Petroleum and Coal

Article **Open Access**

Pore and Fracture Pressures Real Time Prediction during Drilling Operation Using a Regression Ensemble Machine Learning Model

Tarek M. Barghoth[1], Shady G. Elramah[1], Adel M. Salem[1,2]*

[1] *Petroleum Engineering Department, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt*

[2] *Department of Petroleum Engineering, Faculty of Engineering and Technology, Future University in Egypt (FUE), Cairo, Egypt*

Received July 18, 2024; Accepted October 16, 2024

Abstract

Estimating pore and fracture pressures is critical in achieving a successful drilling operation both onshore and offshore. With proper estimation of pore pressure and fracture pressure, drilling engineers can achieve accurate mud weight design, balancing and stabilizing the formation without fracturing it, determining casing setting depths, determining casing pressure loads, testing and analyzing the integrity of barriers and doing well control assessments. Meanwhile the instability of wellbore can cause several problems in drilling operations. So optimizing mud weight design through accurate pore and fracture pressures can prevent these potential problems. There are various predictive mathematical techniques and models that are used in predicting pore and fracture pressures but usually these techniques are limited and cannot predict them comprehensively and accurately.

This research paper shows a model based on the machine learning techniques for predicting pore and fracture pressures that is a regression ensemble (mean) model aggregating six intelligent predictive models and algorithms, which are Random Forest, Decision Tree, XGBoost, Extra Trees, Adaboost and Gradient Boosting models and algorithms. Which shows a blind test accuracy of prediction of the final model of 0.999 for Pore Pressure and 0.998 for Fracture Pressure from only mud logs, surface drilling parameters, evaluation of drill cuttings and gas levels at surface that are available easily. The aforementioned six algorithms, as well as the ensemble (mean) model, are trained using the 5 fold cross-validation (CV) to optimize the algorithm structure, while the statistical accuracy parameters (e.g. R2, MSE, MAE, MDAE, MAPE %, MDAPE %) are used as the criterion of selection the best algorithm in terms of the prediction accuracy.

Keywords: *Pore pressure; Fracture pressure; Real time prediction; Modelling; Data analysis; Artificial intelligence; Machine learning techniques; Nile Delta of Egypt.*

1. Introduction

Accurate and well-timed pore and fracture pressures prediction is a "Driller's Tool", as it is an aid to the drillers of a gas or oil well that lets them enhance the drilling process for time and cost, and even optimize the final implemented well design [1].

Pore pressure gradient and fracture gradient are the most crucial variables practically employed to establish the window of (mud weight or mud density) in drilling engineering. Prior to installing a casing, the mud weight design should be properly chosen based on the wellbore stability, fracture gradient and pore pressure gradient, (Figure 1) [2-3].

2. Pore pressure gradient concept

Forming pressure is another name for pore pressure. It is sometimes referred to as the mud weight of well and is comparable to hydraulic potential. Pore pressure can be classified as normal, subnormal, or abnormal. Depending on the area and magnitude, the typical values for the normal pore pressure are 9 ppg and 0.465 psi/ft. The subnormal pressure, on the other hand, is below 9 ppg while the abnormal pressure is larger than 9 ppg [4].

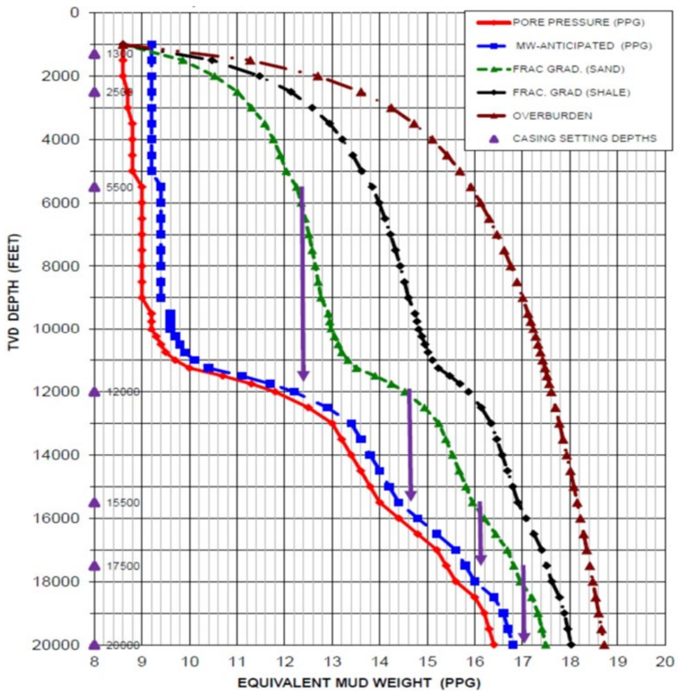

Figure 1. Depth (TVD) versus pore pressure gradient, anticipated mud weight, overburden pressure gradient, fracture gradient (shale & sand) and casing shoes.

2.1. Pre-drilling phase pore pressure prediction

In general, the evaluation of pore pressure is based primarily on the correlation of available data from nearby wells (well history) and seismic data, from analyzing the offset well data using various predictive techniques and from planning for development wells involves using information from previous regional drilling experiences [5].

There are some limitations and challenges for such methods as for exploration wells, our dependency is only on seismic data. Field data must be collected from previously drilled wells in the formation. Raw filed data should be processed into a form where interpretations can be made [6].

2.2. Drilling phase pore pressure estimation

In general, there are three ways to estimate and quantify pore pressure during the drilling phase. First mud logging techniques, these take into account drilling parameters as well as drill cuttings and gas levels at the surface. Logging while drilling (LWD), wireline logging, and measurement while drilling (MWD). Direct techniques, RFT (Repeat Formation Tester), DST (Drill Stem Test), and production testing [7-8].

Terzaghi published the first research on the prediction of pore pressure in 1943, and an empirical equation was created to calculate the pore pressure [9]. In 1965, Hottmann and Johnson conducted research to estimate pore pressure, taking into account the characteristics of shale and the variation in sound velocity noted by sonic logs [10].

Biot Willis proposed an empirical link (Eq. 1) between effective stress, overburden pressure, and pore pressure [11].

$$\mathrm{PP} = \frac{\sigma_{\mathrm{ov}} - \sigma_{\mathrm{eff}}}{\beta} \tag{1}$$

Eaton method [12] estimate pore pressure through the following equations (Eq. 2).

$$\mathrm{PP} = \sigma_{\mathrm{ov}} - (\sigma_{\mathrm{ov}} - \mathrm{P_n}) \times \left(\frac{\mathrm{d_{co}}}{\mathrm{d_{cn}}}\right)^{1.2} \tag{2}$$

The ratio method is simpler and does not need overburden values and pore pressure can be estimated through the following equation (Eq. 3).

$$\mathrm{PP} = \mathrm{P_n} \times \left(\frac{\mathrm{d_{cn}}}{\mathrm{d_{co}}}\right) \tag{3}$$

The sonic log method (Eq. 4) has proved to be the most applicable calculation for the majority of sedimentary sequence compared with use of a trend curve that has been developed from one or several specific regions [13].

$$PP = \sigma_{ov} - (\sigma_{ov} - P_n) \times \left(\frac{\Delta t_n}{\Delta t_o}\right)^3 \tag{4}$$

The pore pressure can be estimated according to the Resistivity log theory for the depth of interest from the following equation (Eq. 5) [14-15].

$$\mathrm{PP} = \sigma_{ov} - (\sigma_{ov} - P_\mathrm{n}) \left(\frac{R_{\log}}{R_n}\right)^n \tag{5}$$

Pore pressure estimation through the direct methods usually includes RFT (Repeat Formation Tester), DST (Drill Stem Test) and production testing. The limitation of direct methods (RFT and DST) are the DST and RFT data provide definitive values of pore pressure for the well, these direct measurements are only possible in permeable formations, obtained after the well is drilled and not applicable to largely impermeable shale sections where the majority of overpressure is developed. In HPHT wells, the RFT and DST tools should be considered for use prior to performing potentially problematic drilling operations, such as coring [7].

2.3. Post drilling phase pore pressure prediction

Depends on empirical models in general that are a field-specific, i.e. these equations provide accurate predictions for the parameter of interest only for the field. To overcome this issue, many studies have been performed to develop predictive models for forecasting diverse parameters within the oil and gas industry applying artificial intelligence techniques [16].

Wang *et al.* performed a study for the prediction of the pore pressure in 2010, employing three methods: the trend line technique (TLM), the hybrid genetic algorithm without a mutation rate (HGANM) and the original Fillippone formula method (OFFM) [17]. Hu *et al.* employed the Propagation Artificial Neural Network (BPANN) approach to forecast the pore pressure based on data collected from five wells in two separate fields in 2013 with the average error recorded for the model being 7.15% [18]. Abidin employed an artificial neural network (ANN) technique to forecast the pore pressure in fields with abnormal and normal pressure in 2014 with performance accuracy is very high, at around 5.0048% [19]. Aliouane *et al.* used the multi-layer perceptron neural network (MLPANN) and fuzzy logic (FL) to estimate the pore pressure value in horizontal wells [20]. Haris *et al.* used the probabilistic neural network (PNN) approach to forecast pore pressure in 2017, The precision of the PNN model was 98% higher than relations created using seismic data [21]. Kiss *et al.* used an ANN system to forecast the PP parameter in 2018 based on two essential parameters: drilling efficiency (DE) and mechanical specific energy drilling (MSE) [22]. Rashidi and Asadi applied an ANN to a set of drilling data, drilling efficiency (DE) and mechanical specific energy (MSE), acquired from three wells drilled in an Iranian sandstone reservoir in 2018 [23]. Karmakar and Maiti created predictive models in 2019, for the pore pressure in well U1343E of the IODP at the Bering Sea slope area with a reduction error (RE) of nearly 0.98 [24]. Yu *et al.* used four inelegant prediction models to anticipate the PP in 2020; support vector machine (SVM), gradient boosting (GB), random forest method and multilayer perceptron (MLP) [25]. Andrian *et al.* used adaptive neuro fuzzy inference system (ANFIS) methods to estimate the pore pressure parameter with a 70% accuracy. Abdelaal *et al.* used 3100 drilling data records to create three models to forecast pore pressure during the drilling operation with of prediction accuracy (AAPE = 2% and R = 0.98) [26]. Radwan *et al.* predicted the PP parameter at the Mangahewa gas field in New Zealand using 25,935 data records [27].

3. Fracture pressure gradient concept

The fracture gradient in gas and oil wells drilling can be calculated by dividing the minimum horizontal in situ stress (σ_h) by the depth. Below the earth's crust there are three independently operating, perpendicularly arranged stresses to each one another that are exist; Two horizontal stresses σ_H and σ_h and a vertical stress σ_V are collectively referred to as the normal stresses. Generally, the drilled formations are subjected to in-situ stresses with no shear stresses in the majority of gas and oil wells applications. Principal stresses are normal stresses that are perpendicular to one another and do not have any accompanying shear stresses; the maximum, intermediate and minimum principal stress (σ_1 ,σ_2 , σ_3).

The size of the min stress underground will therefore play a significant role in determining the pressure needed to fracture the formation. Since all subsurface stresses are connected, the following regular connections can be observed. The fracture gradient rises with increasing overburden. The fracture gradient rises as pore pressure rises. Large pore pressure drops decrease the fracture gradient. A number of parameters, including formation type, mineralogy, rock strength, permeability, and the direction of weak planes such as bedding planes affect the fracture gradient [28].

3.1. Pre-drilling phase fracture pressure prediction

In general, there are two major methods for determining fracture pressure. The first approach is the direct method, in which the pressure necessary to fracture the formation and cause propagation is measured directly. Many direct approaches rely on the Formation Integrity Test (FIT), the formation Leak-off Test (LOT) or the Extended Leak-off Test (XLOT) [29]. The second approach is the indirect method that uses correlations between rock and formation properties as well as stress analysis to predict fracture pressure. These correlations are carried out using rock and formation properties including (overburden stress and Poisson's ratio), as well as density and porosity obtained from well logs [5]. There are some Limitations and Challenges for such methods as for Exploration wells; our dependency is only on seismic data. In addition, field data must be collected from previously drilled wells in the exact formation strata. Raw filed data should be processed into a form where interpretations can be made [6].

3.2. Drilling phase fracture pressure estimation

The fracture pressure can be measured by two methods. The first method is the direct methods that includes Leak off Test (LOT) & Formation Integrity Test (FIT). The second method is the indirect methods, which rely on correlations and models.

During the drilling stage, Formation Integrity Tests (FIT, LOT, XLOT, etc.) are performed to identify the approximate amount of fracture gradient below each casing shoe. The FIT pressure is adjusted to an equivalent mud weight (EMW) to establish a maximum limit of the primary well control for the next hole section [30].

The second approach is the indirect method that uses correlations between rock and formation properties as well as stress analysis to predict fracture pressure. These correlations are carried out using rock and formation properties including (overburden stress and Poisson's ratio), as well as density and porosity obtained from well logs.

According to the Hubbert and Willis approach [31], it was believed that 33% of the overburden stress would serve as the lower limit of the fracture pressure (Eq. 6) and that the upper limit of the fracture pressure would not exceed 50% of the overburden stress (Eq. 7).

$$FG_{min} = \frac{1}{3}\left(\frac{\sigma_{\text{ov}}}{D} + 2\,\frac{P_f}{D}\right) \qquad (6)$$

$$FG_{max} = \frac{1}{2}\left(\frac{\sigma_{\text{ov}}}{D} + \frac{P_f}{D}\right) \qquad (7)$$

Matthews and Kelly [32] proposed a quantity known as the "matrix stress coefficient-K_i" (Eq. 8) which is equivalent to the effective stress coefficient and depends generally on the depth and formation pressure.

$$FG = \frac{P_f}{D} + K_i \frac{\sigma}{D} \quad (8)$$

Pennebaker modified Matthews and Kelly's matrix stress coefficient (K_i) correlation by assuming that the matrix stress coefficient-K_i varies with depth and formation type (Eq. 9) [33].

$$FG = \frac{P_f}{D} + K_P \frac{\sigma}{D} \quad (9)$$

Eaton calculated the fracture gradient using Poisson's ratio of the formation and the idea of the minimal injection pressure provided by Hubbert and Willis [34].

$$\mathrm{FG} = \left(\frac{v}{1-v}\right)\left(\frac{(\sigma_{\mathrm{ov}} - \mathrm{P_f})}{\mathrm{D}}\right) + \frac{\mathrm{P_f}}{\mathrm{D}} \quad (10)$$

Anderson *et al.* (1973) proposed an empirical equation for predicting fracture pressure that is a function of formation pressure, overburden stress, Poisson's ratio, depth and the compressibility ratio of the porous to bulk rock matrix. This concept is dependent on Biot's stress-strain relationships (Eq. 11) [35].

$$\mathrm{FP} = \alpha \mathrm{P_f} + \frac{2v}{1-v}(\sigma_{ov} - \alpha \mathrm{P_f}) \quad (11)$$

Daines superimposed Eaton's equation with a horizontal tectonic stress (σ_t). In terms of stress, he defined it as "the min pressure within the wellbore to extend and hold open an existing fracture" which can be expressed in the following equation (Eq. 12) [36].

$$\mathrm{FP} = \frac{v}{1-v}(\sigma_{\mathrm{v}} - P_{\mathrm{f}}) + P_{\mathrm{f}} + \sigma_{\mathrm{t}} \quad (12)$$

The tensile failure pressure in case of vertical well can be determined from Kirsch's borehole solution which is applicable for the impermeable case (non-penetrating fluid), as expressed in the following equation (Eq. 13) by Haimson and Fairhurst. This pressure was known as the formation breakdown pressure [37].

$$\mathrm{FBG} = 3\sigma_{\mathrm{h}} - \sigma_{\mathrm{H}} - \mathrm{P_f} + \mathrm{T_0} \quad (13)$$

(Eq. 13) can be simplified into the following form (Eq. 14) if we assume that $(\sigma_{max} - T_0)$ is nearly equal to the (σ_{min}) and ignoring the temperature effects [38].

$$FP_{\max} = 2\sigma_{\mathrm{h}} - P_{\mathrm{f}} \quad (14)$$

$$FP_{\min} = \sigma_{\mathrm{h}} = \frac{v}{1-v}(\sigma_{\mathrm{V}} - P_{\mathrm{f}}) + P_{\mathrm{f}} \quad (15)$$

The most likely fracture pressure is the average of the fracture pressures (or gradients) upper and lower bounds as shown in (Eq. 16) [38-39].

$$FP_{\mathrm{avg}} = \frac{3v}{2(1-v)}(\sigma_{\mathrm{V}} - P_{\mathrm{f}}) + P_{\mathrm{f}} \quad (16)$$

where FP_{avg} represents the most likely fracture pressure.

3.3. Post drilling phase fracture pressure prediction

Reviewing the results viewed in the majority of the studies in which an empirical model was proposed for predicting the parameters included in the gas and oil industry. It was discovered that these models give accurate predictions for the parameter of interest only for the field whose data was used in the generation of the empirical equations (i.e. such empirical models are in general field specific) [40]. So many studies have recently been conducted to develop predictive models for forecasting various parameters in the gas and oil industry using artificial intelligence and machine learning techniques [16].

Sadiq and Nashawi used depth, Poisson's ratio and overburden stress gradient to forecast fracture pressure using RBF and ANN algorithms [29]. Malallah and Nashawi estimated the fracture gradient using feed-forward artificial neural networks (ANN) with average absolute relative error of 6.5% and the average relative error of 3.7% [41]. Keshavarzi *et al.* predicted the fracture gradient using (ANN) the neural network with R = 0.9962 for the training, R = 0.9928 for the validation and R = 0.9827 for the testing and was obtained by combining a feed-forward neural network with back propagation neural network [42]. Abdulmalek *et al.* use support vector machine (SVM) to estimate fracture and pore pressures with high accuracy as the determination coefficient (R2) is higher than 0.995 with utilizing various parameters such as well logs and real-time surface drilling parameters [43]. Elkatatny *et al.* use an artificial neural network (ANN) for estimating fracture pressure using over 3900 real field data points based solely on real time surface drilling parameters. The artificial neural network model (ANN) was compared to the Matthews and Kelly model that is one of the most commonly used models in the field for predicting fracture pressure [44]. Ahmed *et al.* predict the fracture pressure with using five ML and AI tools, named ANN, FN, FL, RBF, and SVM, using only real-time surface drilling parameters which are widely available. The pros of this method is that the fracture pressure can be obtained without having to access the measurement logs [45].

4. Problem statement

Pore and fracture pressures determination is critical in optimizing the mud weight design for a successful oil or gas well as it has a direct impact on the stability of wellbore which itself can lead to several problems in drilling operation [46].

There are various predictive mathematical techniques and models that have been used in determining and predicting the fracture and pore pressures but usually these techniques are limited and cannot predict them comprehensively and accurately.

Generally, these models are empirical models that are a field-specific, i.e. these equations provide accurate predictions for the parameter of interest only for the field and have a lot of limitations and restrictions. In the other hand, the direct tests (which itself can't be done unless the drilling has already started) like FIT, LOT, DST, RFT, etc. are costly, consume a lot of time, effort and money, have limitations and restrictions itself and can lead to severe unexpected problems.

So with the progress of technology of both hard and software, and by using artificial intelligence and machine learning techniques which reduce the computational time, build more reliable models, and able to interpret the data effectively. This helps a lot in forecasting the fracture and pore pressures based on field data for the developed wells.

This study aims to analyze the different methods that are used for predicting fracture and pore pressures. Introduce a new model using machine learning techniques for predicting real time pore and fracture pressures based on field data containing (mud logs & surface drilling parameters) that are available easily for "BE Filed" in The Nile Delta, Egypt.

5. Methodology

In this study, the methodology is going to be in the following order: collecting field data, cleaning and quality assuring the data. Developing a machine learning model by implementing a regression ensemble (mean) model aggregating six intelligent predictive models and algorithms named as (RF, DT, Adaboost, Xgboost, Extra trees, and Gradient boosting) for the Prediction and Analysis of pore and fracture pressures.

The six algorithms, as well as the ensemble (mean) model, are trained using the 5 fold cross-validation (CV) to optimize the algorithm structure, while the statistical accuracy parameters (e.g. R2, MSE, MAE, MDAE, MAPE %, MDAPE %) are used as the criterion of selection the best algorithm in terms of the prediction accuracy. Figure 2 shows the flow chart of steps that would be followed until the models are built.

Figure 2. Machine learning process train till developing the model.

5.1. Data set and facilities

The case is based on dataset for "BE Filed" in The Nile Delta, Egypt. For fracture & pore pressures prediction. The dataset includes one file. In particular, file Study on BE Field

Data.csv is used for this case. There are 2518 number of sample points. The target columns are “FP kg/cm^2 & PP kg/cm^2”. There are 21 feature columns which are: 'Depth', 'SGR', 'CGR', 'PEF', 'RHOB', 'NPHI’. Assistant Tools was used in this study include Open sources python libraries & SML [47]. Table 1 and Table 2 show detailed statistics and data visualization

Table 1. Detailed statistics and data visualization of features data.

Statistical	Depth	SGR	CGR	PEF	RHOB	NPHI	DTC	Vs	UCS	Porosity
Count	2518	331	331	331	331	331	331	331	331	331
Mean	2506.5	51.363	35.349	4.897	2.435	33.153	150.89	2444.48	1965.33	24.211
Std	727.028	13.838	14.488	4.065	0.146	11.778	12.55	0	0	12.454
Min	1248	14.73	-5.5	2.395	1.666	9.568	0	2444.48	1965.33	10.766
20%	1751.4	39.784	27.866	3.645	2.36	23.971	149	2444.48	1965.33	17.417
40%	2254.8	52.7	36.684	3.882	2.458	28.669	150.75	2444.48	1965.33	20.176
50%	2506.5	56.094	39.263	3.965	2.485	32.804	151.5	2444.4	1965.3	21.235
60%	2758.2	58.509	41.344	4.02	2.499	35.434	152.12	2444.4	1965.3	22.412
80%	3261.6	62.971	46.118	4.2	2.533	40.01	156.06	2444.4	1965.3	25.282
max	3765	71.805	56.752	26.09	2.588	83.687	170.09	2444.4	1965.3	78.149

Table 2. Detailed statistics and data visualization of features data

Statistical	INC	Azimuth	WOB	RPM	TQ	ROP m/hr	PP kg/cm^2	FP kg/cm^2
count	2518	2518	2518	2518	2518	2518	2518	2518
mean	4.974	227.391	19.62	118.16	6331.748	17.975	277.877	420.086
std	6.265	87.287	7.469	29.877	2211.405	9.941	69.104	124.916
min	0.1	2.4	0	0	0	1.356	128.47	198.645
20%	0.3	99.7	12.858	117.81	4776.249	10.384	198.692	290.003
40%	0.7	264.16	17	125.42	6196.683	14.44	280.21	378.04
50%	1	268	18.972	127.27	6643.324	16.212	293.76	419.946
60%	3.08	269.4	21.726	128.94	7040.368	18.331	303.842	462.634
80%	12	280.47	27.294	132.51	8052.346	23.747	344.128	550.408
Max	18.2	327.5	41.632	156.65	13136.45	91.044	382.91	649.583

5.2. Features correlation

The matrix correlation (Figure 3) shows the Spearman correlation among features; Spearman coefficient is to measure the rank correlation. Both very large positive coefficient and very small negative coefficient indicate strong correlation between the features. Coefficient close to zero indicates weak ranking correlation. A positive value close to 1 means the two features have nearly the same ranking orders.

Pairwise correlation (Figure 4) shows the pair plot map presents the pairwise relationships in this dataset. It created a grid of Axes such that each feature in dataset shared the same y-axis across a single row and the same x-axis across a single column. Only the feature columns most correlated with the target columns are selected.

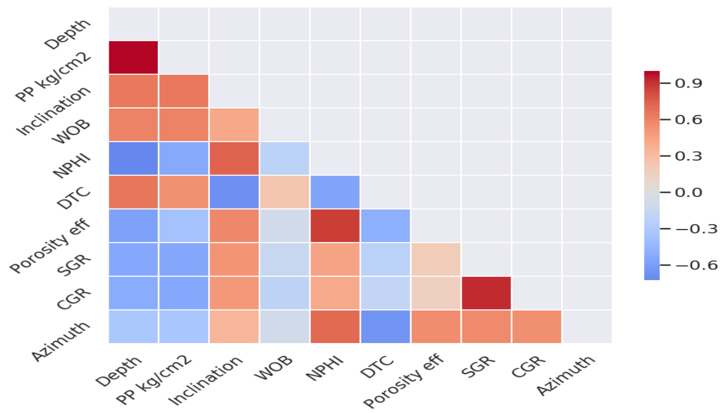

Figure 3. Matrix correlation chart among different features for the developed model.

5.3. Model development

The first step to start your machine-learning model. Select the data features file, a target column that will be used with your model to make predictions; one for PP and the other for FP and your problem type (It will be regression).

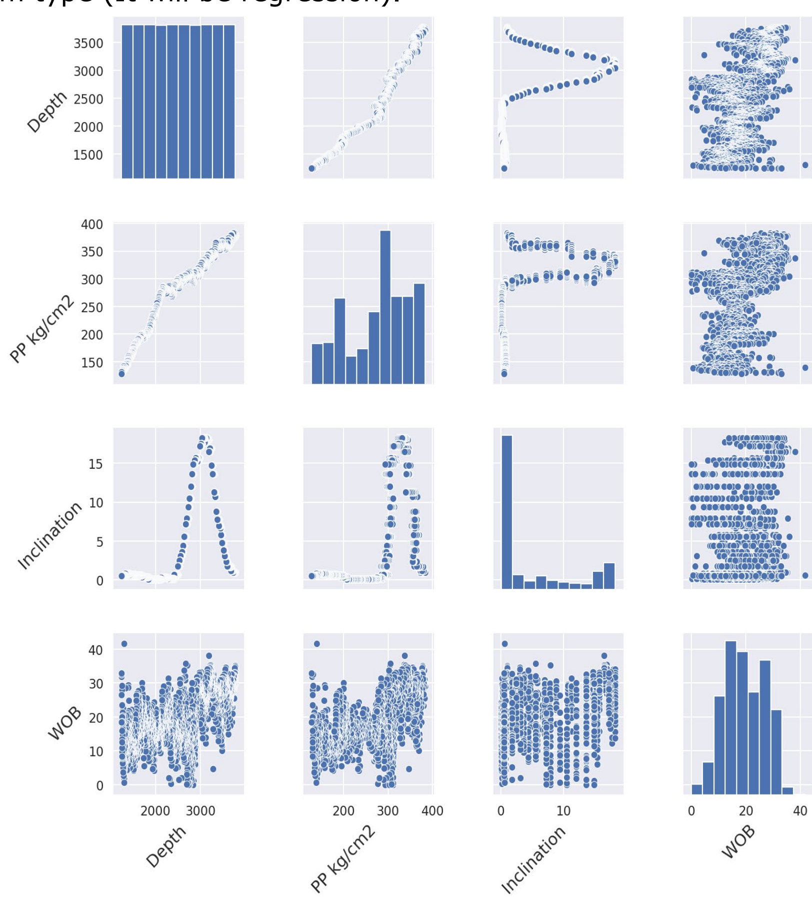

Figure 4. Pairwise correlation chart among different features for the developed model.

5.4. Data preprocessing

Data preprocessing is a technique and a necessary step to process the input data to help machine-learning algorithms easier to uncover the information hidden in the data.

Four preprocessing steps took places. Train/Test Split, Impute Missing, Categorical Encoder, and Treat Outliers. (Table 3) summarizes the statistical data of string feature columns after preprocessing step. First, train/test split is performed using stratified 5-fold CV method. Second, categorical features (lithology, resistivity and fluid type) are encoded using one hot encoding. Third, the missing values of features UCS, PEF, DTC, Porosity eff., RHOB, Vs., SGR, NPHI and CGR are imputed based on using mean mode method. Finally, the outliers has been removed using the probability threshold method.

5.5. Model training

Training a machine learning model is to uncover the relationship between features and target and to have the model learn the true underlying pattern in the data while stay robust again inevitable noise. The data is split into training data (2014 samples) and test data (504 samples) stratified based on feature ''FP & PP kg/cm^2''. The training data is further split into true training data and validation data points using k-fold with k=5. In this study, the 6 models are trained, separated with preprocessed data and then ensembled into a final model using conducts hyperparameter tuning trying to find the best model that has low bias and low variance.

Table 3. Detailed statistics of string features data after preprocessing.

Statistical	Lithology _shale	Lithology _shaly sand	Lithology _sandy shale	Lithology _anhydrite	Res_low	Res_ intermediate	Fluid type - water
Count	2014	2014	2014	2014	2014	2014	2014
Mean	0.015	0.049	0.922	0	0.955	0.028	0.983
std	0.121	0.215	0.269	0.022	0.207	0.164	0.129
min	0	0	0	0	0	0	0
20%	0	0	1	0	1	0	1
40%	0	0	1	0	1	0	1
50%	0	0	1	0	1	0	1
60%	0	0	1	0	1	0	1
80%	0	0	1	0	1	0	1
max	1	1	1	1	1	1	1

5.6. Random forest

Random forests is an ensemble learning method for classification and regression that operates by constructing a multitude of decision trees attraining time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [48].

5.7. Decision tree

DT is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements [49].

5.8. XGBoost

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. XXX total number of hyperparameter tuning iterations were performed. Each hyperparameter configuration are sampled from the following user specified range [50].

5.9. Extra trees

Extra Tree is a meta-estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various subsamples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [51].

5.10. AdaBoost

AdaBoost algorithm is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases [52].

5.11. Gradient boosting

Gradient Boostng produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boostng methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function [53].

5.12. Ensemble (mean) model

Model ensemble is a process of combining different machine models into a final model. There are different ways of combining the models. An ensemble model usually has more stable performance, low bias and less variance when predicting on new data points that it has not yet seen before [54]. In this case, we use EA method to combine the based models: randomforest, decisiontree, xgboost, extratrees, adaboost, gradient boosting.

6. Results and discussion

6.1.Model evaluation

Model evaluation is one of the most critical component of a machine learning study. By evaluating the trained machine-learning model on training, validation and test dataset, we can obtain an unbiased impression on how dependable and generalizable our trained machine-learning model could be.

Table 4 and Table 5 represent the statistical results and model evaluation obtained from the six developed models plus the ensemble one for training and testing of the total dataset presented for pore and fracture pressures respectively.

Table 4. Pore pressure statistical results & model evaluation.

Method	Dataset	Explained variance	Neg_AE	Neg_MAPE	Neg_MSE	Neg_MDAE	Neg_MDAPE	R2
Randomforest	Train	0.9999	-0.5629	-0.2009	-0.6077	-0.3986	-0.1487	0.9999
Randomforest	Test	0.9993	-1.3628	-0.4912	-3.3825	-1.0213	-0.385	0.9993
Decisiontree	Train	0.9994	-1.2322	-0.4317	-2.987	-0.79	-0.319	0.9994
Decisiontree	Test	0.9989	-1.7116	-0.6147	-5.2421	-1.3104	-0.5085	0.9989
Xgboost	Train	1	-0.329	-0.125	-0.198	-0.2369	-0.0892	1
Xgboost	Test	0.9993	-1.3768	-0.4938	-3.5041	-0.9988	-0.4059	0.9993
Extratrees	Train	0.9999	-0.3796	-0.1381	-0.277	-0.2612	-0.099	0.9999
Extratrees	Test	0.9993	-1.3123	-0.4756	-3.058	-1.0306	-0.3901	0.9993
Adaboost	Train	0.9973	-2.8927	-1.0699	-13.139	-2.4525	-0.9126	0.9973
Adaboost	Test	0.9969	-2.9926	-1.0926	-14.564	-2.4618	-0.9588	0.9969
gradientboosting	Train	1	-0.2559	-0.093	-0.1404	-0.163	-0.0632	1
gradientboosting	Test	0.9992	-1.4361	-0.5258	-3.6713	-1.0847	-0.4124	0.9992
ensemble_mean	Validation	0.9998	-0.7955	-0.2871	-1.042	-0.6353	-0.2423	0.9998
ensemble_mean	Test	0.9995	-1.4962	-0.537	-3.7598	-1.1706	-0.4495	0.9995

Table 5. Fracture pressure statistical results & model evaluation.

Method	Dataset	Explained variance	Neg_AE	Neg_MAPE	Neg_MSE	Neg_MDAE	Neg_MDAPE	R2
Randomforest	Train	0.9998	-0.8396	-0.1893	-3.6062	-0.3657	-0.098	0.9998
Randomforest	Test	0.9975	-2.5217	-0.5413	-38.7657	-1.0887	-0.2874	0.9975
Decisiontree	Train	0.9984	-2.4175	-0.535	-25.4809	-1.1928	-0.3115	0.9984
Decisiontree	Test	0.9968	-3.2327	-0.7013	-49.7875	-1.5	-0.3716	0.9968
Xgboost	Train	0.9997	-1.3075	-0.3102	-4.5737	-0.827	-0.2104	0.9997
Xgboost	Test	0.9975	-2.6515	-0.5662	-38.3338	-1.1206	-0.3048	0.9975
Extratrees	Train	0.9997	-0.8758	-0.1897	-3.9892	-0.3633	-0.0943	0.9997
Extratrees	Test	0.9975	-2.4615	-0.5234	-38.8934	-1.0295	-0.2727	0.9975
Adaboost	Train	0.9952	-6.5356	-1.5266	-76.3138	-5.0212	-1.2538	0.9951
Adaboost	Test	0.9944	-6.7914	-1.5678	-87.2614	-5.3953	-1.3017	0.9943
gradientboosting	Train	1	-0.3212	-0.0798	-0.2459	-0.2048	-0.0505	1
gradientboosting	Test	0.9975	-2.4793	-0.5285	-38.6702	-1.003	-0.2627	0.9975
ensemble_mean	Validation	0.9995	-1.6926	-0.3839	-7.8143	-1.031	-0.2607	0.9995
ensemble_mean	Test	0.9985	-2.8514	-0.6186	-38.498	-1.3476	-0.3386	0.9985

Figure 5 and Figure 6 show the regression plot of the true FP & PP and predicted FP & PP for training dataset Ensemble model, respectively.

Figure 7 and Figure 8 show the regression plot of the true FP & PP and predicted FP & PP for testing dataset Ensemble model, respectively.

For better comparison, the two figures are placed in one plot; Lines 0, +20% and -20% error are provided for reference see Figure 9 & Figure 10, respectively.

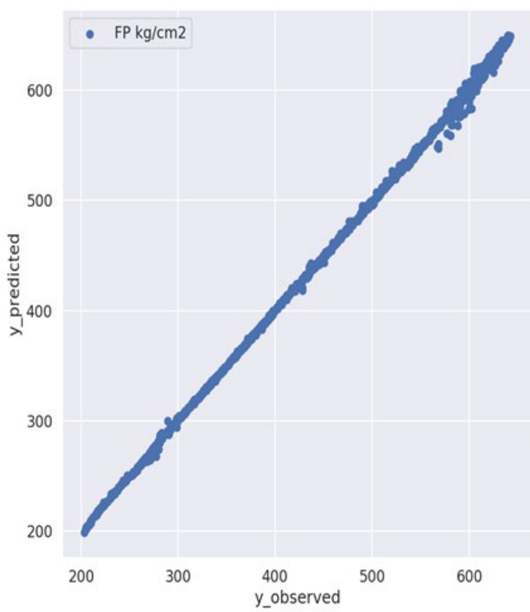

Figure 5. Regression plot of the true FP and predicted FP for train dataset for Ensemble model.

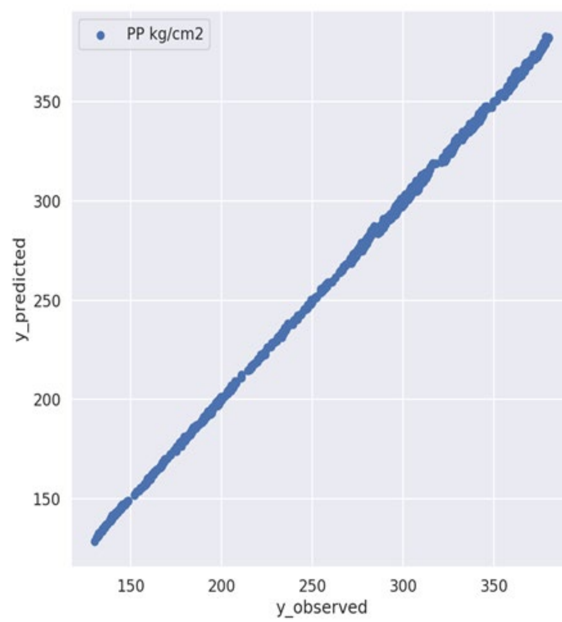

Figure 6. Regression plot of the true PP and predicted PP for train dataset for Ensemble model.

According to the above shown results, the Ensemble (Mean) model can be considered the best accurate model compared to the other developed model. In addition, this regression plot of the ensemble model has the highest correlation between the predicted and true values.

Figure 11 and Figure 12 show feature contribution to the end value (PP & FP). Figure 13 & Figure 14 show a comparison between predicted and true FP & PP using ensemble (mean) model for the dataset Features with top contribution are selected in this plot. In other words, the feature, which has the largest value of each class, contributes the most impact on determining final corresponding class.

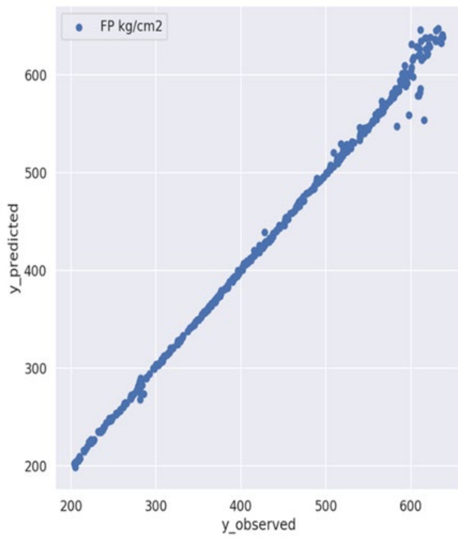

Figure 7. Regression plot of the true FP and predicted FP for test dataset for Ensemble model.

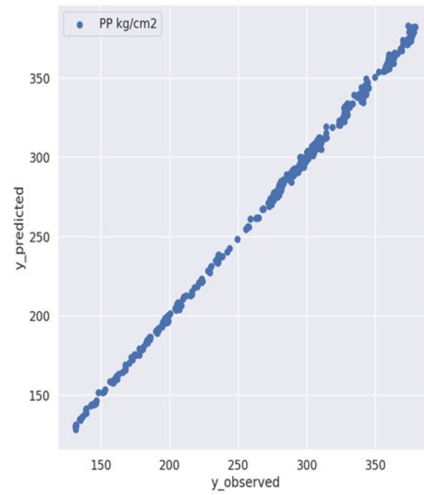

Figure 8. Regression plot of the true PP and predicted PP for test dataset for Ensemble model.

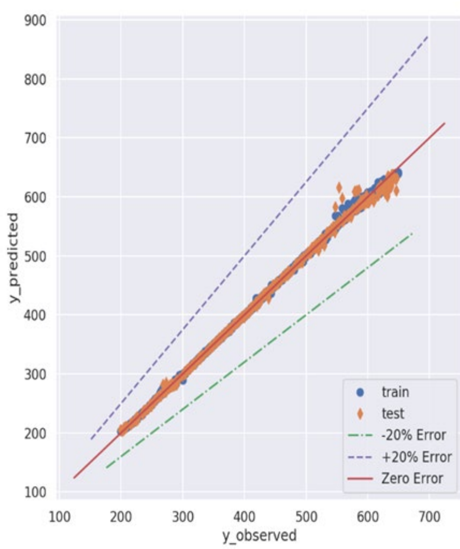

Figure 9. Regression plot of the true FP and predicted FP for train/test dataset for Ensemble model.

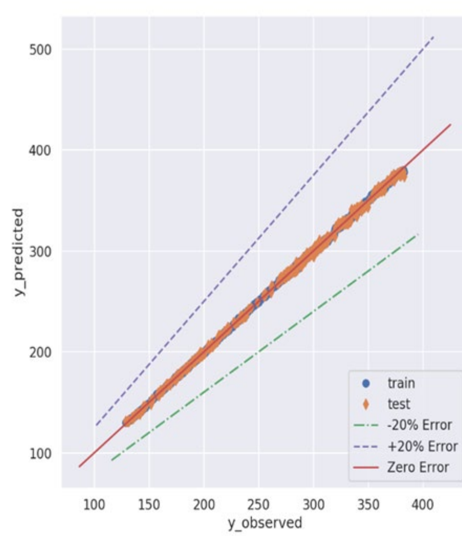

Figure 10. Regression plot of the true PP and predicted PP for train/test dataset for Ensemble model.

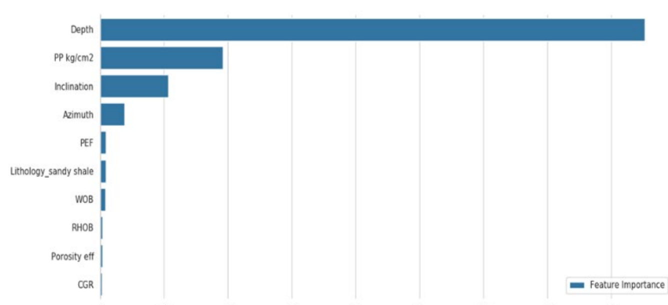

Figure 13. Features contribution to the fracture pressure prediction

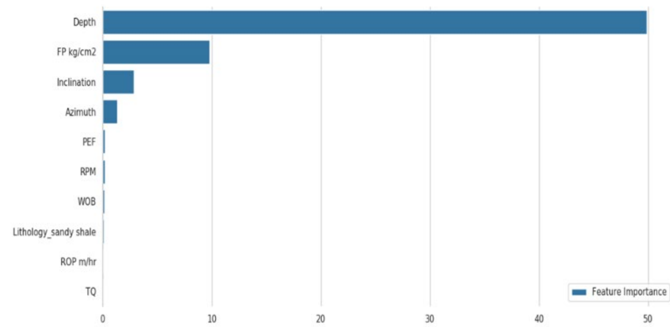

Figure 14. Features contribution to the pore pressure prediction.

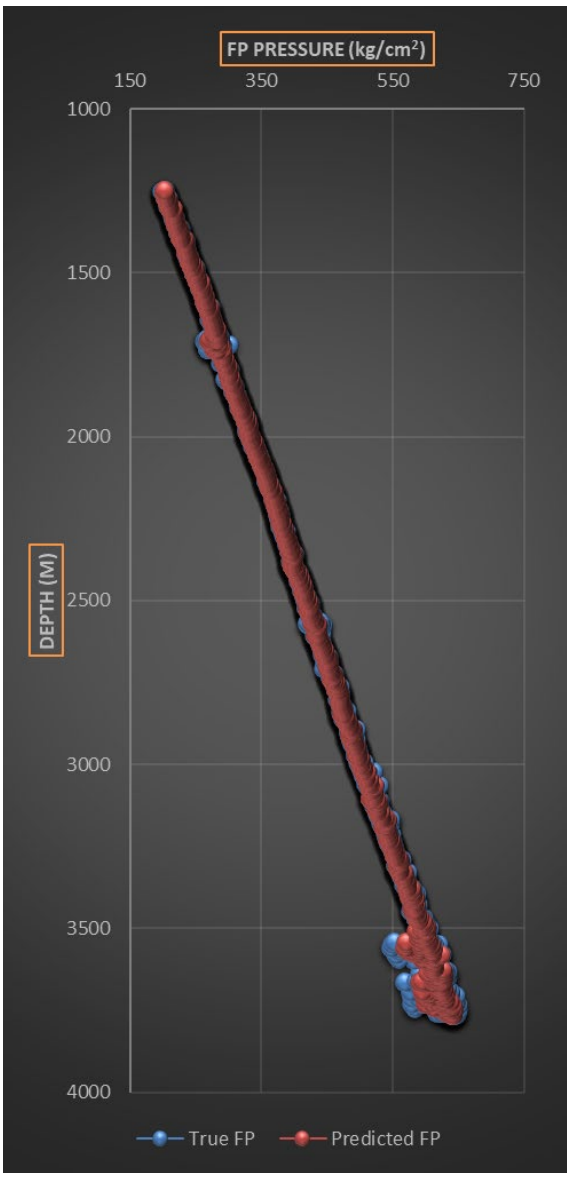

Figure 11. FP predicted vs. FP true for the developed model.

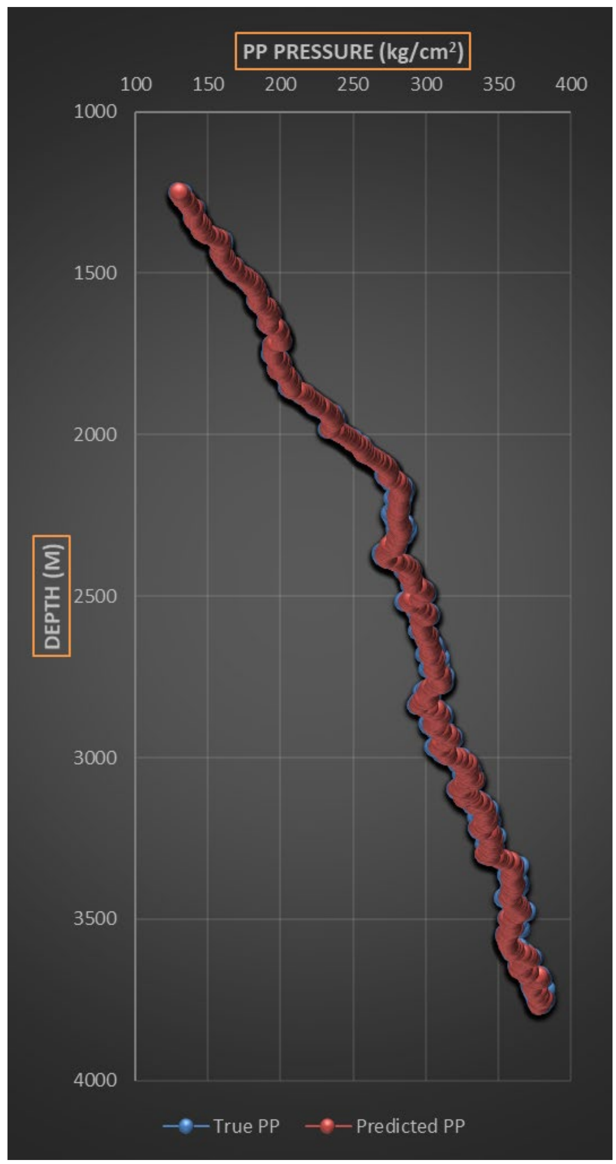

Figure 12. PP predicted vs. PP true for the developed model.

7. Conclusions

This work shows a model based on the machine learning techniques for predicting pore and fracture pressures that is a regression ensemble (mean) model which shows a blind test accuracy of 0.999 for pore Pressure and 0.998 for fracture Pressure from only mud logs and surface drilling parameters that are easily available.

The ensemble model is very accurate, comprehensive, low bias, has more stable performance and the highest test accuracy when predicting on new data points that it has not yet seen before.

There is no need to depend on the empirical models ,that are in general a field-specific, or any additional tests any more like (FIT, LOT, DST, RFT) that are costly, consume a lot of time and effort, have limitations and restrictions themselves and can lead to severe unexpected problems.

Nomenclature

ppg	*Pound per gallon*
psi	*Pound per square inch*
σ_h	*The minimum horizontal in-situ stress*
σ_H	*The maximum horizontal in-situ stress*
σ_V	*The vertical stress*
σ_1	*The maximum principal stress*
σ_2	*The intermediate principal stress*
σ_3	*The minimum principal stress*
TVD	*True vertical depth*
D	*Depth*
Min	*Minimum*
Max	*Maximum*
PP	*Pore pressure*
P_f	*Formation pressure*
P_n	*Normal pore pressure*
FP	*Fracture pressure*
FG	*Fracture gradient*
ρ_f	*Fluid density*
σ_{ov}	*Overburden stress*
σ	*Matrix stress*
β	*Biot coefficient*
d_c	*D exponent*
d_{co}	*Observed D exponent*
d_{cn}	*Normal trendline D exponent*
RFT	*Repeat Formation Tester*
DST	*Drill Stem Test*
Δt_n	*Normal* Δt *value at the depth of interest*
Δt_o	*Observed* Δt *value at the depth of interest*
R_n	*Normal* R_{es} *value at the depth of interest*
R_{log}	*Observed* R_{es} *value at the depth of interest*
FIT	*Formation integrity test*
LOT	*Leak-off test*
XLOT	*Extended leak-off test*
K_i	*Matrix stress coefficient*
v	*Poisson's ratio*
v_s	*Shear wave velocity*
v_p	*Compressional wave velocity*
α	*Compressibility ratio of the porous to the bulk rock matrix*
FBG	*Formation breakdown gradient*
T_0	*Rock tensile strength*
FP_{avg}	*Average fracture pressure*
MAE	*Mean absolute error*

MAPE	*Mean absolute percentage error*
MSE	*Mean squared error*
MDAE	*Median absolute error*
MDAPE	*Median absolute percentage error*
R^2	*Correlation coefficient*

Acknowledgment

All gratitude to Suez University, the Faculty of Petroleum and Mining Engineering, and the staff of Petroleum Engineering department for supporting this research and providing an excellent research environment.

References

[1] Fooshee JS. The development of a pore pressure and fracture gradient prediction model for the Ewing Banks 910 area in the Gulf of Mexico. (2009). *LSU Master's Theses*. 3198. https://repository.lsu.edu/gradschool_theses/3198
[2] Hu L, Deng J, Zhu H, Lin H. A new pore pressure prediction method-back propagation artificial neural network. Electronic Journal of Geotechnical Engineering, 2013; 18: 4093-4107.
[3] Bandara MK, Al-Ameri NJ. Wellbore Instability Analysis to Determine the Failure Criteria for Deep Well/H Oilfield. Petroleum and Coal, 2024; 66(1): 08-121.
[4] Radwan A, Abudeif A. Pore and fracture pressure modeling using direct and indirect methods in Badri Field Gulf of Suez. Elsevier, 2019.
[5] Hossain ME, Al-Majed AA. Fundamentals of Sustainable Drilling Engineering. John Wiley & Sons, 2015.
[6] Kankanamge T. Pore Pressure and Fracture Pressure Modelling With- Offset Well Data and Its Application To-Surface Casing Design of a Developmet Well Deep Panuke Gas Pool Offshore Nova Scotia. Published online, 2013.
[7] Mouchet JP, Mitchell A. Abnormal Pressures While Drilling, 1989.
[8] Bourgoyne Jr AT, Chenevert ME, Millheim KK, Young Jr FS. Applied Drilling Engineering, second printing. Textb Ser SPE, Richardson, Texas, 1991.
[9] Terazaghi K. Theoretical soil mechanics. John Wiley Sons, 1965.
[10] Hottman CE, Johnson RK, Marquis GL, et al. Estimation of formation pressures from log-derived shale properties. pubs.geoscienceworld.org. Published online, 1965.
[11] Biot MA, Willis DG. The Elastic Coefficients of the Theory of Consolidation. J Appl Mech, 1957; 24(4): 594-601.
[12] Eaton BA. The Equation for Geopressure Prediction from Well Logs. Published online, 1975.
[13] Yoshida C, Ikeda S. An investigative study of recent technologies used for prediction, detection, and evaluation of abnormal formation pressure and fracture pressure in North and South. onepetro.org, 1996.
[14] Yoshida C, Ikeda S. An investigative study of recent technologies used for prediction, detection, and evaluation of abnormal formation pressure and fracture pressure in North and South. onepetro.org, 1996.
[15] Farsi M, Mohamadian N, Ghorbani H, et al. Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms. Nat Resour Res, 2021; 30(5): 3455-3481.
[16] Hazbeh O, Ahmadi Alvar M, Khezerloo-ye Aghdam S, Ghorbani H, Mohamadian N, Moghadasi J. Hybrid computing models to predict oil formation volume factor using multilayer perceptron algorithm. J Pet Min Eng., 2021; 23(1):17-30.
[17] Wang Y, Fu W. Formation pressure prediction based on hybrid genetic algorithm. Nat Comput Ser, 2010; 28(3): 21-42.
[18] Aliouane L, Ouadfeul SA, Boudella A. Pore Pressure prediction in shale gas reservoirs using neural network and fuzzy logic with an application to Barnett Shale. Geophys Res Abstr, 2015; 17: 2015-2723.
[19] Abidin MH. Pore pressure estimation using artificial neural network. Published online, 2014.
[20] Aliouane L, Ouadfeul SA, Boudella A. Pore Pressure prediction in shale gas reservoirs using neural network and fuzzy logic with an application to Barnett Shale. Geophys Res Abstr, 2015; 17: 2015-2723.
[21] Haris A, Sitorus RJ, Riyanto A. Pore pressure prediction using probabilistic neural network: Case study of South Sumatra Basin. IOP Conf Ser Earth Environ Sci, 2017; 62(1).

[22] Kiss A, Fruhwirth RK, Pongratz R, Maier R, Hofstätter H. Formation breakdown pressure prediction with artificial neural networks. Soc Pet Eng - SPE Int Hydraul Fract Technol Conf Exhib, 2018.
[23] Rashidi M, Asadi A. An Artificial Intelligence Approach in Estimation of Formation Pore Pressure by Critical Drilling Data. Published online, 2018.
[24] Karmakar M, Maiti S. Short term memory efficient pore pressure prediction via Bayesian neural networks at Bering Sea slope of IODP expedition 323. Measurement, 2019; 135: 852-868.
[25] Yu H, Chen G, Gu H. A machine learning methodology for multivariate pore-pressure prediction. Comput Geosci. 2020; 143: 104-548.
[26] Abdelaal A, Elkatatny S, Abdulraheem A. Data-Driven Modeling Approach for Pore Pressure Gradient Prediction while Drilling from Drilling Parameters. ACS Omega, 2021; 6(21): 13807-13816.
[27] Radwan AE, Wood DA, Radwan AA. Machine learning and data-driven prediction of pore pressure from geophysical logs: A case study for the Mangahewa gas field, New Zealand. J Rock Mech Geotech Eng, 2022.
[28] Jaeger J, Cook N, Zimmerman R. Fundamentals of Rock Mechanics., 2009.
[29] Sadiq T, Nashawi IS. Using Neural Networks for Prediction of Formation Fracture Gradient. Published online, 2000.
[30] Postler DP. Pressure integrity test interpretation. Proc Drill Conf, 1997:169-183.
[31] Hubbert MK, Willis DG. Mechanics Of Hydraulic Fracturing. Trans AIME, 1957; 210(01): 153-168.
[32] Matthews WR, Kelly J. How to predict formation pressure and fracture gradient from electric and sonic logs. Oil Gas, 1967.
[33] Pennebaker ES. Detection of abnormal-pressure formation from seismic field data. In: Drilling and Production Practice. OnePetro, 1968.
[34] Eaton BA. Fracture gradient prediction and its application in oilfield operations. SPE Repr Ser, 1969; (49): 88-95.
[35] Anderson RA, Serviecs S, Ingram DS, Services SW. Determining Fracture Pressure tiradlents From Well Logs. Soc Pet Eng, 1973.
[36] Daines SR. Prediction of Fracture Pressures for Wildcat Wells. Soc Pet Eng AIME, SPE, 1982.
[37] Haimson B, Fairhurst C. Initiation and Extension of Hydraulic Fractures in Rocks. Soc Pet Eng J, 1967; 7(03): 310-318.
[38] Zhang J. Pore pressure prediction from well logs: Methods, modifications, and new approaches. Earth-Science Rev, 2011; 108(1-2): 50-63.
[39] Zhang J, Wieseneck J. Challenges and surprises of abnormal pore pressures in shale gas formations. Proc - SPE Annu Tech Conf Exhib, 2011; 2: 908-916.
[40] Behesht Abad AR, Mousavi S, Mohamadian N, et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs. J Nat Gas Sci Eng, 2021; 95: 104210.
[41] Malallah A, Nashawi IS. Estimating the fracture gradient coefficient using neural networks for a field in the Middle East. J Pet Sci Eng, 2005; 49(3-4): 193-211.
[42] Keshavarzi R, Jahanbakhshi R, Rashidi M. Predicting Formation Fracture Gradient In Oil And Gas Wells: A Neural Network Approach. Published online, 2011.
[43] Abdulmalek Ahmed S, Mahmoud AA, Elkatatny S, Mahmoud M, Abdulraheem A. Prediction of pore and fracture pressures using support vector machine. Int Pet Technol Conf, 2019.
[44] Abdulmalek Ahmed S, Elkatatny S, Ali AZ, Abdulraheem A, Mahmoud M. Artificial neural network ANN approach to predict fracture pressure. SPE Middle East Oil Gas Show Conf MEOS, 2019.
[45] Ahmed A, Elkatatny S, Ali A. Fracture Pressure Prediction Using Surface Drilling Parameters by Artificial Intelligence Techniques. J Energy Resour Technol Trans ASME, 2021; 143(3).
[46] Chukwuemeka AO, Amede G, Alfazazi U. A Review of Wellbore Instability during Well Construction: Types, Causes, Prevention and Control. Petroleum and Coal, 2017; 59: 590-610.
[47] Quantum Reservoir Impact, LLC. SpeedWise Machine Learning. Published online, 2024.
[48] Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J, 2020; 20(1): 3-29.
[49] Sivananda M., Kumar DGK. Classification and Regression Based on Decision Tree Algorithm for Machine Learning. Interantional J Sci Res Eng Manag, 2024; 08(02): 1-13.
[50] Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing. PeerJ Comput Sci, 2017; 3(7): 127.
[51] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn, 2006; 63(1): 3-42.

[52] Cao Y, Miao QG, Liu JC, Gao L. Advance and Prospects of AdaBoost Algorithm. Acta Autom Sin, 2013; 39(6): 745-758.
[53] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev, 2021; 54(3): 1937-1967.
[54] Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 2022; 10: 99129-99149.

To whom correspondence should be addressed: Tarek M. Barghoth, Petroleum Engineering Department, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt, E-mail: Tarek.Barghoth@pme.suezuni.edu.eg