Petroleum and Coal

Article Open Access

Predicting Equivalent Circulating Density While Primary Cementing Job by Employing Machine Learning Techniques

Elfakeur Abidi Saad[1], *Aissa Abidi Saad*[1], *Salaheldin Elkatatny*[2], *Khalil Moulay Brahim*[1], *Mohamed Nadhir Abid*[3], *Hichem Horra*[4]

[1] *Department of Drilling and Rig mechanics, University of Ouargla, 30000 Ouargla, Algeria*
[2] *Department of Petroleum Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia*
[3] *Faculty of Technology, University of Hama Lakhdar, 39000 El-Oued, Algeria*
[4] *Department of Engineering, Integrated Well Services, Baker-Hughes, Al-khobar 31952, Saudi Arabia*

Received February 28, 2025; Accepted May 28, 2025

Abstract

Equivalent Circulating Density (ECD) is a key parameter in drilling and cementing operations. Previous research on ECD prediction has primarily focused on drilling rather than cementing. In practice, the miscalculation of ECD is more significant in terms of cost, leading to large losses and time consumption compared to drilling. This study aims to address this gap by applying machine learning techniques used in drilling to the primary cementing process. Due to the lack of available pressure logging tools during casing running and cementing, along with high downhole measurement costs, alternative approaches are necessary. Several regression models, including Decision Tree Regressor (DTR), Linear Regression (LR), Random Forest Regressor (RFR), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), K-Neighbor Regressor (KNN), Ridge Regressor (RR), Artificial Neural Network (ANN), and an ensemble ML method, were applied to predict ECD before cementing. A dataset of 1,036 simulation data points was used, with 70% for training, 15% for validation, and 15% for testing. Model evaluation through statistical and graphical analyses showed that the DT model achieved the highest accuracy (R^2 = 0.994) for testing, followed by the RF and Ensemble models.

Keywords: *Equivalent circulation density; Machine learning (ML); Primary cementing; Simulation data; Drilling operation.*

1. Introduction

Primary cementing plays an essential role in well integrity for saving and extending their lives by establishing a protective zone between the rock formation and the casing string, isolating the migration of fluids between the formation layers and surface. A successful cement job can be challenging due to its complexity during the job and the interaction between the casing, the cement slurry, and the formation heterogeneities and properties [1].

Equivalent circulating density (ECD) is a crucial parameter to calculate and predict before drilling and primary cementing operations. ECD is the typical main objective for determining the impact of friction pressure from fluid flow and total hydrostatic pressure from fluid density at depth [2-4]. An appropriate selection and optimization of ECD, particularly in narrow window exploration and deep horizontal wells, is vital to achieving an adequate isolation zone and reducing non-productive time (NPT) [5-7]. In contrast, mismanagement in predicting ECD can lead to various issues in both drilling and cementing, such as lost circulation, bad cementation, collapse of wellbore, kicks, stuck pipe or casing, and blowouts. These problems make it a challenge for drilling engineers to escape while preparing well programs and during operations since they pose a significant difficulty in well budgeting by costing millions of dollars each year for the drilling activity in non-productive time (NPT) and material costs [8-12]. ECD is a measured parameter affected by several factors, including fluid rheology characteristics (density, viscosity), annular friction pressure loss (AFPL), well bore geometry (hydraulic diameter), pumping rate, and downhole pressure [13-14]. ECD estimation can be conducted using traditional mathematical models and downhole measurement tools; however, in cementing, only mathematical models are typically used. In drilling, both methods are employed. First, for well planning, several mathematical correlation models are used with different calculation processes, a variety of input parameters, and fluid types. Analytical computer codes are frequently utilized for modeling, performance prediction, and control. Typically, complex differential equation solutions are solved through the use of complex algorithms [15-16]. Secondly, in real time, there are different downhole sensors that measure the ECD with accurate values [17-18]. The limitation of mathematical models is that some applications and missing input parameters have an impact on ECD prediction. Usually, these systems require a large amount of computing power and time to generate accurate projections. In the face of numerous correlations of ECD in previous studies, they are limited by inadequate or insufficient data or are only relevant to precise areas [19-21]. The limitation of using downhole sensors is due to their sensitivity and failures exposed to downhole conditions [17]. According to precedent limitations, a novel strategy has emerged. This new strategy is necessary for ensuring alignment with advancements in relevant technologies, processes, and scientific development. Artificial intelligence (AI) systems can identify key information patterns within a multidimensional information field [22]. Furthermore, many AI systems, like neural networks (NN) and machine learning (ML), are durable, noise-immune, and fault-tolerant. Artificial intelligence (AI) is a technique that employs powerful computers to analyze complicated algorithms by replicating how the human brain thinks [23-24]. In the oil and gas industry, which often deals with huge amounts of data, AI techniques offer significant advantages in data modeling, prediction, and management. Over the past few years, artificial intelligence techniques have been widely applied in the oil and gas industry. Table 1 represent literature review related to machine learning in petroleum engineering and ECD prediction.

Writers have made an effort to overcome the shortcomings and limitations of mathematical correlations, downhole measurement failure, and high cost by including drilling parameters through different AI techniques to predict ECD. Recent studies are summarized below:

Alkinani, Al-Hameedi [25] designed an ANN model featuring a single hidden layer with 12 neurons for predicting ECD. This model incorporated surface drilling parameters, including weight on bit (WOB) and drill pipe revolution (RPM). Additionally, it considered mud hydraulics and properties such as flow rate (Q), total flow area (TFA), mud weight (MW), yield point (YP) and plastic viscosity (PV).

Abdelgawad, Elzenary [26]proposed a model for predicting equivalent circulating density (ECD) by using ANN an ANFIS. The first model with ANN used a hidden layer and 20 neurons, while the second model, ANFIS, employed 5 Gaussian membership functions in the inputs and a linear type for the outputs. They used rate of penetration (ROP), drill pipe pressure and MW as prediction features.

Rahmati, Tatar [27] utilized a Radial Basis Function (RBF) for robust prediction performance, achieving an AAPE of 0.22 and an R^2 of 0.98, which included input variables such as mud density and type, pressure and temperature.

Kandil, Khaled [28] focused on enhancing efficiency and reducing human error. The authors employed three ML algorithms—ANN with a Levenberg-Marquardt backpropagation algorithm, PAR, and KNN—to predict ECD. These models leveraged 14 critical operational parameters obtained from downhole sensors. The authors analysed 4663 data points, with 80-85% for training and validation. The ANN model demonstrated remarkable accuracy, achieving an R2 value close to 0.999.

Ahmadi, Shadizadeh [29] developed a robust model for predicting mud density at using a hybrid of particle swarm optimization (PSO) and (ANN). Additionally, two competitive machine learning models, FIS and GA-FIS, were tested. Statistical analysis revealed that the PSO-ANN model outperformed other methods, making it a reliable tool for predicting drilling fluid density in HPHT conditions using data from previous literature

Alsaihati, Elkatatny [30] introduced artificial intelligent machines, incorporating support vector machines (SVM), random forests (RF), and functional networks (FN), for the estimation of ECD in real-time in horizontal wells using field data. The RF model demonstrated superior performance with high accuracy. They used seven drilling parameters as inputs to predict ECD, including pumping rate, ROP, SPP, WOB, RPM, torque, and hook load.

Gamal, Abdelaal [31] utilised ML techniques, specifically ANNs and ANFISs, to predict ECD with high accuracy using drilling data. The ANN model achieved an average absolute percentage error (AAPE) of 0.3% and a coefficient of correlation (R) above 0.98, while ANFIS recorded an AAPE of 0.7% and an R of 0.96. Additionally, a new equation for real-time ECD determination was introduced. Focused on six drilling parameters for ECD prediction, such as ROP, flow rate (GPM), drill string speed (RPM), standpipe pressure (SPP), WOB, and torque (T).

Al-Rubaii, Al-Shargabi [32] integrated a combination of mud properties and drilling parameters into the ECD prediction model. They used flow rate (GPM), ROP, drill string rotation (RPM), standpipe pressure (SPP), mud weight (MW), low shear yield point (LSYP), yield point (YP), and plastic viscosity (PV) as inputs. Additionally, they considered azimuth, borehole angles, modified hole geometry factors, and average cuttings concentration in an annulus, among other factors.

Employed mathematical models to develop hydraulic programs, which are crucial for hydraulic calculations during cementing operations. The programs may utilize different rheological models, such as Bingham plastic (BPM), Power-law (PLM), or Herschel–Bulkley (HBM), as the foundation for hydraulic calculations. Metwally [33] proposed a model that improves the prediction accuracy of ECD using Herschel-Bulkley model for water-based mud (WBM). Importantly, ECD predictions for formulated WBM match well with those of OBM which signifies the strength and dependability of the proposed methodology but different data inputs are needed for each rheological model, and these are usually acquired by means of extensive laboratory testing. The software then uses these data to compute drilling fluid hydraulics and related parameters like ECD [34]. A tree-based ensemble method is used by employing XGBoost methodology to predict ECD. With R^2 and RMSE for the testing/blind data set of 0.989 and 0.023, respectively, the findings demonstrated a significant prediction capacity [35].

The previous research has predominantly concentrated on predicting equivalent circulating density (ECD) in drilling operations rather than in cementing. In practice, the equivalent circulating density (ECD) parameter is more critical during cementing operations. Consequently, it is imperative to prioritize and focus on ECD management throughout the cementing process. Whereas, losses incurred during cementing operations are significantly more costly and time-consuming compared to those encountered during drilling. This is primarily due to the need for multiple additional procedures, such as pulling out and replacing the casing, running the drill string, sealing the losses with lost circulation materials (LCM) or cement plugs, waiting on cement (WOC), and eventually cementing. These steps involve pulling out the drill string, running a new casing, and cementing again, all of which contribute to increased operational expenses and extended non-productive time. However, the machine learning techniques developed for drilling applications can be adapted for cementing. The limited availability of pressure logging tools for use during the running and cementing of casing strings, combined with the high costs and potential failures associated with downhole measurements in challenging hole conditions, calls for alternative approaches. These alternatives should aim to overcome these limitations while minimizing effort and cost. In this study, various regression models of ML, including Decision Tree Regressor (DTR), Linear Regression (LR), Random Forest Regressor (RFR), K-Neighbor Regressor (KNN), Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Ridge Regressor (RR), and Artificial neural network (ANN), were employed to forecast ECD before cementing, yielding promising results. The performance of these regression models was evaluated and compared with existing methodologies, demonstrating their effectiveness in ECD prediction. The software user has an improved ability to apply constructive design modifications to affect the real-time operation's result. With the software, the user may monitor and evaluate certain variables to make informed judgments based on iterative simulation situations, whether it is ECD management or displacement efficiency enhancement.

Table 1. Literature review related to machine learning in petroleum engineering and ECD prediction.

Topic	References	Details
ML in Petroleum Engineering	**[36]**	Machine learning techniques, including ANNs, have been applied since the early 1990s to tasks like identifying seismic reflections, showcasing potential for improved efficiency and accuracy.
ML for Drilling Engineering	**[37-39]**	ML techniques like ANNs have been used since 1990 for assessing drill bit wear and predicting ROP. Sprunger *et al.* (2022) highlighted ML's advantages in hydraulic fracturing operations.
Definition & Importance of ECD	**[40, 41]**	ECD represents the hydrostatic pressure exerted by drilling mud under dynamic conditions, crucial for maintaining wellbore stability and managing formation and fracture pressures.
Role of Drilling Mud & ECD	**[13, 16, 26, 42]**	Discusses the impact of mud density on drilling conditions, emphasizing ECD optimization to prevent issues like poor drilling rates and increased costs.
Downhole ECD Measurements	**[17, 18]**	Erge et al. and Rommetveit *et al.* highlighted the reliability and precision of downhole ECD measurements but noted high costs and operational constraints.
Limitations of Mathematical Models for ECD	**[10, 15, 43-46]**	Traditional models often overlook crucial factors, leading to inaccurate ECD predictions. ANNs offer a more advanced approach by considering a broader range of input variables.
ANNs for ECD Prediction	**[47-49]**	Baranthol *et al.* conducted oil field measurements and validated their findings. Ahmed et al. created an AI models to estimate mud density and forecast the impact of drill-string rotation on ECD.
Advanced AI Techniques for ECD	**[50-52]**	Elzenary *et al.* used ANFIS and ANN to develop ECD predictive models. Ahmadi employed PSO-ANFIS and LLSVM algorithms for accurate ECD predictions.
High Accuracy ECD Models	**[25, 27]**	Rahmati and Tatar used a radial basis function model for ECD prediction, achieving high accuracy. Alkinani et al. used an ANN model with various drilling parameters for ECD prediction.
ML Models in Drilling	**[30, 53]**	Alsaihati *et al.* developed models using SVM, RF, and FN with high accuracy. Gamal et al. used ANNs and ANFIS to predict ECD with high R2 values.

2. Methodology

2.1. Data collection

Based on laboratory and field data, data was created for this study to model a primary cementing job. This made it possible to create datasets that closely mimic data from the actual world. Simulated data can also be used by ML to create accurate models and forecasts.

The data gathered for this investigation was from a simulation of the primary cementing job of a 13 3/8-inch section in Hassi Messaoud field, Algeria. The strategy that was followed while conducting this study is depicted in Figure 2 A total of 1036 points were obtained from five vertical wells. The cementing parameters that were gathered and used as model inputs came from the simulation reports of the cementing program that represent fluid weight (mud, spacer, lead slurry, and tail slurry) in g/cm^3, flow rate in L/s, fluid volume (V) in m^3, pump pressure (SPP) in KPa, yield point (Yp) in Pa, and plastic viscosity (PV) in cp. ECD data were collected from the same reports and used for the model output estimation. These parameters were selected based on their influence on equivalent circulating density (ECD).

2.2. Statistical analysis of parameters

The dataset consists of 1036 samples with seven parameters related to cementing operations. The extended statistical summary of the dataset reveals key insights into the distribution and variability of each parameter. Table 2 show the input and output statistical parameters. The variation in fluid density is from 1.25 to approximately 1.9 g/cm^3, since the variability lies low and finally flow rate varies between nearly about only 8.3 to 88.33 L/s indicating a broader distribution fluid volume range from 01 to 296 m^3, and stand pipe pressure (SPP) shows some very wide range values like 827 to 8673 KPa which indicates outliers. Plastic viscosity (PV) distributed with a heavy-tail, ranges between 11-78 cP. More stable, 4.79 to 11.01 Pa of yield point (YP). The target set, equivalent circulating density (ECD) ranges reasonably within the scale of 1.25 to 1.65 g/cm^3.

Table 1. Statistical analysis parameters.

Statistical parameters	Flow rate (L/s)	Fluid volume (m^3)	SPP (KPa)	Fluid density (g/cm^3)	PV (cP)	YP (Pa)	ECD (g/cm^3)
mean	56.89	137.19	2282.23	1.41	36.56	9.11	1.35
std	20.66	87.04	1709.14	0.24	31.07	2.09	0.08
min	8,30	1.00	827.37	1.25	11.00	4.79	1.25
25%	50.47	62.21	1709.14	1.25	11.00	6.70	1.30
50%	56.78	124.21	2282.23	1.30	14.00	9.10	1.33
75%	63.09	215.03	1709.14	1.35	76.00	11.01	1.37
max	88.33	296.00	8673.60	1.90	78.00	11.01	1.65
Kurtosis	-0.50	1.50	-0.30	2.10	3.50	-1.00	1.20
Skewness	0.90	0.60	0.70	1.10	1.40	-0.30	1.00

2.3. Heatmap correlation

The correlation heatmap shows the relationships between various parameters in the dataset as shown in Figure 1.

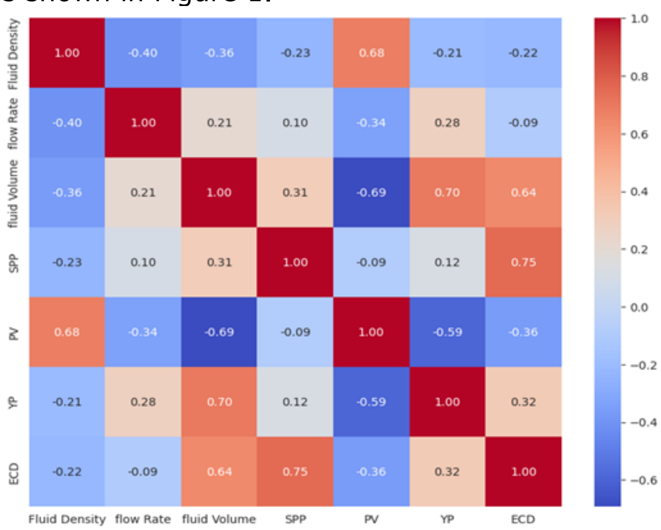

Figure. 1. Heatmap correlations of input and output parameters.

Fluid density has a strong positive correlation with PV (0.68) and a moderate negative correlation with flow rate (-0.40). Fluid volume has a strong positive correlation with YP (0.70) and a moderate positive correlation with SPP (0.31). SPP has a strong positive correlation with ECD (0.75). PV has a moderately negative correlation with fluid volume (-0.69) and YP (-0.59). ECD shows a strong positive correlation with both SPP (0.75) and fluid volume (0.64). These correlations indicate which parameters are most closely related, with strong positive values suggesting a direct relationship and negative values indicating an inverse relationship.

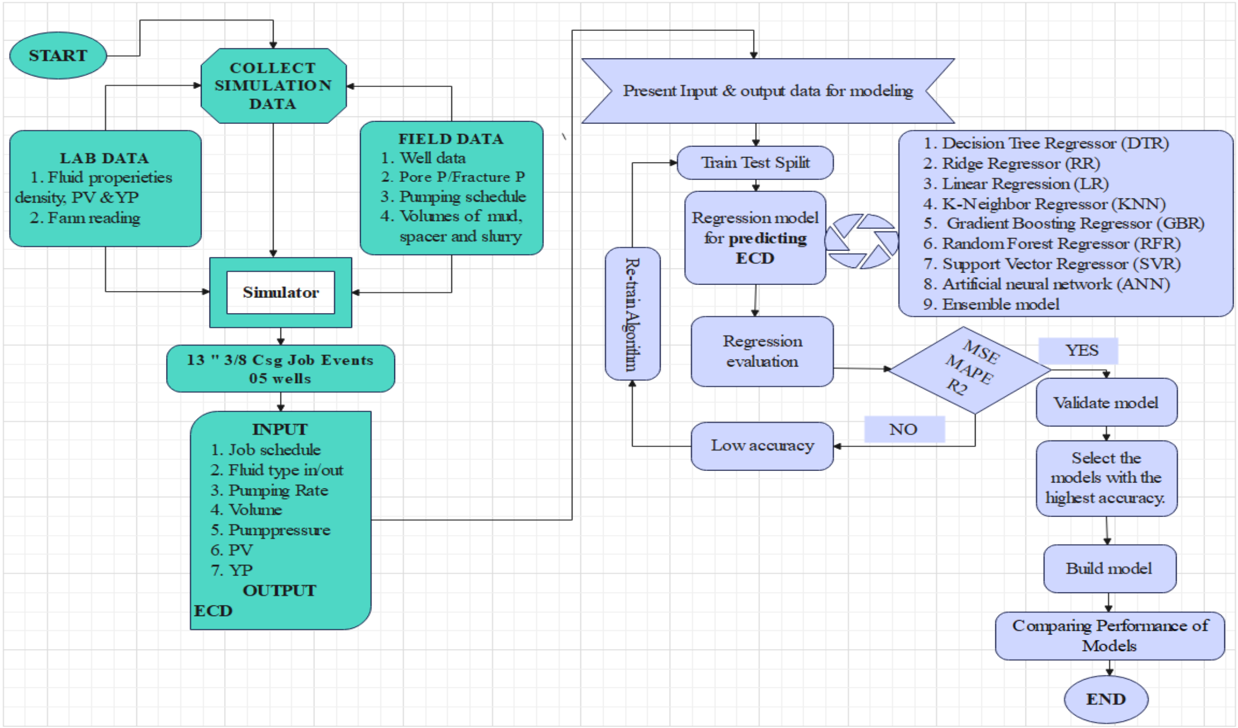

Figure 2. Flow chart processing for predicting ECD with ML models.

2.4. Data modelling

With the use of the simulation, researchers will be able to examine how a system behaves by building a model that consists of a structure and the rules that regulate it to generate different outputs. For AI systems to be trained, simulated data is frequently the only choice available, particularly in situations where obtaining real data would be difficult, costly, or take a long time. The process of creating ECD models begins with the collection of data to produce model input parameters of high quality implemented in python. Next, the ML model is trained, the model parameters are optimized using the trained algorithm, and the model results are tested for accuracy. If the accuracy is low, a re-training process is conducted to obtain the ideal model parameters for high accuracy performance for the ECD prediction. The dataset is divided into three sections: 70% for training, 15% for testing, and 15% for validation, with ECD selected as the target variable.

3. Regression models

Several regression models are employed to predict ECD from the dataset, have been summarized in Table 3 definition, use, and objective for each technique. The performance of each model is assessed using metrics such as R^2, MSE, and RMSE error were calculated by Equations 1,2 and 3.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(ECD_i - ECD_{ipredicted})^2 \quad (1)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|ECD_i - ECD_{ipredicted}|}{|ECD_i|} \quad (2)$$

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(ECD_i - ECD_{ipredicted})^2}{(ECD_i - \overline{ECD})^2} \quad (3)$$

Table 3. Regression models used in the study.

Technique	Definition	Use	Objective
Linear regression (LR)	Assumes a linear relationship between input variables and a single output variable.	Predictive modeling and forecasting in various domains like economics, engineering.	Minimize the sum of squared differences between observed and predicted values to achieve the best-fit linear line.
Ridge regression (RR)	Type of linear regression with a regularization term (L2 penalty) to prevent overfitting.	Cases with multicollinearity, providing more reliable predictions by penalizing large coefficients.	Balance the fit of the model with complexity, minimizing both the sum of squared errors and the magnitude of coefficients.
K-Neighbors regressor (KNR)	Predicts the output for a data point by averaging the outputs of its K nearest neighbors.	Regression problems with complex relationships between input and output.	Make predictions based on the similarity of new data points to known data points, capturing local patterns in the data.
Decision Tree Regressor (DTR)	Builds a tree structure model to make predictions by splitting data into subsets based on feature values.	Regression and classification tasks with non-linear relationships between input features and output.	Create a model that predicts the output by learning simple decision rules inferred from data features.
Random Forest Regressor (RFR)	Ensemble method using multiple decision trees to make predictions by averaging their outputs.	Domains requiring improved predictive accuracy and control overfitting.	Enhance model accuracy and robustness by combining predictions of multiple trees trained on different subsets of data.
Gradient Boosting Regressor (GBR)	Sequentially builds an ensemble of trees where each corrects the errors of the previous one.	Machine learning competitions and applications due to high predictive accuracy.	Minimize prediction error by iteratively adding models that correct residuals of previous models using gradient descent.
Support Vector Regression (SVR)	Finds a hyperplane in high-dimensional space that best fits the data, minimizing error within a threshold.	Non-linear relationships and high-dimensional spaces.	Maximize the margin around the hyperplane while keeping prediction error within a defined tolerance level.
Artificial Neural Networks (ANN)	A computational model inspired by the human brain, consisting of interconnected units (neurons) that process information.	Complex nonlinear relationships, pattern recognition, and time series prediction.	Learn intricate patterns in the data through multiple layers of neurons, optimizing the network to minimize prediction error.
Multi-Layer Perceptron (MLP)	A type of ANN with one or more hidden layers between input and output layers, using backpropagation for training.	Deep learning tasks requiring feature learning and complex pattern recognition.	Learn complex relationships by adjusting neuron weights through backpropagation, minimizing prediction error across multiple hidden layers.
XGBoost	An optimized version of gradient boosting that uses parallel processing and regularization to improve speed and accuracy.	Widely used in Kaggle competitions and real-world machine learning applications.	Reduce prediction errors and training time through parallelized tree boosting and regularization.
Ensemble model	Combines predictions from multiple base models (e.g., MLP, GBR, DTR, XGBoost) and trains a meta-model (e.g., Ridge) to improve accuracy.	Applications requiring high accuracy by leveraging multiple models.	Improve predictive accuracy by combining outputs of different models, allowing the meta-model to learn from their individual strengths and weaknesses.

3.1. Building ML models

A straightforward and easily comprehensible technique, linear regression maximizes the fit by reducing the squared variations between the observed and predicted values. This is reinforced by Ridge regression, which includes a regularization term to avoid overfitting, particularly in datasets with intricate features. Recursively dividing the data according to the values of the input features is how Decision Trees, a non-linear method, forecast results. Artificial Neural Networks (ANNs) are very good at identifying intricate non-linear patterns in the data because of their linked layers of neurons. Even though they are quite successful, K-Nearest Neighbors (KNN), Random Forest, and Gradient Boosting are more difficult to understand because they rely on more complex mathematical frameworks that emphasize data point similarity or repeated error correction.

An ensemble model was used, especially employing a stacking strategy, to further improve forecast accuracy. In order to improve the final forecast, this entails merging predictions from many base models, including MLP, Gradient Boosting, Decision Trees, and XGBoost, with a meta-model (Ridge regression). By combining the advantages of many models, this approach produces more reliable and accurate ECD forecasts

3.2. Building linear model

The function provides a mathematical equation describing the model for linear models, particularly Ridge regression and linear regression. In order to perform this, the fitted model's coefficients and intercept have to be extracted, and a usable equation string needs to be developed.

3.3. Building ANN and ensemble models

The first step in the approach utilized in this work to develop the model via ANN is data preparation, where the dataset is split into sets for training, testing and validation. To handle missing values and standardize features, a preprocessing pipeline is created and applied uniformly to all datasets. The first model architecture makes use of an Artificial Neural Network (ANN) with the scikit-learn MLPRegressor. Using GridSearchCV, a thorough hyperparameter tuning procedure is carried out to maximize the model's performance. Through the use of a 5-fold cross-validation approach, 2,160 model fits are produced for each of the 432 possible parameter combinations. The methodology applies ensemble learning techniques to improve prediction power beyond baseline model tuning. The modeling technique incorporates the Gradient Boosting and Ridge Regression algorithms. As the last phase of the model design, a StackingRegressor is constructed, building upon these ensemble approaches. To provide a more reliable and accurate prediction model, this stacked ensemble integrates many basic models, such as Multi-Layer Perceptron (MLP), Gradient Boosting, Decision Tree, and XGBoost.

4. Results and discussion

4.1. Model training and validation

The dataset was divided into three categories: 70% for training, 15% for validation, and 15% for testing. Each model's performance was evaluated using R-squared (R2), mean square error (MSE), and mean absolute percentage error (MAPE) metrics.

Table 4 compares various machine learning models on a dataset, evaluating their performance using metrics. Linear regression shows moderate performance, while K-Nearest Neighbors (KNN) shows high accuracy.

The decision tree shows near-perfect performance, with an R^2 of 0.993. Random Forest achieves high accuracy, with a validation R^2 of 0.990 and a low MSE. Gradient boosting performs well, with R^2 values close to 0.98 and a low MSE. Support Vector Regression (SVR) shows poorer performance, with lower R^2 values and higher MSE values. Ridge regression is similar to linear Regression, with R^2 values around 0.81-0.87 and MSE values around 0.0011. Artificial Neural Networks (ANN) show strong performance, with a validation R^2 of 0.897 and a low MSE. In summary, decision trees and random forest models show the highest accuracy, but decision trees show signs of overfitting. ANN also performs well with lower MAPE values, indicating robust predictive capability. Calculated and predicted plots for models (training, validation, and testing) of LR, RFR and DTR are illustrated in In Figure 4 and Figure 5.

Table 4. Summary of model's performance for training.

Model	Dataset	MSE	R^2	MAPE
Linear regression	Training	0.001130	0.810069	0.021927
Linear regression	Validation	0.000989	0.871384	0.019876
Linear regression	Test	0.001147	0.823291	0.021742
K-Nearest neighbors	Training	0.000103	0.982627	0.003625
K-Nearest neighbors	Validation	0.000156	0.979689	0.005162
K-Nearest neighbors	Test	0.000141	0.978195	0.004492
Decision tree	Training	0.000001	0.999909	0.000082
Decision tree	Validation	0.000126	0.983577	0.003052
Decision tree	Test	0.000039	0.993962	0.002186
Random forest	Training	0.000018	0.996982	0.001402
Random forest	Validation	0.000076	0.990055	0.003256
Random forest	Test	0.000049	0.992459	0.002977
Gradient boosting	Training	0.000076	0.98718	0.004323
Gradient boosting	Validation	0.000110	0.985709	0.005474
Gradient boosting	Test	0.000099	0.984669	0.004959
Support vector r	Training	0.003437	0.422468	0.037043
Support vector r	Validation	0.003376	0.561007	0.036823
Support vector r	Test	0.003411	0.47427	0.036368
Ridge regression	Training	0.001131	0.810054	0.021953
Ridge regression	Validation	0.000991	0.871079	0.019935
Ridge regression	Test	0.001149	0.822933	0.021798
ANN (MLP)	Training	0.000783	0.874319	0.021927
ANN (MLP)	Validation	0.000743	0.897772	0.019876
ANN (MLP)	Test	0.000824	0.854801	0.021742
Ensemble model	Training	0.000058	0.991196	0.004466
Ensemble model	Validation	0.000101	0.980891	0.005270
Ensemble model	Test	0.000083	0.984474	0.004976

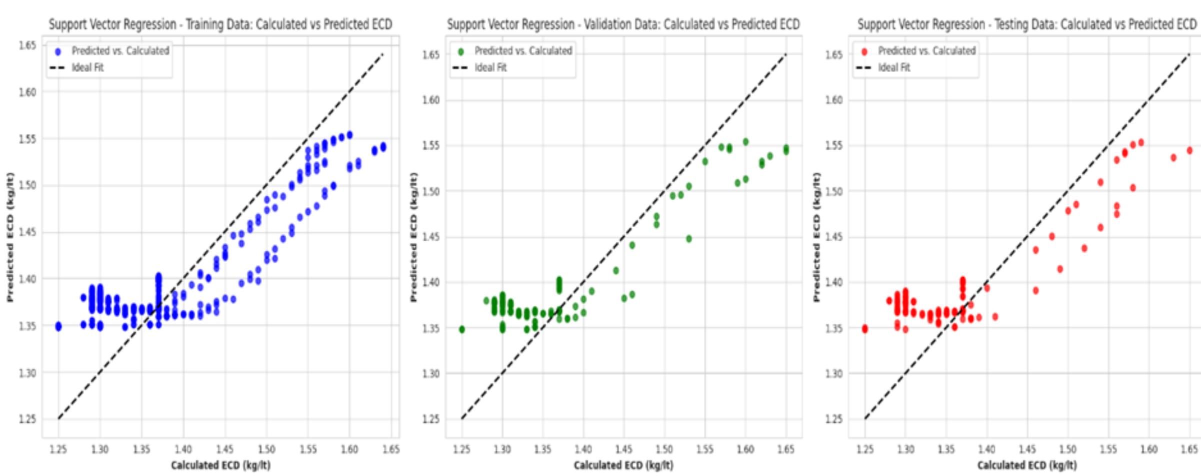

Figure 3. Calculated and predicted plots using Support Vector Regressor model (training, validation and testing).

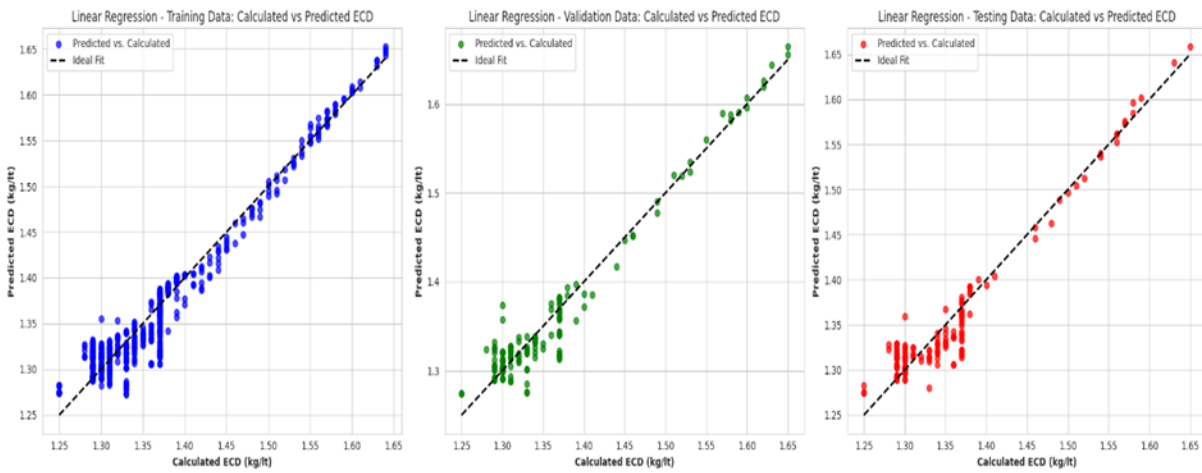

Figure 4. Calculated and predicted plots using linear regression model (training, validation and testing).

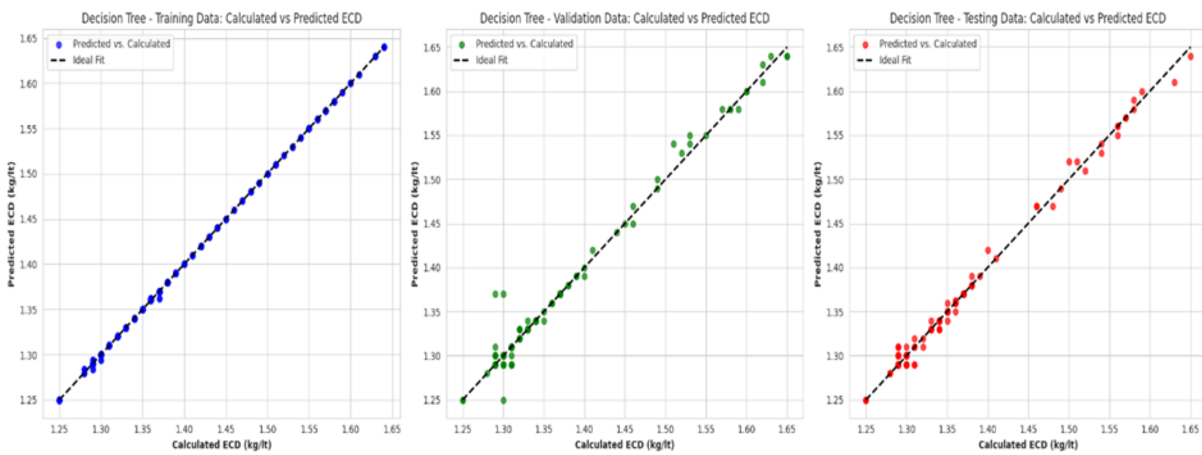

Figure 5. Calculated and predicted plots using Decision Tree model (training, validation and testing).

4.2. Model performance

Table 5 presents a comprehensive comparison of regression methods on a dataset using metrics.

Table 5. Model performance metrics.

Regression method	MSE	R2	MAPE
Linear regression	0.001147	0.823291	0.021742
K-nearest neighbors	0.000141	0.978195	0.004492
Decision tree	0.000038	0.994159	0.002233
Random forest	0.000055	0.991578	0.003031
Gradient boosting	0.000099	0.984696	0.004961
Support vector regression	0.003411	0.474270	0.036368
Ridge regression	0.001149	0.822933	0.021798
Artificial neural network (MLP)	0.000824	0.854801	0.017628
Ensemble model	0.000836	0.984474	0.004976

Linear regression shows moderate performance, with an MSE of 0.0011, an R^2 of 0.823, and a MAPE of 0.0217. K-Nearest Neighbors (KNN) shows excellent performance, achieving a low MSE of 0.00014, a high R^2 of 0.978, and a low MAPE of 0.00449. Decision tree regression has the highest accuracy, with an MSE of 0.000038, an R^2 of 0.994, and a MAPE of 0.00223. Random forest also performs well, with an MSE of 0.00005, an R^2 of 0.991, and a MAPE of 0.0030. Gradient boosting shows high predictive accuracy and a good fit, although not as high as a decision tree or random forest. Support vector regression (SVR) has the poorest performance, with an MSE of 0.003411, a low R^2 of 0.474, and a high MAPE of 0.036. Ridge regression presents similar performance to linear regression, but regularization does not significantly enhance predictive accuracy. An artificial neural network (ANN) shows strong performance, providing a good balance of error and fit and demonstrating its capability to handle complex relationships in the data.

4.3. Ensemble model results

The model performance improved over time using our machine learning approach. On the validation set, the first Artificial Neural Network (ANN) model with MLPRegressor had a R^2 score of 0.85. This increased to 0.88 after using GridSearchCV to adjust the hyperparameters. With the addition of ensemble learning strategies like Ridge Regression and Gradient Boosting, performance was further improved to an R^2 score of 0.94. Using a StackingRegressor—which incorporated MLP, Gradient Boosting, Decision Tree, and XGBoost as basic models—and achieving an R^2 score of 0.97 was a major advancement. An R^2 score of 0.984 was obtained in the last optimization stage utilizing the most advanced hyperparameter optimization framework Optuna as shown in Figure 6 and Figure 7.We assessed performance using a variety of measures during this process, throughout this process, we evaluated performance using multiple metrics including MSE, R^2, and MAPE, with each modeling step demonstrated substantial gains in each of these criteria.

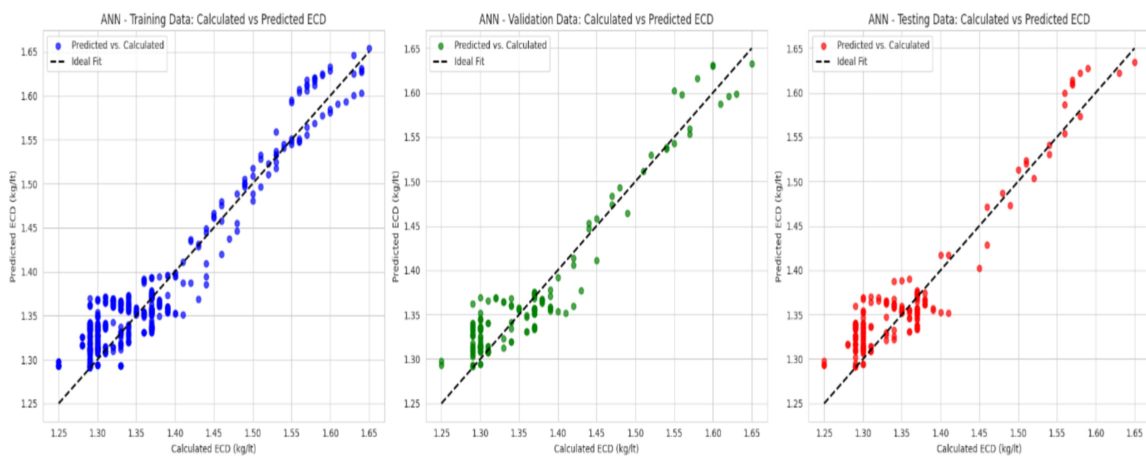

Figure 6. Calculated and predicted plots using Artificial neural network model (training, validation and testing).

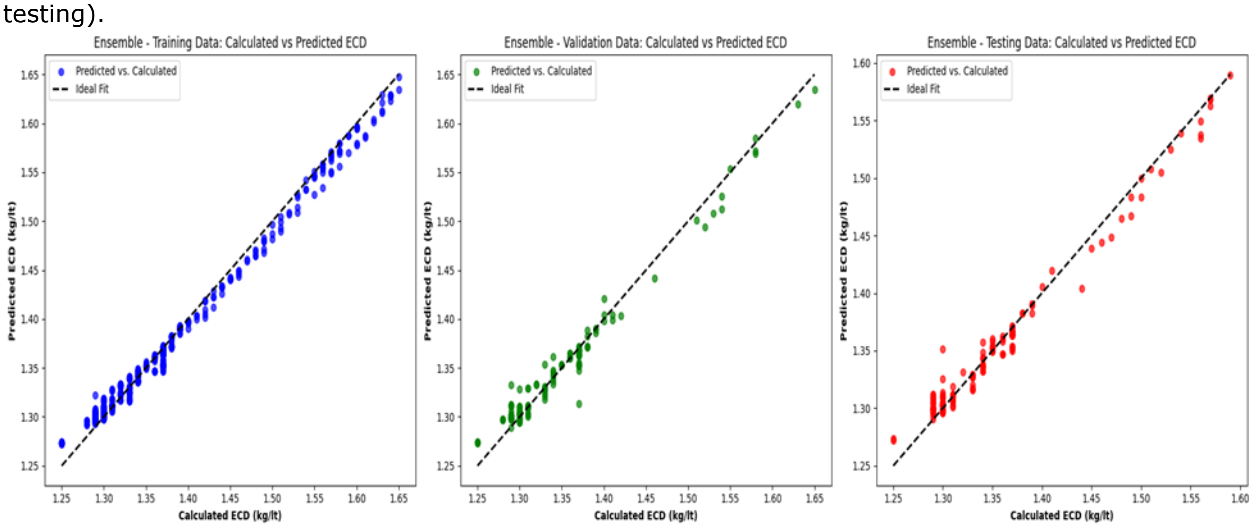

Figure 7. Calculated and predicted plots using Ensemble model (training, validation and testing).

5. Conclusion

The ECD was forecasted using machine learning algorithms based on fluid weights (mud, spacer, slurry), flow rate, fluid volume, SPP, YP, and PV. The models were developed using a dataset of 1,036 simulated data points, with 70% of the data used for training and the remainder for validation and testing.

This study utilized a range of machine learning models, including DTR, RR, LR, KNN, GBR, RFR, SVR, ANN, and an Ensemble approach, demonstrating that these techniques can accurately predict ECD. The Decision Tree model achieved the highest accuracy with an R^2 value of 0.994 for the testing dataset, with Random Forest, K-Nearest Neighbors, and the ensemble model also showing strong performance. The ensemble approach combining MLP, Gradient Boosting, Decision Tree, and XGBoost provided validation accuracies of 94–97% (R^2), showcasing robust prediction capabilities. The study's findings highlighted good agreement between predicted and calculated ECD values, confirming the models' reliability. This work demonstrated the utility of simulated data for building accurate predictive models in the absence of real-time downhole data during cementing operations. The study was conducted under controlled conditions focusing on a vertical well profile and a 13 3/8-inch casing section, ensuring a detailed preliminary evaluation of the models. The models developed can serve as an initial point for further research and may aid in real-time ECD management, offering users an enhanced ability to monitor and make informed decisions for operational improvements. Future research is recommended to expand the applicability of these models by incorporating real-time downhole data, addressing abnormal conditions, and considering various well profiles to improve generalizability.

List of symbols

NPT	*Non-productive time*
AI	*Artificial intelligence*
ML	*Machine learning*
DTR	*Decision tree regressor*
RR	*Ridge regressor*
LR	*Linear regression*
KNN	*K-nearest neighbors*
GBR	*Gradient boosting regressor*
SVR	*Support vector regressor*
RF	*Random forest*
ANN	*Artificial neural network*
MLP	*Multilayer perceptron regressor*
ANFIS	*Adaptive network-based fuzzy interference system*
PSO	*Particle swarm optimization*
SVM	*Support vector machine*
LLSVM	*Least square support vector machine*
GA	*Genetic algorithm*
FIS	*Fuzzy inference system*
FN	*Functional networks*
RBF	*Radial basis function*
PAR	*Passive aggressive regressor*
R2	*Coefficient of determination*
MSE	*Mean squared error*
RMSE	*Root mean squared error*
AAPE	*Average absolute percentage error*
MAPE	*Mean absolute percentage error*
MW	*Mud weight*
WOB	*Weight on bit*
RPM	*Rotating speed in revolutions per minute*
ROP	*Rate of penetration*
GPM	*Gallon per minute*
SPP	*Standpipe pressure*
LSYP	*Low shear yield point*

TFA	*Total flow area*
PV	*Plastic viscosity*
YP	*Yield point*
T	*Torque*
BPM	*Bingham plastic model*
PLM	*Power-law model*
GBM	*Herschel–Bulkley model*
LCM	*Lost circulation materials*
WOC	*Wait on cement*
HPHT	*High Pressure High Temperature*
XGBoost	*Extreme gradient-boosting*

References

[1] Creel P, McKenzie D, Briney M, Pipes H, editors. Real-Time Cementing Designs vs. Actual Jobs in Progress. IADC/SPE Drilling Conference; 2006. SPE-98079-MS.

[2] Khan JA, Irawan S, Dan IBM, Cai B. Determining the difference of kick tolerance with single bubble and dynamic multiphase models: Evaluation of well-control with water/synthetic based muds. Ain Shams Engineering Journal, 2022;13(4):101678. https://doi.org/https://doi.org/10.1016/j.asej.2021.101678

[3] Rostami SA, Gumus F, Simpkins D, Pobedinski I, Kinik K, Rajabi MM, editors. New Generation of MPD Drilling Software - From Quantifying to Control. SPE Annual Technical Conference and Exhibition; 2016. D031S050R005.

[4] AlBahrani H, Alsheikh M, Wagle V, Alshakhouri A, editors. Designing Drilling Fluids Rheological Properties with a Numerical Geomechanics Model for the Purpose of Improving Wellbore Stability. IADC/SPE International Drilling Conference and Exhibition; 2022. D011S009R003.

[5] Roy S, Kamal SZ, Frazier R, Bruns R, Hamlat YA, editors. Inline Drilling Fluid Property Measurement, Integration, and Modeling to Enhance Drilling Practice and Support Drilling Automation. Abu Dhabi International Petroleum Exhibition & Conference; 2021. D041S108R003.

[6] Abdelaal A, Ibrahim A, Elkatatny S, editors. Rheological Properties Prediction of Flat Rheology Drilling Fluids. 56th US Rock Mechanics/Geomechanics Symposium; 2022. ARMA-2022-0822.

[7] Yang M, Luo H, Xie R, Wu P, Yang M, Yang L, et al. Estimating and optimizing wellbore pressure during primary cementing in low pressure and leakage formations. 2022;208:109236.

[8] Alkinani HH, Al-Hameedi AT, Flori RE, Dunn-Norman S, Hilgedick SA, Alsaba MT, editors. Updated Classification of Lost Circulation Treatments and Materials with an Integrated Analysis and their Applications. SPE Western Regional Meeting; 2018. D031S003R005.

[9] Alkinani HH, Al-Hameedi AT, Dunn-Norman S, Alkhamis MM, Mutar RA, editors. Prediction of Lost Circulation Prior to Drilling for Induced Fractures Formations Using Artificial Neural Networks. SPE Oklahoma City Oil and Gas Symposium; 2019. D021S008R001.

[10] Costa SS, Stuckenbruck S, da Fontoura SA, Martins AL, editors. Simulation of transient cuttings transportation and ECD in wellbore drilling. SPE Europec featured at EAGE Conference and Exhibition?; 2008: SPE.

[11] Caicedo H, Pribadi MA, Bahuguna S, Wijnands F, Setiawan NB, editors. Geomechanics, ECD management and RSS to manage drilling challenges in a mature field. SPE Oil and Gas India Conference and Exhibition?; 2010: SPE.

[12] Liu N, Zhang D, Gao H, Hu Y, Duan L. Real-Time Measurement of Drilling Fluid Rheological Properties: A Review. Sensors, 2021;21(11):3592. https://www.mdpi.com/1424-8220/21/11/3592

[13] Zhang H, Sun T, Gao D, Tang HJC, fuels to, oils. A new method for calculating the equivalent circulating density of drilling fluid in deepwater drilling for oil and gas. 2013;49:430-8.

[14] Igwilo K, N U, Okoro E, Adenubi S, Obibuike U. Article Open Access Improved Model for Equivalent Circulating Density and Improved Annular Pressure Drop Estimation: Drilling Approach. Petroleum and Coal, 2022; 64: 167-76.

[15] Bybee K. Equivalent-Circulating-Density Fluctuation in Extended-Reach Drilling. Journal of Petroleum Technology, 2009; 61(02): 64-7. https://doi.org/10.2118/0209-0064-jpt

[16] Ataga E, Ogbonna J, editors. Accurate estimation of equivalent circulating density during high pressure high temperature (HPHT) drilling operations. SPE Nigeria Annual International Conference and Exhibition; 2012: SPE.

[17] Rommetveit R, Ødegård SI, Nordstrand C, Bjørkevoll KS, Cerasi P, Helset HM, et al., editors. Drilling a challenging HPHT well utilizing an advanced ECD Management System with decision support and real time simulations. SPE/IADC Drilling Conference and Exhibition; 2010: SPE.
[18] Erge O, Vajargah AK, Ozbayoglu ME, van Oort E, editors. Improved ECD prediction and management in horizontal and extended reach wells with eccentric drillstrings. SPE/IADC Drilling Conference and Exhibition; 2016: SPE.
[19] Al-Hameedi AT, Alkinani HH, Dunn-Norman S, Flori RE, Hilgedick SA, Alkhamis MM, et al., editors. Predictive Data Mining Techniques for Mud Losses Mitigation. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition; 2018: SPE.
[20] Andagoya Carrillo KI, Avellán FJ, Camacho G, editors. ECD and Downhole Pressure Monitoring While Drilling at Ecuador Operations. SPE Latin American and Caribbean Petroleum Engineering Conference; 2015. D011S003R006.
[21] Diaz H, Miska S, Takach N, Yu M, editors. Modeling of ECD in Casing Drilling Operations and Comparison with Experimental and Field Data. IADC/SPE Drilling Conference; 2004. SPE-87149-MS.
[22] Razi M, Arz A, Naderi AJEJoSR. Annular pressure loss while drilling prediction with artificial neural network modeling. Eur J Sci Res., 2013; 95(2): 272-88.
[23] Kalogirou SA. Artificial intelligence for the modeling and control of combustion processes: a review. Progress in Energy and Combustion Science. 2003;29(6):515-66. https://doi.org/https://doi.org/10.1016/S0360-1285(03)00058-3
[24] Kalogirou SA, Panteliou S, Dentsoras A. MODELING OF SOLAR DOMESTIC WATER HEATING SYSTEMS USING ARTIFICIAL NEURAL NETWORKS. Solar Energy. 1999;65(6):335-42. https://doi.org/https://doi.org/10.1016/S0038-092X(99)00013-4
[25] Alkinani HH, Al-Hameedi AT, Dunn-Norman S, Al-Alwani MA, Mutar RA, Al-Bazzaz WH, editors. Data-driven neural network model to predict equivalent circulation density ECD. SPE Gas & Oil Technology Showcase and Conference; 2019: SPE.
[26] Abdelgawad KZ, Elzenary M, Elkatatny S, Mahmoud M, Abdulraheem A, Patil S. New approach to evaluate the equivalent circulating density (ECD) using artificial intelligence techniques. Journal of Petroleum Exploration and Production Technology. 2019; 9(2): 1569-78. https://doi.org/10.1007/s13202-018-0572-y
[27] Rahmati AS, Tatar AJO, Science G, nouvelles TRdIE. Application of Radial Basis Function (RBF) neural networks to estimate oil field drilling fluid density at elevated pressures and temperatures. Oil & Gas Science and Technology - Rev. IFP.,2019; 74:50.
[28] Kandil A, Khaled S, Elfakharany TJAE. Prediction of the equivalent circulation density using machine learning algorithms based on real-time data. AIMS Energy, 2023; 11(3): 425-453.
[29] Ahmadi MA, Shadizadeh SR, Shah K, Bahadori AJEJoP. An accurate model to predict drilling fluid density at wellbore conditions. Egyptian Journal of Petroleum, 2018; 27(1): 1-10.
[30] Alsaihati A, Elkatatny S, Abdulraheem AJAo. Real-time prediction of equivalent circulation density for horizontal wells using intelligent machines. ACS Omega, 2020; 6(1): 934-42.
[31] Gamal H, Abdelaal A, Elkatatny S. Machine Learning Models for Equivalent Circulating Density Prediction from Drilling Data. ACS Omega, 2021; 6(41): 27430-27442. https://doi.org/10.1021/acsomega.1c04363
[32] Al-Rubaii M, Al-Shargabi M, Aldahlawi B, Al-Shehri D, Minaev KMJS. A developed robust model and artificial intelligence techniques to predict drilling fluid density and equivalent circulation density in real time. Sensors, 2023; 23(14): 6594. https://doi.org/10.3390/s23146594
[33] Metwally M. Advancing ERW Efficiency: Developing High-Performance Water-Based Mud with Friction Reducer and Precision Prediction for ECD and Open Hole Extension Limits using Herschel-Bulkley Rheological Model. J Petro Chem Eng., 2024; 2(1): 9-19.
[34] Liao M, Wang Y, Ma J, Liu Z, Ji CJG. Calculation of Cementing ECD for Well Annuli with Narrow or Varied Radial Sizes. Geofluids, 2023; (1): 1227197. http://dx.doi.org/10.1155/2023/1227197
[35] Ekechukwu G, Adejumo A. Explainable machine-learning-based prediction of equivalent circulating density using surface-based drilling data. Scientific Reports. 2024; 14(1): 17780. https://doi.org/10.1038/s41598-024-66702-w
[36] Hansen KV. Neural networks for primary reflection identification. SEG Technical Program Expanded Abstracts 1993;10.1190/1.18224501993. p. 242-5. https://doi.org/10.1190/1.1822450
[37] Sprunger C, Muther T, Syed FI, Dahaghi AK, Neghabhan S. State of the art progress in hydraulic fracture modeling using AI/ML techniques. Modeling Earth Systems and Environment. 2022; 8(1): 1-13. https://doi.org/10.1007/s40808-021-01111-w

[38] Moran D, Ibrahim H, Purwanto A, Osmond J, editors. Sophisticated ROP Prediction Technologies Based on Neural Network Delivers Accurate Drill Time Results. IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition; 2010. SPE-132010-MS.

[39] Arehart RA. Drill-Bit Diagnosis With Neural Networks. SPE Computer Applications. 1990;2(04):24-8. https://doi.org/10.2118/19558-pa

[40] Haciislamoglu M, Cartalos U, editors. Practical Pressure Loss Predictions in Realistic Annular Geometries. SPE Annual Technical Conference and Exhibition; 1994. SPE-28304-MS.

[41] Arshad U, Jain B, Ramzan M, Alward W, Diaz L, Hasan I, et al., editors. Engineered solution to reduce the impact of lost circulation during drilling and cementing in Rumaila Field, Iraq. International Petroleum Technology Conference; 2015: IPTC.

[42] Al-Rubaii MM, Al-Nassar FY, Al-Harbi S, editors. A new real time prediction of equivalent circulation density from drilling surface parameters without using pwd tool. SPE Symposium: Unconventionals in the Middle East; 2022: SPE.

[43] Hoberock L, Thomas D, Nickens H. Here's how compressibility and temperature affect bottom-hole mud pressure. Oil Gas J., 1982; 80(12).

[44] Peters EJ, Chenevert ME, Zhang C. A model for predicting the density of oil-based muds at high pressures and temperatures. SPE drilling engineering. 1990;5(02):141-8.

[45] Hemphill T, Ravi K, Bern P, Rojas JC, editors. A simplified method for prediction of ECD increase with drillpipe rotation. SPE Annual Technical Conference and Exhibition?; 2008: SPE.

[46] Ahmed R, Enfis M, Miftah-El-Kheir H, Laget M, Saasen A, editors. The effect of drillstring rotation on equivalent circulation density: modeling and analysis of field measurements. SPE Annual Technical Conference and Exhibition?; 2010: SPE.

[47] Baranthol C, Alfenore J, Cotterill M, Poux-Guillaume G, editors. Determination of hydrostatic pressure and dynamic ECD by computer models and field measurements on the directional HPHT well 22130C-13. SPE/IADC Drilling Conference and Exhibition; 1995: SPE.

[48] Osman E, Aggour M, editors. Determination of drilling mud density change with pressure and temperature made simple and accurate by ANN. SPE Middle East Oil and Gas Show and Conference; 2003: SPE.

[49] Harris O, Osisanya S, editors. Evaluation of equivalent circulating density of drilling fluids under high-pressure/high-temperature conditions. SPE Annual Technical Conference and Exhibition?; 2005: SPE.

[50] Elzenary M, Elkatatny S, Abdelgawad KZ, Abdulraheem A, Mahmoud M, Al-Shehri D, editors. New technology to evaluate equivalent circulating density while drilling using artificial intelligence. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition; 2018: SPE.

[51] Ahmadi MAJN. Toward reliable model for prediction drilling fluid density at wellbore conditions: A LSSVM model. 2016; 211: 143-9.

[52] Ahmadi MA, Shadizadeh SR, Shah K, Bahadori A. An accurate model to predict drilling fluid density at wellbore conditions. Egyptian Journal of Petroleum. 2018; 27(1): 1-10.

[53] Gamal H, Abdelaal A, Elkatatny S. Machine Learning Models for Equivalent Circulating Density Prediction from Drilling Data. ACS Omega. 2021;6(41): 27430-27442.

To whom correspondence should be addressed: Elfakeur Abidi Saad, Department of Drilling and Rig mechanics, University of Ouargla, 30000 Ouargla, Algeria, E-mail: elfakeur@yahoo.fr