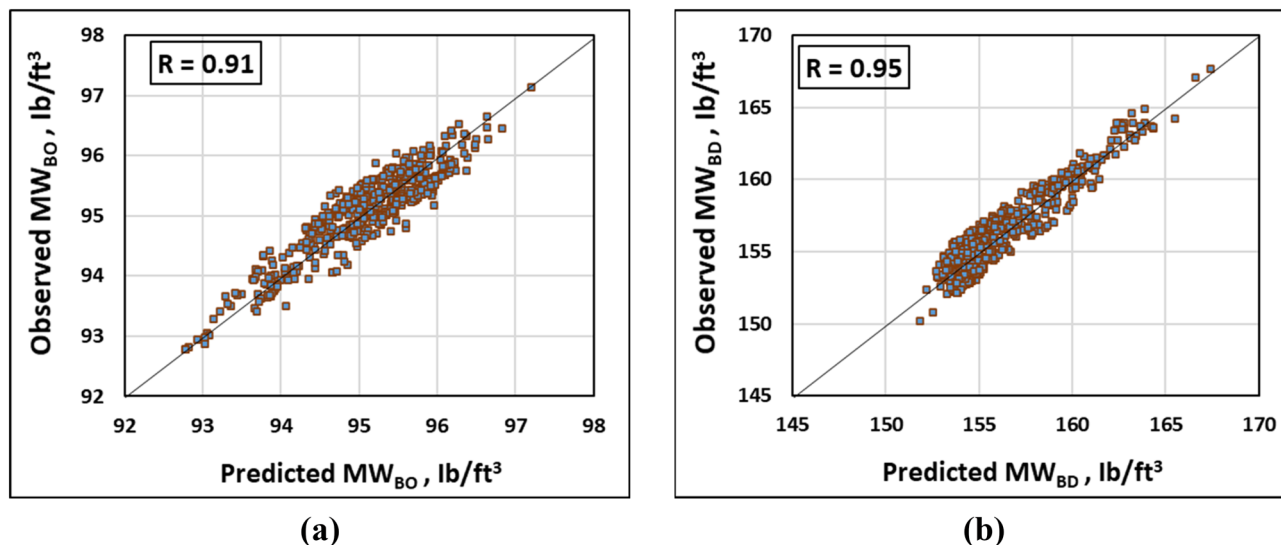


Figure 7. Crossplots between the actual and predicted output values for (a) MW_{BO} and (b) MW_{BD} models using the testing dataset.

i	W_{1ij}			W_{2i}	$b_{1,i}$	b_2
	$j=1$	$j=2$	$j=3$			
1	-0.146	3.033	-0.626	0.972	-1.907	-0.598
2	0.136	-0.915	-1.555	-2.913	-0.630	
3	3.116	-0.749	0.169	-1.230	-0.454	
4	0.587	0.957	2.175	-0.990	0.613	
5	-4.532	0.451	-0.565	-0.768	0.770	
6	-0.593	1.210	4.230	-0.779	2.736	
7	-1.410	0.579	2.169	-0.957	0.371	
8	-0.398	2.369	-0.515	-1.743	-1.829	
9	-1.864	-0.601	0.769	0.501	0.446	
10	1.022	-1.752	-0.466	1.067	-0.439	

Table 4. The optimized weights and biases of the developed ANN-based model to estimate the MWBO.

i	W_{1ij}			W_{2i}	$b_{1,i}$	b_2
	$j=1$	$j=2$	$j=3$			
1	-0.445	0.393	2.084	-0.598	0.546	1.398
2	-0.743	-0.797	-0.110	1.689	-0.369	
3	-1.822	-1.811	-1.066	-0.705	-0.537	
4	-0.704	1.211	-1.903	1.188	-1.668	
5	0.553	-2.369	1.021	1.069	1.565	
6	-0.266	-1.558	-0.526	-1.628	1.186	
7	-0.123	0.160	-1.913	-1.418	-0.981	

Table 5. The optimized weights and biases of the developed ANN-based model to estimate the MWBD.

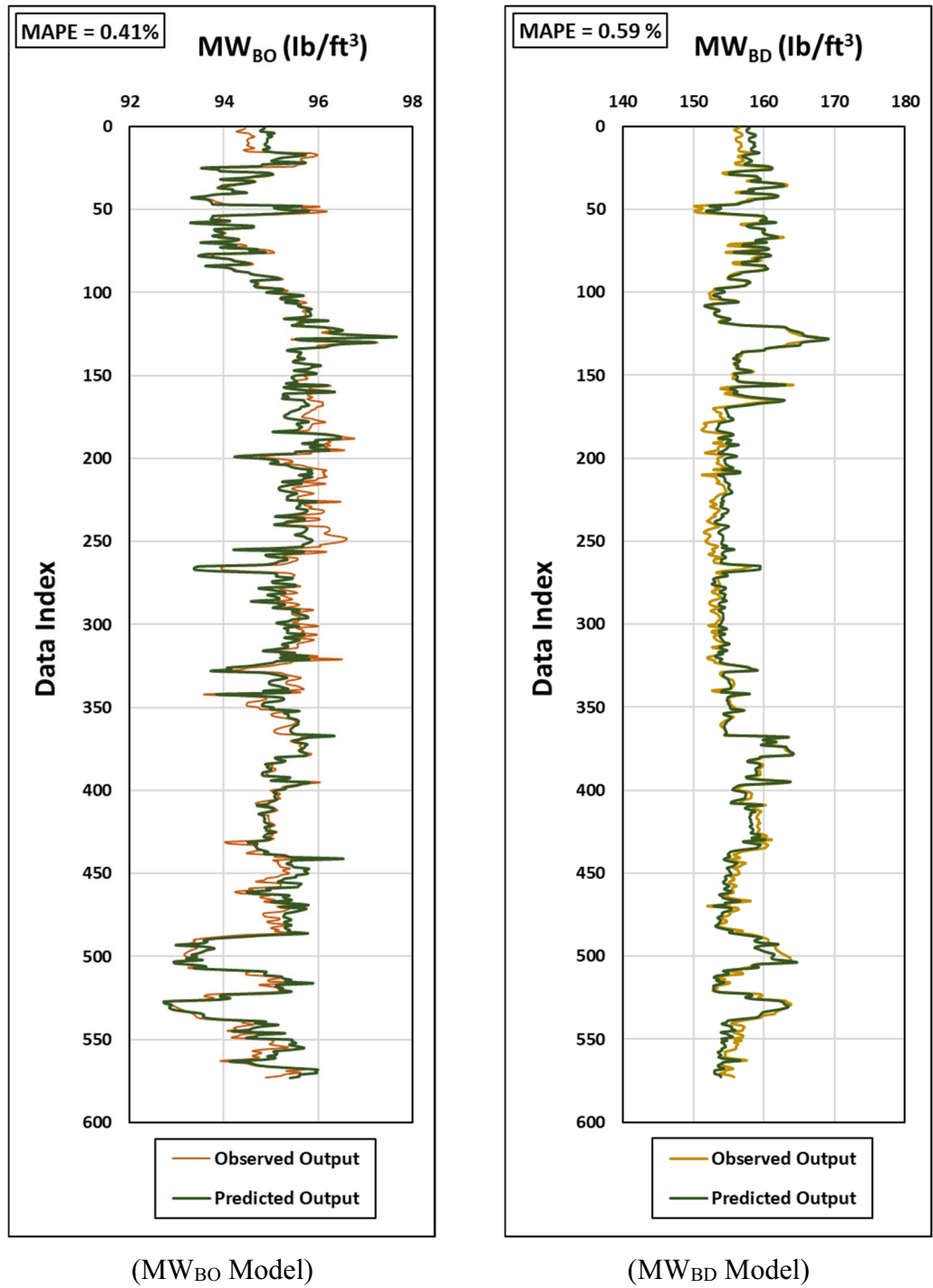


Figure 8. Graphical representations for the Observed vs. predicted outputs for both MW_{BO} (left) and MW_{BD} (right) models for the validation process.

	MW _{BO}			MW _{BD}		
	Prediction error			Prediction error		
	MAPE	MSE	RMSE	MAPE	MSE	RMSE
Training process	0.21	0.003	0.057	0.41	0.004	0.065
Testing process	0.26	0.035	0.186	0.53	0.357	0.598
Validation process	0.41	0.044	0.209	0.59	0.675	0.821

Table 6. Prediction accuracy of the developed equations to estimate MW_{BO} and MW_{BD}.

Received: 23 May 2022; Accepted: 9 September 2022

Published online: 21 September 2022

References

- Zoback, M. D. *et al.* Determination of stress orientation and magnitude in deep wells. *Int. J. Rock Mech. Min. Sci.* **40**, 1049–1076 (2003).
- Zoback, M. D. *Reservoir Geomechanics* (Cambridge University Press, 2007).
- Zahiri, J., Abdideh, M. & Ghaleh Golab, E. Determination of safe mud weight window based on well logging data using artificial intelligence. *Geosyst. Eng.* **22**, 193–205 (2019).
- Aslannezhad, M. & Jalalifar, H. Determination of a safe mud window and analysis of wellbore stability to minimize drilling challenges and non-productive time. *J. Pet. Explor. Prod. Technol.* **6**, 493–503 (2016).
- Kamran, M. A probabilistic approach for prediction of drilling rate index using ensemble learning technique. *J. Min. Environ.* **12**, 327–337 (2021).
- Abbas, A. K., Almohammed, H. H., Alqatran, G., Mohammed, H. Q. & Mohammed, A. Determination of safe operating mud weight window from well logging data using machine learning algorithms. in *Offshore Technology Conference Asia* (Offshore Technology Conference, 2020).
- Phan, D. T., Liu, C., AlTammar, M. J., Han, Y. & Abousleiman, Y. N. Application of artificial intelligence to predict time-dependent mud-weight windows in real time. *SPE J.* **27**, 39–59 (2022).
- Ibrahim, A. F., Gowida, A., Ali, A. & Elkhatny, S. Machine learning application to predict in-situ stresses from logging data. *Sci. Rep.* **11**, 1–14 (2021).
- Gowida, A., Ibrahim, A. F., Elkhatny, S. & Ali, A. Prediction of the least principal stresses using drilling data: A machine learning application. *Comput. Intell. Neurosci.* **2021**, 1–13 (2021).
- Ibrahim, A. F., Gowida, A., Ali, A. & Elkhatny, S. Real-time prediction of in-situ stresses while drilling using surface drilling parameters from gas reservoir. *J. Nat. Gas Sci. Eng.* **97**, 104368 (2022).
- Fjaer, E., Holt, R. M., Horsrud, P. & Raaen, A. M. *Petroleum Related Rock Mechanics* (Elsevier, 2008).
- Dawson, R. How significant is a boxplot outlier?. *J. Stat. Educ.* <https://doi.org/10.1080/10691898.2011.11889610> (2011).
- Rabbani, E., Sharif, F., Koolivand Salooki, M. & Moradzadeh, A. Application of neural network technique for prediction of uniaxial compressive strength using reservoir formation properties. *Int. J. Rock Mech. Min. Sci.* **56**, 100–111 (2012).
- Gowida, A., Elkhatny, S. & Gamal, H. Unconfined compressive strength (UCS) prediction in real-time while drilling using artificial intelligence tools. *Neural Comput. Appl.* **33**, 8043–8054 (2021).
- Chemmakh, A. *Machine Learning Predictive Models to Estimate the UCS and Tensile Strength of Rocks in Bakken Field* (OnePetro, 2021).
- Jamshidi, E., Arabjamaloei, R., Hashemi, A., Ekramzadeh, M. A. & Amani, M. Real-time estimation of elastic properties of formation rocks based on drilling data by using an artificial neural network. *Energy Sources Part A* **35**, 337–351 (2013).
- Gowida, A., Moussa, T., Elkhatny, S. & Ali, A. A hybrid artificial intelligence model to predict the elastic behavior of sandstone rocks. *Sustainability* **11**, 5283 (2019).
- Acar, M. C. & Kaya, B. Models to estimate the elastic modulus of weak rocks based on least square support vector machine. *Arab. J. Geosci.* **13**, 1–12 (2020).
- Tabaeh, H. M. & Mohammad, A. Estimation of in-situ horizontal stresses using the linear poroelastic model and minifract test results in tectonically active area. *Russ. J. Earth Sci.* **16**, 1–9 (2016).
- Yegnanarayana, B. *Artificial Neural Networks* (PHI Learning Pvt. Ltd, 2009).
- Xu, Y. & Goodacre, R. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* **2**, 249–262 (2018).
- Pereira, L. G., Gandelman, R. A., Clemente, R. G., Teixeira, P. H. S. & Teixeira, G. T. Development of software to predict mud weight for pre-salt drilling zones using machine learning. in *OTC Brasil* (OnePetro, 2013).
- Zhou, H., Niu, X., Fan, H. & Wang, G. Effective calculation model of drilling fluids density and ESD for HTHP well while drilling. in *IADC/SPE Asia Pacific Drilling Technology Conference* (OnePetro, 2016).
- Okpo, E. E., Dosunmu, A. & Odagme, B. S. Artificial neural network model for predicting wellbore instability. in *SPE Nigeria Annual International Conference and Exhibition* (OnePetro, 2016).

Author contributions

S.E. conceived the idea and collected the required data and participated in the methodology design. A.I., A.G., conducted the data analysis, designed the methodology, run the algorithms and performed the sensitivity and the optimization of the results. S.E., also participated in methodology design and results validation supervising. The original manuscript was written by A, I., and all authors participated in the manuscript revision and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-20195-7>.

Correspondence and requests for materials should be addressed to S.E.

Reprints and permissions information is available at www.nature.com/reprints.

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



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/>.

© The Author(s) 2022