ABSTRACT: Real-time prediction of the formation pressure gradient is critical mainly for drilling operations. It can enhance the quality of decisions taken and the economics of drilling operations. The pressure while drilling tool can be used to provide pressure data while drilling, but the tool cost and its availability limit its usage in many wells. The available models in the literature for pressure gradient prediction are based on well logging or a combination of some drilling parameters and well logging. The well-logging data are not available for all wells in all sections in most wells. The objective of this paper is to use support vector machines, functional networks, and random forest (RF) to develop three models for real-time pore pressure gradient prediction using both mechanical and hydraulic drilling parameters. The used parameters are mud flow rate ($Q$), standpipe pressure, rate of penetration, and rotary speed (RS). A data set of 3239 field data points was used to develop the predictive models. A different data set unseen by the model was utilized for the validation of the proposed models. The three models predicted the pore pressure gradient with a correlation coefficient ($R$) of 0.99 and 0.97 for training and testing, respectively. The root-mean-squared error (RMSE) ranged from 0.008 to 0.021 psi/ft for training and testing, respectively, between the predicted and the actual pore pressure data. Moreover, the average absolute percentage error (AAPE) ranged from 0.97% to 3.07% for training and testing, respectively. The RF model outperformed the other models by an $R$ of 0.99 and RMSE of 0.01. The developed models were validated using another data set. The models predicted the pore pressure gradient for the validation data set with high accuracy ($R$ of 0.99, RMSE around 0.01, and AAPE around 1.8%). This work shows the reliability of the developed models to predict the pressure gradient from both mechanical and hydraulic drilling parameters while drilling.

1. INTRODUCTION

Petroleum engineering is an engineering field concerned with the activities related to hydrocarbon production, which can be either crude oil or natural gas. As a part of petroleum engineering, drilling is considered the only way for the actual discovery of reservoirs to produce hydrocarbons. While drilling a well through a certain geological column, various formations with different properties and pressures (pore and fracture pressures) are encountered. The knowledge of the drilling window, which contains both pore and fracture pressures versus depth, may secure reaching the target with minimum time and costs.

The pore pressure or formation pressure is the pressure exerted by the fluids contained in the porous media. The normal pore pressure at any depth results from the weight of the water column extended from a certain depth to the surface. The normal pressure gradient ranges from 0.433 psi/ft for fresh water to 0.465 psi/ft for salt water. The abnormal pressure term can be used to describe the deviation from the normal gradient, which can be either overpressure or subnormal pressure. Overpressure, also called geopressed, is the pore pressure greater than the normal pressure. This pressure results from an extra pressure source added to the normal pressure and may cause kicks and blowouts. The excess pressure source may be due to geological, mechanical, geochemical, geothermal, and combined reasons. However, the subnormal pressure is the pressure less than the normal pressure and may cause differential-pressure pipe sticking and loss of circulation.

Real-time pore pressure prediction may enhance well trajectory, casing, and mud program designs and provide better wellbore stability analysis, resulting in reducing the overall drilling time and cost.

Real-time pore pressure prediction can be either qualitative or quantitative. The proposed methods in the literature used well-logging data, formation properties and few studies combined logging and drilling parameter data. Hottman and Johnson predicted the pore pressure based on logging data from Miocene and Oligocene shales. The authors established Cartesian cross-
plots relating the pressure gradient and the difference in sonic travel time or resistivity ratio between the observed and the normal trends. Pennebaker\textsuperscript{1} used the sonic travel time ratio instead of the difference in sonic travel time used by Hottman and Johnson.\textsuperscript{2,3} Matthews and Kelly\textsuperscript{4} utilized a semilog scale instead of the Cartesian one for the same correlation of Hottman and Johnson. Eaton\textsuperscript{5} mentioned that both pore and overburden pressures control log-derived data. Eaton\textsuperscript{6} developed an empirical sonic-based model to estimate the pore pressure gradient in shales. Gardner et al.\textsuperscript{7} introduced a new attempt to predict the pore pressure by including the overburden pressure data after he had analyzed the data provided by Hottman and Johnson. Bowers\textsuperscript{8} predicted the pore pressure from sonic data (slowness) after replacing the effective stress \((\alpha_c)\) with \((\alpha_c - \text{pore pressure})\).

Foster and Whalen\textsuperscript{9} used the equivalent depth method for the first time to predict the pore pressure from electrical logs. Additionally, Ham\textsuperscript{10} used the equivalent depth method with resistivity, sonic, and density data to estimate the pore pressure and calculate the required mud density. Eaton\textsuperscript{11,12} proposed empirical resistivity- and conductivity-based models to predict the formation pressure gradient in shales from well logs. Bingham\textsuperscript{13} introduced the \(d\) exponent to correct the rate of penetration (ROP) for the changes in the hole diameter, weight on bit (WOB), and rotary speed (RS). Jorden and Shirley\textsuperscript{14} modified the Bingham approach by suggesting a new term called \(d_{exp}\). Rehm and McClendon\textsuperscript{15} modified Jorden and Shirley’s \(d_{exp}\) by considering the mud weight effect. Eaton\textsuperscript{16} noticed that the corrected \(d\) exponent plot is similar to the resistivity log plot. Consequently, the author introduced a model to estimate the pore pressure at any depth using observed \(d\), normal trendline value of \(d\), overburden pressure, and normal pore pressure gradients.

2. MACHINE LEARNING STUDIES IN PETROLEUM ENGINEERING

Machine learning has various tools such as artificial neural networks (ANNs), support vector machines (SVMs), adaptive neurofuzzy inference system (ANFIS), and functional networks (FNs), which provide robust performance and high accuracy for classification and prediction problems.\textsuperscript{17} Machine learning is widely used in many disciplines of engineering, medicine, economics, military, and marine sectors.\textsuperscript{18} It has been widely used in petroleum engineering as it not only has the ability to solve complex problems and deal with the big data but also perfectly represents them with high accuracy compared to other models.\textsuperscript{19} Different models were introduced for different purposes such as ROP prediction and optimization for different drilled formations and well profiles,\textsuperscript{20} estimating the oil recovery factor,\textsuperscript{21} lithology classification,\textsuperscript{22} well planning,\textsuperscript{23} the formation lithology,\textsuperscript{24} prediction of formation tops,\textsuperscript{25} estimating the properties of reservoir fluids,\textsuperscript{26} fracture density estimation,\textsuperscript{27} detecting the downhole abnormalities during horizontal drilling,\textsuperscript{28} wellbore stability,\textsuperscript{29} predicting the compressional and shear sonic times,\textsuperscript{30} fracture pressure prediction while drilling,\textsuperscript{31} estimating the content of total organic carbon,\textsuperscript{32} identifying and estimating the rock failure parameters,\textsuperscript{33} estimating the wear of a drill bit from the drilling parameters,\textsuperscript{34} predicting the rheological properties of drilling fluids in real time,\textsuperscript{35−37} and estimating the rock static Young’s modulus.\textsuperscript{38−41}

A few studies used machine learning tools to predict the pore pressure gradient. Ahmed et al.\textsuperscript{42} applied ANN to develop a pore pressure white box prediction model using seven parameters. The used parameters were porosity \((\rho)\), rotation speed (rpm), WOB, mud weight (MW), bulk density, rate of penetration (ROP), and interval transient time (\(\Delta t\)). Ahmed et al.\textsuperscript{43} used five artificial intelligence tools (ANN, ANFIS, RBF, SVM, and FN) to estimate the pore pressure with the same seven input variables used in Abdulfalek et al.\textsuperscript{44} study. Aliouane et al.\textsuperscript{45} proposed fuzzy logic (FL)- and ANN-based models to predict the pore pressure from logging data in shale gas reservoirs. Hu et al.\textsuperscript{46} applied the ANN to predict the formation pressure. The authors used four input parameters, which are depth, density, interval transit time, and gamma ray. Li et al.\textsuperscript{47} used the ANN to predict the pore pressure by including input parameters, such as gamma ray, interval transit time, natural potential, and pipe pressure test data.

Rashidi and Asadi\textsuperscript{48} introduced the ANN-based model to predict the pore pressure using two parameters (mechanical specific energy and drilling efficiency). The proposed model does not consider bit hydraulics and bit wear, which may cause wrong predictions in very soft and very hard abrasive formations. Keshavari and Jahanbakhshi\textsuperscript{49} applied the back-propagation neural network (BPNN) and the general regression neural network (GRNN) to predict the pressure gradient in the Asmari reservoir in Iran. The input parameters were depth, permeability, rock density, and porosity.

The objective of this work is to use SVMs, FNs, and RF to develop three models to predict the pore pressure gradient while drilling using the available mechanical and hydraulic drilling parameters. Unlike the other empirical models, the proposed models do not require a pressure trend (such as normal pressure trend) to predict the pore pressure gradient. The high cost and low availability of the pressure while drilling (PWD) tool limit its usage in many wells. The available models in the literature are based on well logging or a combination of some drilling parameters and well logging. The well-logging data are not available for all wells in all sections in most wells. Moreover, the logging while drilling (LWD) tool is located tens of feet above the drilling bit in case it is there which actually does not reflect the formations being drilled instantaneously.

3. METHODOLOGY

The study started with data acquisition from vertical wells in the Middle East followed by the filtration and cleaning process. The data set went through the analysis stage to obtain more information about the inputs and the target. Then, random division of the data set took place for training and testing. The model development stage started with running the initial case, and the parameters were updated until obtaining the best results.

Finally, the optimum parameters were extracted, and the models were validated using unseen data, which were not included in training and testing. Figure 1 summarizes the methodology followed in this study to build the different models. In this work, three machine learning models were developed to predict the pore pressure gradient of the downhole formations while drilling using three machine learning tools: SVM, FNs, and RF.

4. DATA PROCESSING AND ANALYSIS

4.1. Data Description. A data set of around 3239 points was collected from some vertical wells in a field in the Middle East. The data included hydraulic and mechanical drilling parameters in addition to the pore pressure gradient data with their corresponding depths. These drilling parameters include hydraulic measurements such as pump rate and standpipe
pressure (SPP) and mechanical measurements such as RS, ROP, torque (T), and WOB. The drilling parameters were used as model inputs to predict the pore pressure gradient as a target. These drilling parameters can be measured at either the surface or the downhole during normal drilling operations. Additionally, these parameters are affected by formations being drilled and their fluid content. The field data were statistically analyzed showing data variability, as the data cover a wide range of the inputs and the outputs, as shown in Table 1. For example, the data have a wide range of the pore pressure gradient as it covers normal (around 0.465 psi/ft), abnormal (greater than 0.465 psi/ft), and subnormal (lower than 0.465 psi/ft) pressure.

4.2. Data Processing. In machine learning, the data quality is as important as the quality of a prediction or a classification model. The field data should be filtered and analyzed for better prediction. A specially designed MATLAB program was used to remove all values that are not representative, such as missing values, −999 values, not a number (NAN), and any unrealistic data points. Consequently, the data were cleaned by removing all values that are not representative, such as negative values, −999 values, NAN, and any unrealistic data points. Then, the outliers should be excluded as they may result in major issues in statistical analysis.

Outliers may be as a result of human errors and/or instrumental errors. Outlier detection can be performed by some techniques such as the Z score and a box-and-whisker plot. The reliability of the input data was tested by different strategies such as comparing the collected data with the working ranges of the tools and with the same parameters in offset wells in the same field. Additionally, the actual pore pressure gradient values were compared to the pore pressure values generated using the common trends of the pressure gradient for different formations in this field. This comparison ensured a reasonable match between the actual and the generated values, confirming the reliability of the collected data.

4.3. Input Selection. The formation properties affect the drillability of the different geological strata during drilling operations, as these properties dictate the resistance to drill through these formations. The drilling parameters can somehow reflect the resistance encountered during drilling the geological column. The generated cuttings have effects on the pump pressures and their rates during drilling. The formation types and the drilling parameters play a significant role in controlling the ROP. As a result, the aforementioned drilling parameters can somehow reflect the nature of the drilled formations, and in turn, their pore pressure gradients. The ROP can be utilized as an indicator to detect overpressurized formations during drilling. The increased porosity (trapped fluids) and the reduced density of the under-compacted formations make them more drillable. The ROP was used to build the models as it reflects the effects of other drilling parameters such as WOB. Additionally, RS was used as it indirectly includes the torque effect. Two mechanical drilling parameters (ROP and RS) are used in addition to two hydraulic drilling parameters (SPP and pump rate).

5. RESULTS AND DISCUSSION OF THE DEVELOPED PREDICTION MODELS

5.1. SVM Model. SVM is a supervised learning tool used to analyze data for regression and classification problems with a high degree of complexity. Moreover, SVM has advantages over other artificial intelligence algorithms, such as generalization capability, strong interference capacity, and less learning time. SVM transfers the data from a lower-dimensional to a higher-dimensional space, called kernel space, which provides more space for training examples to find a support vector classifier (hyperplane), which reduces the number of misclassified points. SVM uses kernel functions to move the data to the kernel space by systematically finding the support vector classifiers in the higher dimension. The kernel function selection is based on the nature of the data. The performance of the SVM-based model relies on the optimization process of many parameters to develop the desired predictive model with a high accuracy.

The filtered data were split into two groups, with a ratio of 75:25 for training and testing the model, respectively. The SVM model parameters, including various combinations of different options available for SVM parameters, were optimized by running different cases for each parameter. The standard practice to optimize the hyperparameters is to use grid search or random search with cross-validation. These methods search through a fixed (grid search) or random (random search) set of values for the hyperparameters and select the one that provides the best modeling performance after being evaluated by cross-validation and testing the model parameters.
validation. In this study, the hyperparameters were optimized one by one to select the optimum. Moreover, many metrics were calculated for every run when we changed one hyperparameter at a time to look at all of them and then compare to decide about the optimum one.

Different kernel functions, like Gaussian and polynomial functions, with different tuning parameters, such as regularization parameter (C) (from 1 to 3000), kernel option (from 1 to 40), epsilon, and lambda, were tested to obtain the best performance. The regularization or penalty parameter characterizes the generalization ability of the machine that controls the sensitivity of the machines to outliers or, in other words, it tells the algorithm how much it should care about the misclassified points. The kernel option is a scalar or a vector containing the options for the kernel function selected. In the case of the polynomial kernel, the kernel option is a scalar that gives the degree of the polynomial or a vector in which the first element is the degree of the polynomial and other elements give the bandwidth of each dimension; thus, the vector is of size $n + 1$, where $n$ is the dimension of the problem. For the Gaussian kernel, the kernel option defines the bandwidth of each dimension.

The model performance was evaluated using R and AAPE between the predicted and actual target values. The correlation coefficient and AAPE were calculated by eqs 1 and 2, as shown in Appendix 1. The parameters giving the highest R and the lowest errors (RMSE and AAPE) between the actual and the predicted pore pressure gradient values were chosen. The SVM-based model, with its optimized parameters, as listed in Table 2, predicted the pore pressure gradient with an R value of 0.98 and an AAPE of 1.53% for training and an R value of 0.97 and an AAPE of 1.98% for testing. Moreover, the RMSE was 0.016 and 0.018 for training and testing, respectively.

5.2. FN Model. FNs are considered a powerful tool that have high capability like artificial neural networks (ANNs) for prediction and classification engineering problems.57,58 FNs are the generalization of ANNs in which the activation functions associated with neurons are learnt from data (not fixed). In ANNs, the weights of the neurons should be learnt, while they are suppressed in FNs.52 Unlike ANNs, FNs do not require weights on the neuron connections as they use multiargument functional models and the weight effect inherently exists within these functions.59,60 The outputs of the neurons are forced to converge to an equivalent output. The specification of the initial topology in the FN is based on the features of the problem. As a result, understanding the problem can assist in developing the structure of the network.

The same set of data was used to develop the FN predictive model. The data were randomly divided into a ratio of 75/25 for training and testing, respectively. The inputs were pump rate, ROP, RS, and SPP. Five FN methods with four types per each were examined: exhaustive search (FNESM), forward selection (FNFSM), backward elimination (FNBEM), forward—backward (FNFBM), and backward—forward (FNBFM). The performance of each method was compared to that of others based on the R and AAPE to select the optimum method. Based on the optimization process, as listed in Table 3, FNFBM and FNBFM with type 3 (nonlinear) resulted in the highest R of 0.97 for training and 0.96 for testing between the actual and predicted output values. Moreover, the RMSE was around 0.019 and 0.021 for training and testing, respectively, and the AAPE was around 2.8% for training and 3.07 for testing. Figure 3 shows the cross-plots between the predicted and the actual pore pressure gradients, in which the points coincide with the 45° line.

5.3. RF Model. RF is considered an ensemble learning technique that can be utilized for regression and classification problems.61 It combines hundreds or thousands of decision trees to train each tree on a slightly different set of observations. It uses a process, called bootstrapping, as an iterative resampling technique to estimate statistics of a population by sampling a data set with replacement. The predictions of each tree are averaged to provide the final prediction in a process called aggregation. About 33% of the original data sets are not included in bootstrapping, and this sample, called out-of-bag data, is used to internally check the RF-based model accuracy.62 The RF is better than a single decision tree due to its ability of limiting the overfitting without an increase in the error margin.63

### Table 2. SVM-Based Model Optimized Parameters

| Parameter       | Value |
|-----------------|-------|
| Kernel function | Gaussian |
| C               | 10 |
| Kernel option   | 30 |

![Figure 2](https://example.com/f2.png)  
Figure 2. Cross-plots between the predicted and actual pressure gradient results: (a) training and (b) testing (SVM model).
After the optimization process in which the tuning parameters are adjusted, the optimum parameters are listed in Table 4.

The Max_features option determines the number of features to be considered when searching for the best split (e.g., if it is "sqrt," then max_features = sqrt (n_features) and if it is "log2," then max_features = log2(n_features)). The option n_estimators determines the number of trees in the forest, while max_depth determines the maximum depth of the tree. The RF model predicted the pore pressure gradient using the same input parameters with an $R^2$ value of 0.99 for training and 0.98 for testing, with an AAPE not exceeding 1.79%. Moreover, the RMSE did not exceed 0.011 psi/ft for both training and testing. Figure 4 shows the cross-plots between the predicted and actual pore pressure gradients, in which the points significantly coincide with the 45° line.

The optimum results of the three models were compared, as listed in Table 5, which shows that the RF model is the most appropriate technique in predicting the pore pressure gradient, with the highest $R^2$ value and the lowest error for training and testing. The SVM model came in the second place, followed by the FN model, as the least accurate model. The histogram of errors for the RF model (the least errors) was constructed, as shown in Figure 5. Moreover, the RMSE was calculated for the testing data set for the RF model after dividing it into three categories (subnormal, normal, and supernormal). The RMSE was 0.016, 0.02, and 0.015 for subnormal, normal, and supernormal, respectively. Based on the results we got from the analysis performed in each category, it is found that the RMSE as an evaluation metric is more important than $R^2$ to check the models’ performance in our problem. To prove this, we took a small sample from the supernormal pressure data set and calculated the RMSE and $R^2$ for this sample. Then, the results were compared to the overall RMSE and $R^2$ of the full supernormal pressure data set. The sample had an RMSE of 0.015 and an $R^2$ value about 0.94. Based on the results, both gave small RMSE and different $R^2$ values.

### Table 3. FN Model Performance for Various Methods

| FN method | relationship type | $R^2_{\text{training}}$ | $R^2_{\text{testing}}$ | AAPE_{\text{training}} | AAPE_{\text{testing}} |
|-----------|------------------|-----------------------|-----------------------|------------------------|-----------------------|
| FNEM      | 1                | 0.9430                | 0.9394                | 4.3283                 | 4.4767                |
| FNEM      | 2                | 0.9622                | 0.9562                | 3.1991                 | 3.4261                |
| FNFSM     | 1                | 0.9430                | 0.9394                | 4.3283                 | 4.4767                |
| FNFSM     | 2                | 0.9622                | 0.9562                | 3.1991                 | 3.4261                |
| FNFSM     | 3                | 0.9673                | 0.9574                | 3.0901                 | 3.3141                |
| FNFSM     | 4                | 0.9631                | 0.9574                | 3.0436                 | 3.3243                |
| FNFSM     | 1                | 0.9430                | 0.9394                | 4.3283                 | 4.4767                |
| FNFSM     | 2                | 0.9603                | 0.9537                | 3.2099                 | 3.4565                |
| FNBSM     | 3                | 0.9649                | 0.9588                | 3.0436                 | 3.3243                |
| FNBSM     | 4                | 0.9599                | 0.9539                | 3.2387                 | 3.5710                |
| FNFBM     | 1                | 0.9430                | 0.9394                | 4.3283                 | 4.4767                |
| FNFBM     | 2                | 0.9622                | 0.9562                | 3.1939                 | 3.4201                |
| FNFBM     | 3                | 0.9696                | 0.9631                | 2.8032                 | 3.0798                |
| FNFBM     | 1                | 0.9430                | 0.9394                | 4.3283                 | 4.4767                |
| FNFBM     | 2                | 0.9622                | 0.9562                | 3.1939                 | 3.4201                |
| FNFBM     | 3                | 0.9696                | 0.9631                | 2.8032                 | 3.0798                |

The bold numbers represent the optimum runs.

### Table 4. RF-Based Model Optimized Parameters

| parameter      | available options     | optimum option |
|----------------|-----------------------|----------------|
| max_features   | ["auto", "sqrt", "log2"] | sqrt           |
| max_depth      | [3, 4, 5, ..., 30]    | 11             |
| n_estimators   | [3, 4, 5, ..., 150]   | 100            |

Figure 3. Cross-plots between the predicted and actual pressure gradient results: (a) training and (b) testing (FN model).
5.4. Model Application and Validation. The performance of the proposed three models was validated for predicting the pressure gradient using a total of 92 unseen data points from the same field that were not included in building the models. Table 6 shows the statistical analysis for the selected data set, which seems very similar to the training and testing data set analysis. A comparison between the actual versus the predicted pressure gradients from the models was established. The models predicted a continuous profile of the pressure gradients based on the continuous profiles of the available drilling parameters. The RF and SVM models estimated the pore pressure gradient with high R values of about 0.99 and 0.98 for the FN model between the actual and predicted output values for validation. Moreover, the AAPE did not exceed 1.2% for RF, 1.6% for SVM, and 2.6% for FN. The RF model gave the least RMSE of 0.01 psi/ft compared to SVM and FN models. Figure 6 shows the cross-
Figure 6. Cross-plots between the predicted and the actual pressure gradient results for 92 unseen data points: (a) RF model, (b) SVM model, and (c) FN model.

Figure 7. Pore pressure gradient profile for 92 unseen data points: (a) RF model, (b) SVM model, and (c) FN model.
plots between the predicted and the actual pressure gradient results for the validation data set. Moreover, the measured and predicted pore pressure gradient values were plotted on the same graph to track the differences along the selected sections, as shown in Figure 7, showing a high accuracy of pore pressure gradient prediction.

6. CONCLUSION

In this study, a new approach for pore pressure gradient prediction while drilling from the available hydraulic and mechanical drilling parameters using different techniques of AI was proposed. The developed models used around 3239 field data points from some wells to build the models. The developed models do not require a pressure trend (such as normal pressure trend) to estimate the pore pressure. The proposed models can be included in any automated drilling system to predict the pressure gradient in real time at a reasonable cost. Additionally, it can economically replace the PWD tool and can reduce the nonproductive time by minimizing some drilling problems by predicting and avoiding them before their occurrence. These models can technically and economically enhance the drilling operations while drilling and in predrilling design by taking the right actions and avoiding possible problems such as kicks, blowouts, and loss of circulation. The outcomes of this study can be summarized as follows:

- The optimum SVM model used the Gaussian kernel function with a C value of 10 and kernel option of 30.
- The optimum FN model used either FNFBM or FNBFM with type 3 (nonlinear) to obtain the best results.
- The optimum parameters of the RF model were sqrt as max_features, max_depth of 11, and n_estimators of 100.
- The three models predicted the pore pressure gradient with a correlation coefficient (R) between 0.99 and 0.97 for training and testing.
- The AAPE ranged from 0.97% to 3.07% for training and testing between the predicted and the actual pore pressure data. Moreover, the RMSE did not exceed 0.021 psi/ft for all models.
- The RF model outperformed the other models by an R of 0.99, RMSE of 0.011 psi/ft, and an AAPE of 0.97%. Furthermore, it predicted the pore pressure gradient for the validation stage with a high accuracy (R of 0.99, RMSE of 0.01 psi/ft, and AAPE around 1.19%).
- The common practice while over balanced drilling is to have a safety margin over the pore pressure gradient, which, in some cases, may be in the range of 0.1 psi/ft. As a result, gradient prediction with 0.02 psi/ft RMSE is reasonable.

### APPENDIX 1

The formula of the Pearson correlation coefficient (R), between any two variables (x, y), used in this study is expressed as

\[
R = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}
\]

AAPE is expressed as

\[
\text{AAPE} = \frac{\sum \frac{P_{\text{measured}} - P_{\text{predicted}}}{P_{\text{measured}}} \times 100}{n}
\]

where \(P_g\) is the pressure gradient and \(n\) is the number of points.

### AUTHOR INFORMATION

#### Corresponding Author
Salaheldin Elkatatny — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; orcid.org/0000-0002-7209-5715; Email: elkatatny@kfupm.edu.sa

#### Authors
Ahmed Abdelaal — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
Abdulazeez Abdulraheem — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c01340

#### Author Contributions
This manuscript was written through the contributions of A.A. and S.E. All authors have approved the final version of this manuscript.

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

#### Notes
The authors declare no competing financial interest.

#### ACKNOWLEDGMENTS

The authors acknowledge the King Fahd University of Petroleum & Minerals for permitting the publication of this work.

#### ABBREVIATIONS

- \(P_g\) pressure gradient
- \(\text{ANN}\) artificial neural network
- \(\text{ANFIS}\) adaptive network-based fuzzy interference system
- \(R\) correlation coefficient
- \(\text{AAPE}\) average absolute percentage error
- \(\text{SVM}\) support vector machine
- \(\text{FN}\) functional network
- \(\text{FNEM}\) functional network exhaustive search method
- \(\text{FNFBM}\) functional network forward method
- \(\text{WOB}\) weight on bit
- \(\text{rpm}\) revolutions per minute
- \(\text{RS}\) rotating speed
- \(\text{RF}\) random forest
- \(\text{FL}\) fuzzy logic
- \(\text{ROP}\) rate of penetration
- \(\text{SPP}\) standpipe pressure
- \(T\) torque
- \(\text{GRNN}\) general regression neural network
- \(\text{BPNN}\) back-propagation neural network
- \(\text{RBF}\) radial basis function
- \(\text{MW}\) mud weight
- \(\text{PWD}\) pressure while drilling
- \(\text{RMSE}\) root-mean-squared error
**REFERENCES**

1. Bourgoyne, J. A. T.; Millheim, K. K.; Cheenvert, M. E.; Young, J. F. S. Applied Drilling Engineering; Society of Petroleum Engineers: Richardson, TX, 1986; Vol. 2.

2. Mouchet, J. P.; Mitchell, A. Abnormal Pressures While Drilling; Moskou: Boussens, 1989.

3. Rabia, H. Well Engineering & Construction Hussain Rabia; Entrac Consulting Limited: London, 2002.

4. Tingay, M. R. P.; Hillis, R. R.; Swarbrick, R. E.; Morley, C. K.; Damit, A. R. Origin of Overpressure and Pore-Pressure Prediction in the Baram Province, Brunei. *Am. Assoc. Pet. Geol. Bull.* 2009, 93, 51–74.

5. Hottman, C. E.; Johnson, R. K. Estimation of Formation Pressures from Log-Derived Shale Properties. *J. Pet. Technol.* 1965, 17, 717–722.

6. Pennebaker, E. S.; Detection of Abnormal-Pressure Formations from Seismic-Field Data. In *Drilling and Production Practice*; American Petroleum Institute, 1968; pp 184–191.

7. Matthews, W. R.; Kelly, J. How to Predict Formation Pressure and Fracture Gradient from Electric and Sonic Logs. *Oil Gas J.* 1967, 65, No. 10, 1066.

8. Eaton, B. A. The Equation for Geopressure Prediction from Well Logs. *Presented at the Fall Meeting of the Society of Petroleum Engineers of AIME*; Dallas, TX, 1975. DOI: 10.2118/5544-m

9. Gardner, G. H. F.; Gardner, L. W.; Gregory, A. R. Formation Velocity and Density—The Diagnostic Basics for Stratigraphic Traps. *Geophysics* 1974, 39, 770–780.

10. Bowers, G. L. Pore Pressure Estimation from Velocity Data: Accounting for Overpressure Mechanisms Besides Undercompaction. *SPE Drill. Completion* 1995, 10, 89–95.

11. Foster, J. B.; Whalen, H. E. Estimation of Formation Pressures from Electrical Surveys—Offshore Louisiana. *SPE Repr. Ser.* 1966, 18, 57–63.

12. Ham, H. H. A. Method of Estimating Formation Pressures from Gulf Coast Well Logs. *Trans. – Gulf Coast Assoc. Geol. Soc.* 1966, 16, 185–197.

13. Eaton, B. A. The Effect of Overburden Stress on Geopressure Prediction from Well Logs. *J. Pet. Technol.* 1972, 24, 929–934.

14. Bingham, M. A New Approach to Interpreting Rock Drillability, Reprinted from Oil and Gas J.; Petroleum Pub. Co., 1965.

15. Jorden, J. R.; Shirley, O. J. Application of Drilling Performance Data to Overpressure Detection. *SPE Repr. Ser.* 1967, 18, 19–26.

16. Rehm, B. Mclendon, R. Measurement of Formation Pressure from Drilling Data. *Presented at the Fall Meeting of the Society of Petroleum Engineers of AIME*, New Orleans, Louisiana, 1971. DOI: 10.2118/3601-m

17. Contreras, O.; Hareland, G.; Aguiler, R. An Innovative Approach for Pore Pressure Prediction and Drilling Optimization in an Abnormally Subpressured Basin. *SPE Drill. Completion* 2012, 27, 531–545.

18. Mohaghegh, S. Virtual-Intelligence Applications in Petroleum Engineering: Part I—Artificial Neural Networks. *J. Pet. Technol.* 2000, 52, 64–73.

19. Elsafi, S. H. Artificial Neural Networks (ANNs) for Flood Forecasting at Dongola Station in the River Nile, Sudan. *Alexandria Engr. J.* 2014, 53, 655–662.

20. Doraissen, H.; Erkkin, T.; Grader, A. S. Key Parameters Controlling the Performance of Neuro-Simulation Applications in Field Development. In *Proceedings – SPE Annual Western Regional Meeting*; Society of Petroleum Engineers, 1998; pp 233–241. DOI: 10.2118/51079-m

21. Al-Abdul Jabbar, A.; Gamal, H.; Elkatatny, S. Artificial Neural Network to Predict the Rate of Penetration for S-Shape Well Profile. *Arabian J. Geosci.* 2020, 13, 1–11.

22. Mahmoud, A.; Elkatatny, S.; Chen, W.; Abdulraheem, A. Estimation of Oil Recovery Factor for Water Drive Sandy Reservoirs through Applications of Artificial Intelligence. *Energies* 2019, 12, 1–13.

23. Moazzeni, A.; Haffar, M. A. Artificial Intelligence for Lithology Identification through Real-Time Drilling Data. *J. Earth Sci. Clim. Change* 2015, 6, 1–4.

24. Fatehi, M.; Asadi, H. H. Data Integration Modeling Applied to Drill Hole Planning through Semi-Supervised Learning: A Case Study from the Dalii Cu-Au Porphyry Deposit in the Central Iran. *J. Afr. Earth Sci.* 2017, 128, 147–160.

25. Ren, X.; Hou, J.; Song, S.; Liu, Y.; Chen, D.; Wang, X.; Dou, L. Lithology Identification Using Well Logs: A Method by Integrating Artificial Neural Networks and Sedimentary Patterns. *J. Pet. Sci. Eng.* 2019, 182, No. 106336.

26. Al-Abdul Jabbar, A.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A. Predicting Formation Tops While Drilling Using Artificial Intelligence. –Presented at *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition* 2018, SATS 2018; Society of Petroleum Engineers, 2018. DOI: 10.2118/192345-ms

27. Elkatatny, S.; Mahmoud, M. Development of New Correlations for the Oil Formation Volume Factor in Oil Reservoirs Using Artificial Intelligent White Box Technique. *Petroleum* 2018, 4, 178–186.

28. Zazoua, R. S. Fracture Density Estimation from Core and Conventional Well Logs Data Using Artificial Neural Networks: The Cambro-Ordovician Reservoir of M-esdar Oil Field, Algeria. *J. Afr. Earth Sci.* 2013, 83, 55–73.

29. Alsaihati, A.; Elkatatny, S.; Mahmoud, A. A.; Abdulraheem, A. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities While Drilling Horizontal Wells, with Real Case Study. *J. Energy Resour. Technol.* 2021, 143, No. 043201.

30. Okpo, E. E.; Dosunmu, A.; Odagme, B. S. Artificial Neural Network Model for Predicting Wellbore Instability. –Presented at the *SPE Nigeria Annual International Conference and Exhibition*; Society of Petroleum Engineers, 2016; DOI: 10.2118/184371-ms

31. Elkatatny, S.; Tariq, Z.; Mahmoud, M.; Mohamed, I.; Abdulraheem, A. Development of New Mathematical Model for Compressional and Shear Sonic Times from Wireline Log Data Using Artificial Intelligence Neural Networks (White Box). *Arab. J. Sci. Eng.* 2018, 43, 6375–6389.

32. Abdulmolek Ahmed, S.; Elkatatny, S.; Abdulraheem, A.; Mahmoud, M.; Ali, A. Z. New Approach to Predict Fracture Pressure Using Functional Networks. –Presented at *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition* 2018, SATS 2018; Society of Petroleum Engineers, 2018; DOI: 10.2118/192317-ms

33. Mahmoud, A.; Elkatatny, S.; Mahmoud, M.; Abouelresh, M.; Abdulraheem, A.; Ali, A. Determination of the Total Organic Carbon (TOC) Based on Conventional Well Logs Using Artificial Neural Network. *Int. J. Coal Geol.* 2017, 179, 72–80.

34. Tariq, Z.; Elkatatny, S.; Mahmoud, M.; Ali, A. Z.; Abdulraheem, A. A New Approach to Predict Failure Parameters of Carbonate Rocks Using Artificial Intelligence Tools. *Presented at Society of Petroleum Engineers – SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition* 2017; Society of Petroleum Engineers, 2017; pp 1428–1440. DOI: 10.2118/187974-ms

35. Arehart, R. A. Drill-Bit Diagnosis With Neural Networks. *SPE Comput. Appl.* 1990, 2, 24–28.

36. Abdelgawad, K.; Elkatatny, S.; Moussa, T.; Mahmoud, M.; Patil, S. Real-Time Determination of Rheological Properties of Spud Drilling Fluids Using a Hybrid Artificial Intelligence Technique. *J. Energy Resour. Technol., Trans. ASME* 2019, 141 (4). DOI: 10.1115/1.4042253

37. Elkatatny, S. Real-Time Prediction of Rheological Parameters of KC1 Water-Based Drilling Fluid Using Artificial Neural Networks. *Arab. J. Sci. Eng.* 2017, 42, 1655–1665.

38. Alsabaa, A.; Gamal, H.; Elkatatny, S.; Abdulraheem, A. Real-Time Prediction of Rheological Properties of Invert Emulsion Mud Using Adaptive Neuro-Fuzzy Inference System. *Sensors* 2020, 20, No. 1669.

39. Elkatatny, S.; Tariq, Z.; Mahmoud, M. Real Time Prediction of Drilling Fluid Rheological Properties Using Artificial Neural Networks Visible Mathematical Model (White Box). *J. Pet. Sci. Eng.* 2016, 146, 1202–1210.
(40) Mahmoud, A. A.; Elkatatny, S.; Al Shehri, D. Application of Machine Learning in Evaluation of the Static Young’s Modulus for Sandstone Formations. *Sustainability* 2020, 12, 1880.

(41) Mahmoud, A. A.; Elkatatny, S.; Ali, A.; Moussa, T. Estimation of Static Young’s Modulus for Sandstone Formation Using Artificial Neural Networks. *Energies* 2019, 12, No. 2125.

(42) Ahmed, A.; Elkatatny, S.; Ali, A.; Mahmoud, M.; Abdulraheem, A. New Model for Pore Pressure Prediction While Drilling Using Artificial Neural Networks. *Arab. J. Sci. Eng.* 2019, 44, 6079–6088.

(43) Ahmed, A.; Elkatatny, S.; Ali, A.; Abdulraheem, A. Comparative Analysis of Artificial Intelligence Techniques for Formation Pressure Prediction While Drilling. *Arabian J. Geosci.* 2019, 12, No. 592.

(44) Alouane, L.; Ouaudef, S.-A.; Boudella, A. Pore Pressure Prediction in Shale Gas Reservoirs Using Neural Network and Fuzzy Logic with an Application to Barnett Shale, Vol. 17. In *EGU General Assembly*, Austria, 2015.

(45) Hu, L.; Deng, J.; Zhu, H.; Lin, H.; Chen, Z.; Deng, F.; Yan, C. A New Pore Pressure Prediction Method-Back Propagation Artificial Neural Network. *Electron. J. Geotech. Eng.* 2013, 18, 4093–4107.

(46) Li, W.; Yan, T.; Liang, Y. Pressure Prediction Technology of the Deep Strata Based On BP Neural Network. In *Advanced Materials Research*, Vol. 143–144; Trans Tech Publications Ltd, 2010; pp 28–31. DOI: 10.4028/www.scientific.net/AMR.143-144.28

(47) Rashidi, M.; Asadi, A. An Artificial Intelligence Approach in Estimation of Formation Pore Pressure by Critical Drilling Data. *52nd U.S. Rock Mechanics/Geomechanics Symposium*; 2018, No. 1959.

(48) Keshavarzi, R.; Jahanbakshi, R. Real-Time Prediction of Pore Pressure Gradient through an Artificial Intelligence Approach: A Case Study from One of Middle East Oil Fields. *Eur. J. Environ. Civ. Eng.* 2013, 17, 675–686.

(49) Hemphill, T. Validation of Drillpipe Rotation Hydraulics Using Drillpipe Eccentricity as a Key Factor. *Presented at Society of Petroleum Engineers – SPE/IATMI Asia Pacific Oil and Gas Conference and Exhibition, APOGCE 2015*; Society of Petroleum Engineers, 2015. DOI: 10.2118/176451-ms

(50) Thunder, M.; Moore, D. S.; McCabe, G. P. Introduction to the Practice of Statistics. *Math. Gaz.* 1995, 79, 252.

(51) Dawson, R. How Significant is a Boxplot Outlier? *J. Stat. Educ.* 2011, 19, 1–13.

(52) Head, A. L.; A Drillability Classification of Geological Formation. *Presented at the 3rd World Petroleum Congress, OnePetro*, The Hague, The Netherlands, May 1951.

(53) Mensa-Wilmot, G.; Callhoun, B.; Perrin, V. P. Formation Drillability-Definition, Quantification and Contributions to Bit Performance Evaluation. *Presented at the SPE/IADC Middle East Drilling Technology Conference*, Abu Dhabi, United Arab Emirates, November 1999. DOI: 10.2118/57558-MS

(54) Anifowose, F.; Abdulraheem, A. Fuzzy Logic-Driven and SVM-Driven Hybrid Computational Intelligence Models Applied to Oil and Gas Reservoir Characterization. *J. Nat. Gas Sci. Eng.* 2011, 3, 505–517.

(55) Vapnik, V. N. *The Nature of Statistical Learning Theory*; Springer: New York, 1995.

(56) Awad, M.; Khanna, R. Support Vector Machines for Classification. In *Efficient Learning Machines*; Apress: Berkeley, CA, 2015; pp 39–66. DOI: 10.1007/978-1-4302-5990-9_3

(57) Castillo, E.; Gutiérrez, J. M.; Hadi, A. S.; Lacruz, B. Some Applications of Functional Networks in Statistics and Engineering. *Technometrics* 2001, 43, 10–24.

(58) Anifowose, F.; Labadin, J.; Abdulraheem, A. A Least-Square-Driven Functional Networks Type-2 Fuzzy Logic Hybrid Model for Efficient Petroleum Reservoir Properties Prediction. *Neural Comput. Appl.* 2013, 23, 179–190.

(59) Castillo, E.; Cobo, A.; Gutiérrez, J. M.; Pruneda, R. E. *Functional Networks with Applications*; Springer US, Boston, MA, 1999. DOI: 10.1007/978-1-4615-5601-5

(60) Ahmed, A.; Khalid, M. An Intelligent Framework for Short-Term Multi-Step Wind Speed Forecasting Based on Functional Networks. *Appl. Energy* 2018, 225, 902–911.

(61) Zhang, C.; Ma, Y. *Ensemble Machine Learning: Methods and Applications*; Springer-Verlag, New York, 2012.

(62) Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32.

(63) Hegde, C.; Gray, K. E. Use of Machine Learning and Data Analytics to Increase Drilling Efficiency for Nearby Wells. *J. Nat. Gas Sci. Eng.* 2017, 40, 327–335.