

## Machine Learning Models for Real-Time Skin Factor Prediction in Foamed Acid Treatments

Nermeen Hana<sup>1</sup> and Said K. Elsayed<sup>2</sup>

<sup>1</sup> Ain Shams University, Cairo, Egypt

<sup>2</sup> Suez University, Suez, Egypt

Received February 9, 2025; Accepted May 21, 2025

---

### Abstract

For matrix acidizing operations an appropriate acid implementation requires careful planning to minimize well damage while preventing acid overconsumption. Machine learning algorithms allow real-time stimulation result assessments to accomplish this objective. This research seeks to establish advanced machine learning models which provide precise real-time readings for skin factor together with bottom-hole pressure at the coiled tubing nozzle during foamed acid treatments. This research built nine advanced machine learning models which used standard matrix acidizing operational parameters. Real bottom-hole pressure measurements obtained from deployed pressure gauges formed the basis for training the developed models. The predicted bottom-hole pressure enabled the determination of real-time skin factor measurements. The development of machine learning algorithms took place through analysis of an extensive dataset obtained from 31 wells. Through comparison of predictive bottom-hole pressure models with actual field measurements the most efficient machine learning methods demonstrated exceptionally low root mean square error. The Neural Network, AdaBoost, Random Forest and K-Nearest Neighbor algorithms produced root mean square error values of 0.047, 0.048, 0.054 and 0.058 respectively. This study introduces novel methods of using advanced machine learning models as predictive tools to track bottom-hole pressure and skin factor throughout matrix acidizing procedures with foamed acid. These predictive models act as accurate and swift replacements for the traditional pre- and post-stimulation well testing, conventional empirical multiphase flow correlations as well as mechanistic models and unified models. The proposed models function as an economical replacement for downhole pressure gauges as well because they eliminate both cost and duration issues.

**Keywords:** Machine learning; Acid; Coiled tubing; Stimulation; Skin; Formation damage.

---

### 1.1. The importance of skin factor

During production, the skin factor acts as a quantitative measure to indicate possible pressure loss increases at the wellbore area [1-2]. The skin factor combines formation damage with perforation damage as well as partial penetration, well deviation and pseudo skin according to [3-4]. The skin factor represents positive values when the well experiences damage near the wellbore, yet negative values manifest when hydraulic fracturing or acid treatments stimulate the formation [5-6]. The assessment of production efficiency together with well stimulation selection depends heavily on the skin factor measurement [7]. Engineers who understand each part of the total skin factor can determine skin damage sources while developing the optimal pressure drop reduction solution around the near wellbore area [8].

### 1.2. The importance of predicting skin factor during acid treatment

Real-time monitoring of skin factor is essential for acid treatment operations because it helps optimize pumping parameters together with maintaining efficient acid diversion and distribution [9-10]. Usually, well tests conducted before and after the treatment allow measurement of skin factors, which evaluate how well the stimulation treatment worked [11-12]. However, the

dynamic analysis of acid injection treatment becomes possible by real-time monitoring of skin factors, which optimizes both acid parameters and volumes of injected acid [13-14]. This dynamic analysis ensures the damage removal process is effective while simultaneously maintaining the formation condition no worse than its initial state. Moreover, real-time monitoring provides essential information about acid formation interactions that enhances both treatment practices and well results [15].

### 1.3. Calculation methods of skin factor

Multiple traditional calculations for skin factor depend on pressure transient analysis through pressure buildup or falloff testing [16]. The examination of pressure behavior during well closures after definite production or injection periods forms the basis of these evaluation methods [17-19]. Pressure and time data collection enables the determination of skin factor through predetermined analytical models including Horner plot and derivative analysis according to [20]. However, the analysis of these methods needs both prolonged periods of well shut-in along with precise interpretation before they produce effective results. Moreover, these methods provide one global skin value across the entire well while not showing the specific skin factor evolution during the acid treatment procedure [21]. Several proposed methods exist for real-time skin factor evaluation in matrix acidizing operations [22-24]. These techniques aim to fix the limitations of pre-treatment and post-treatment well testing methods by providing continuous stimulation data feedback [25-26]. Two primary methods exist for this calculation.

#### 1.3.1. Steady-state model (Pacaloni's technique)

This method calculates skin factors using real-time pressure and rate information under a steady-state approach [27]. Pacaloni's method provides convenient application yet delivers an exaggerated skin factor because it does not account for transient flow effects according to [28]. The method results in a higher-than-actual skin factor, particularly during changes in rate, and it makes no adjustments for diverting agent effects as well. The acid bank radius of 4 ft was chosen by Pacaloni as estimation to simplify the calculations. Pacaloni derived the formula to determine steady-state skin factor as presented in Equation (1).

$$S(t) = \frac{0.00708kh\Delta p}{\mu \cdot q_i} + \ln \frac{r_b}{r_e} \quad (1)$$

#### 1.3.2. Transient model (Prouvost and Economides' technique)

The transient model depends on actual injection rate histories and fluid sequences to create pressure simulations by maintaining a steady skin factor [19]. The difference between actual observed pressure values and simulated pressure readings becomes the basis for determining the skin factor variation. The transient model offers greater accuracy in estimating the skin factor through its ability to deal with flow transients because it considers actual flow rates and fluid sequence changes [29]. However, this method needs a computer simulator with advanced functionalities. The method uses pressure transient analysis under infinite acting boundary and initial conditions according to Equation (2).

$$p(r, t) = p_i - \frac{q_i \mu}{4\pi kh} \ln \left( \frac{4kt}{\gamma \Phi \mu c_t r_w^2} \right) \quad (2)$$

Including skin effect and modifying to acid injection, the equation becomes Equation (3).

$$p_{sim} = p_i + \frac{162.6q_i B \mu}{kh} \left( \log t + \log \frac{k}{\Phi \mu c_t r_w^2} - 3.2275 + 0.87S_o \right) \quad (3)$$

After changing the injection rate, the transient response would follow a superposition relationship; the injection pressure for one rate change would be Equation (4).

$$p_{sim} = p_i + \frac{162.6q_i B \mu}{kh} \left( \log (t + \Delta t) + \log \frac{k}{\Phi \mu c_t r_w^2} - 3.2275 + 0.87S_o \right) + \frac{162.6(q_{i2} - q_i) B \mu}{kh} \left( \log \Delta t + \log \frac{k}{\Phi \mu c_t r_w^2} - 3.2275 + 0.87S_o \right) \quad (4)$$

The difference between the simulated pressure and the measured value is interpreted as due to the difference between the actual skin value and the initial value used for the simulation, Equation (5).

$$S(t) = S_o + \frac{kh}{2.34 \times 10^5 q_i^{B\mu}} [p_{\text{meas}}(t) - p_{\text{sim}}(t, S_o)] \quad (5)$$

#### 1.4. Calculation methods of bottomhole pressure in a dynamic injection of foam fluid

The estimation of bottomhole pressure (BHP) is possible through various multiphase flow correlations. The correlations differ in their prediction accuracy as well as their computational complexity and their suitability for predicting foam flow characteristics. Empirical correlations, mechanistic models, unified models represent the three main categories at hand.

Empirical correlations proposed by Beggs and Brill [30] besides Orkiszewski and Hagedorn and Brown [31] find frequent use because they require minimal computational effort during field work. The use of experimental data in these correlations enables their practical application for fast BHP estimations in field scenarios. The correlations used for BHP estimation generally stem from conventional gas-liquid systems but do not completely address the rheological and stability behavior of foam [32].

A more physics-based modeling method called mechanistic models applies mass, momentum and energy conservation laws as presented in Ansari et al. [33] and Taitel and Dukler [34]. These improvement methodologies deliver more precise BHP predictions through comprehensive calculations of phase interactions along with measurement of slip velocities and flow regime changes. Real-time applications restrict the use of these models because they need details about fluid properties and well geometry as well as extensive input data which make their practicality limited due to increased complexity.

Unified models, these models combine mathematical precision from conservation laws with observations from real-world measurements. The compatibility of OLGA [35] and TUFFP [36] operates at high standards yet their need for detailed wellbore geometry and foam properties demands intensive resources and input. This reduces their performance potential during real-time applications. Moreover, their sophisticated equations do not prevent them from needing adjustments through real-field data measurements to achieve optimal performance.

#### 1.5. Machine learning models for predicting skin factor

The oil and gas industry has shown extensive interest in machine learning algorithms because these models solve problems that traditional techniques cannot handle effectively [37-39]. These models provide real-time processing and tracking features, which help engineers make rapid operational decisions and implement modifications through their systems [40-41]. Therefore, we introduce machine learning models as a powerful replacement for traditional approaches because these models can automatically process and predict critical parameters with accessible data sources [42-45]. Moreover, the automated process decreases human interaction, thus allowing engineers to focus on critical strategic issues [46-49]. During the last several decades, the industry gathered substantial data quantities, yet finding valuable patterns from these big data sets remains challenging [50-52]. However, machine learning algorithms excel in analyzing extensive and intricate datasets to find meaningful correlations for forecasting applications and process enhancement [53-55]. Although many experts consider ML models optimal for real-time parameter tracking, these systems receive minimal research interest. To the best of our knowledge, the machine learning approach has not been implemented to predict real-time bottom-hole pressure and skin in any of the available studies about foamed acid treatments. Therefore, in an effort to address this knowledge gap, this study seeks to develop nine advanced ML algorithms that can predict real-time bottom-hole pressure and skin factor throughout matrix acidizing treatments with foamed acid. These predictive models act as accurate and swift replacements for the traditional pre- and post-stimulation well testing, conventional empirical multiphase flow correlations, mechanistic models, and unified models. The

proposed models function as an economical replacement for downhole pressure gauges as well because they eliminate both cost and duration issues.

## 2. Methodology

This study follows a systematic research approach which is demonstrated in Figure 1 through four connected stages. The strategic planning of every research phase generates a systematic approach that links various steps to fulfill the project objective.

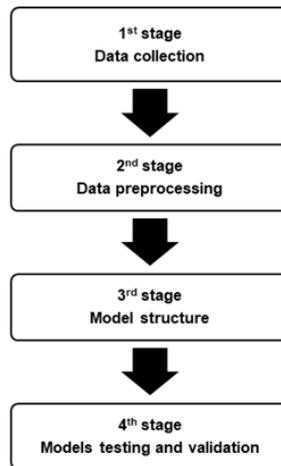


Figure 1. A diagram of the methodology.

### 2.1. Data collection

Surface and downhole sensors are commonly used by operators in modern acid treatment operations for process management and enhancement purposes. To conduct this research, a comprehensive set of 33,453 actual field measurements was collected. The data came from 31 wells present in the Western Desert of Egypt. Effectiveness in machine learning model development depends on large dataset implementation because it enables three fundamental capabilities. Firstly, the implementation of extensive data boosts model generalization effectiveness for unknown cases. The extensive range of examples and edge cases that the model encounters through this method protects against overfitting. Using large training data also enables models to detect actual patterns while eliminating background noise, which results in more accurate predictions. Machine learning algorithms, particularly deep learning models, require access to large amounts of data in order to optimize their parameters for making reliable decisions. The dataset supported model evaluation and looked at disparities between modeled nozzle outlet bottomhole pressures and real pressure measurements conducted by downhole pressure gauges. Table 1 displays the dataset information that includes pressure at a coiled tubing outlet (BHP), coiled tubing depth (CTD), coiled tubing inside diameter (CTID), temperature at a coiled tubing outlet (BHT), acid flow rate at surface (AFR), coiled tubing pressure at surface (CTP), and nitrogen rate at surface (NFR). Various operating conditions appear throughout the presented data set. A generalized model will be developed as part of this paper because the data includes various parameters across different operating conditions.

Table 1. Statistical analysis for the collected database.

Parameter	Unit	Max	Min	Average	Median
Pressure at a coiled tubing outlet	PSI	6565	1006	3557	3408
Acid flow rate at surface	BBL/min	2.17	0.25	1.57	1.73
Temperature at a coiled tubing outlet	°F	300	134	220	244
Coiled tubing pressure at surface	PSI	7342	897	4675	4672
Coiled tubing depth	FT	13971	2920	9103	10554
Nitrogen rate at surface	SCF/M	953	77	672	732
Coiled tubing inside diameter	Inch	1.25	1.23	1.24	1.23

Figure 2 displays a violin plot that reveals the value distributions for all study parameters through box plot and density estimation fusion. Each violin represents probability density distribution across value levels in the data while displaying the distribution patterns. The interquartile range (IQR) occupies the black inner bars of this plot and demonstrates the range where 50% of data lies. The white dot shows the median position of the data points. CTID distribution displays skewness because two different coiled tubing sizes were used during well activities while BHP along with CTP follow a more extensive distribution pattern. The visual presentation aids in observing how data distributions extend and centralize and shape themselves among various parameters.

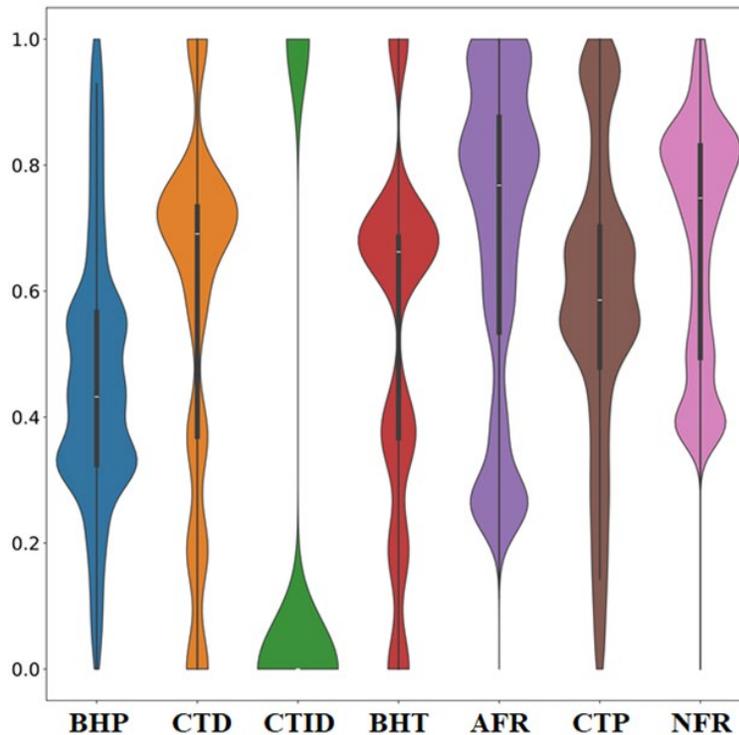


Figure 2. Violin plots for each parameter in the dataset.

## 2.2. Data preprocessing

The core essential requirement for effective machine learning model development is data preprocessing, which transforms unprocessed data into a ready analysis format. At the initial stage of research, we performed strict data cleaning to eliminate both inaccurate values and stuck sensor readings. The analysis dealt with missing data by using complete row elimination or running linear interpolation according to different situations. Outlier detection and removal constituted the subsequent vital step because anomalous data points deteriorate the prediction accuracy and disrupt model training. Box plot analysis served as the main statistical technique for detecting these outliers. The data went through an organization process together with normalization steps that standardized the data format suitable for machine learning algorithm execution. The thorough method used for data preprocessing leads to high-quality input for model training sessions.

Data normalization functions as a preprocessing technique utilized in machine learning to rescale numerical features into a uniform value scope which extends from 0 to 1 or from -1 to 1 without altering the original feature distributions. The normalized rescaling procedure safeguards all features by providing them equal importance in model contribution without any large-scale features surpassing smaller features. The normalization process stands vital be-

cause gradient-based machine learning algorithms which include SVMs and k-means clustering along with logistic regression and neural networks achieve better results when dealing with features normalized to uniform scales. Using normalization makes the model train faster while preventing biases from larger feature values and improves both the consistent and precise outcome of training. The Min-Max Scalar normalization method was adopted here to change all numerical data points into values between 0 and 1 using minimum and maximum values. By using the following equation the data transformation occurs:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (6)$$

where: X is the original data point;  $X_{\text{min}}$  is the minimum value in the feature;  $X_{\text{max}}$  is the maximum value in the feature;  $X_{\text{scaled}}$  is the normalized value within the [0,1] range.

Finally, the preprocessed data underwent a division process that assigned 80% of the data for training purposes and 20% for testing functions. The machine learning model receives training input from the training set so it can detect patterns and data relationships. We separated the testing set from the rest of the data for performance assessment because it contains new inputs the model must function with correctly. Our decision to use an 80/20 split came because the data size falls into the medium category. This ratio provides increased training material to learn from so models can avoid underfitting. Moreover, this ratio is essential for neural network models because of their complexity. The 80/20 ratio achieved optimum performance in our research through applications of cross-validation and hyperparameter tuning methods.

### 2.3. Model structure

The development and evaluation of nine machine learning models through Python 3.10.12 required the implementation of specific hyperparameters as listed in Table 2.

Table 2. Summary of the machine learning models used and their algorithm hyperparameters

Model	Algorithm parameters
GB	<ul style="list-style-type: none"> <li>▪ Method: xgboost</li> <li>▪ Number of trees is 100.</li> <li>▪ Learning rate is 0.099.</li> <li>▪ Regularization is 0.6</li> <li>▪ Limit depth is 6.</li> </ul>
AdaBoost	<ul style="list-style-type: none"> <li>▪ The number of estimators is 50.</li> <li>▪ Learning rate is 1.</li> <li>▪ Regression loss function is linear.</li> </ul>
RF	<ul style="list-style-type: none"> <li>▪ Number of trees in the forest is 10.</li> <li>▪ Minimum subset size is 5.</li> </ul>
SVMs	<ul style="list-style-type: none"> <li>▪ SVM Cost is 1.</li> <li>▪ Regression loss epsilon: 0.1.</li> <li>▪ Kernel type is radial basis function.</li> <li>▪ Numerical tolerance: 0.001</li> <li>▪ Iteration limit is 100.</li> </ul>
DT	<ul style="list-style-type: none"> <li>▪ Minimum instances in leaves are 19.</li> <li>▪ Minimum subset size: 9.</li> <li>▪ Maximal tree depth is 100.</li> <li>▪ The stopping point is at 95% of the majority.</li> </ul>
KNN	<ul style="list-style-type: none"> <li>▪ Number of nearest neighbors is 5.</li> <li>▪ Metric is Euclidean.</li> <li>▪ Weights are uniform.</li> </ul>
LR	<ul style="list-style-type: none"> <li>▪ None.</li> </ul>
NN	<ul style="list-style-type: none"> <li>▪ Neurons per hidden layer are 500.</li> <li>▪ Activation is ReLu.</li> <li>▪ Solver is L-BFGS.</li> <li>▪ Regularization is 0.04.</li> <li>▪ Maximum iterations are 500.</li> </ul>

Model	Algorithm parameters
SGD	<ul style="list-style-type: none"> <li>▪ loss function is hinge.</li> <li>▪ Regularization is Lasso (L1).</li> <li>▪ Learning rate is constant.</li> <li>▪ Initial learning rate is 0.01.</li> <li>▪ Number of iterations is 1000.</li> </ul>

The diverse model structure provides essential benefits to this research. By applying different machine learning models on the same dataset, we gain access to each method's benefits and identify their weaknesses. For example, the linear regression model provides easy interpretation but fails when working with complex data patterns, whereas neural networks excel at relationship detection but tend to overfit data. The analysis of diverse models enables us to determine the best data pattern model that maintains generalization power through a balance between bias and variance. Moreover, different stakeholder priorities can be accommodated through this extensive modeling approach, which provides options from interpretable to efficient to predictive models. Usually, organizations maintain dual perspectives regarding decision-making transparency versus accuracy achievement because they either need simple transparent models (e.g., linear regression) or the precise predictive capabilities (e.g., gradient boosting). The models used regularization techniques to avoid overfitting through penalty terms that were added to the loss function calculations. The additional penalty term in the model prevents complexity from reaching excessive levels to promote accurate predictions on new data points. Perfect training data fitting becomes impossible through regularization even if complex models demonstrate this ability. Regularization methods enforce the restriction of model complexity to allow detection of relevant path.

The Pythagorean Forest depicts the 10 trees that resulted from the RF model application (Figure 3). Each random tree construction appears as a display element in the Pythagorean tree structure. The most accurate tree emerges from branches with maximum length and highest brightness, indicating that a few main attributes can properly divide the branches. Standard deviation served as the basis for data coloring in the resulting trees, which were developed from regression trees. The accuracy and generalization power along with the stability of machine learning models increase through the optimized random forest technique included in the Pythagorean Forest. This methodology improves feature interactions while decreasing bias and variance along with overfitting prevention by implementing optimized distance-based metrics in the framework. This approach delivers more dependable prediction results together with easier interpretation than advanced models such as deep learning systems.

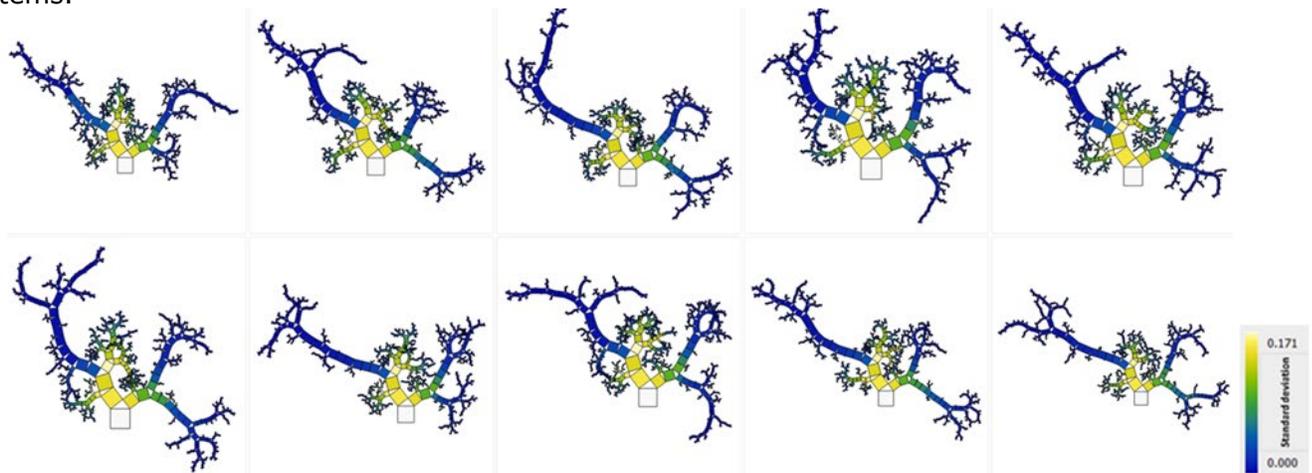


Figure 3. Pythagorean Forest shows all learned Decision Tree models from the RF model.

### 3. Results and discussion

#### 3.1. Model results

The analysis results from machine learning models appear in Table 3. The mean square error (MSE) and root mean square error (RMSE) with mean absolute error (MAE) and coefficient of determination ( $R^2$ ) values appear for each examined model. MSE uses actual and predicted value differences to calculate the average squared variance so it shows higher sensitivity to huge prediction errors from the squaring step. This technique proves useful in regression problems while facing significant influence from outliers. The model achieves better performance when the MSE value remains low since it suggests the predicted values closely match actual values. The square function brings a problem because it introduces metric skew that may not match those of the target variable. The RMSE calculation derives from MSE square rooting to produce error metrics that match the units of the target variable for enhanced interpretability. The implementation of RMSE provides significant measurement benefits because it increases penalty levels for large errors without losing unit accuracy. A lower RMSE value signifies better performance because smaller errors occur. The use of RMSE becomes necessary when highly significant errors need to be penalized. The actual and predicted values comparison through MAE produces an average of their absolute difference measurements. The error calculation in MAE avoids squaring differences, which provides better sensitivity to outliers with a balanced error evaluation. A model demonstrates superior performance when its MAE measurement value reduces. This scoring method shows superior performance against outliers because it avoids squaring error values yet provides less discrimination regarding error size relative to the RMSE method. The  $R^2$  score demonstrates the ability of a model to explain data point variance. The measure extends from 0 to 1, so a value of 1 indicates perfect model fit, and lower scores indicate minimal explanation of target variable variance. Higher  $R^2$  values close to one indicate better model performance because they demonstrate better variability explanation in the data. However, a very high  $R^2$  value does not guarantee an excellent model when data fits the algorithm too closely.

Table 3. Performance comparison of the developed machine learning models.

Model	MSE	RMSE	MAE	$R^2$
Neural network	0.002	0.050	0.017	0.946
AdaBoost	0.003	0.053	0.016	0.940
Random forest	0.003	0.054	0.019	0.937
kNN	0.003	0.058	0.024	0.927
Gradient boosting	0.006	0.078	0.040	0.866
Tree	0.006	0.079	0.033	0.863
SVM	0.009	0.093	0.070	0.813
Linear regression	0.027	0.163	0.125	0.422
SGD	0.029	0.171	0.134	0.362

Figure 4 reveals the normalized anticipated pressure measurements at the coiled tubing nozzle's outlet from neural network predictions align with actual pressure readings obtained from memory gauges. The predictions made by the machine learning model remain distributed equally while located near the 45° straight line in the results. This means that the developed neural network delivered precise predictions for bottomhole pressure values that cover the entire range of operational parameters used in foamed acid treatments. Real-time skin factor calculations can be performed precisely through the modified Prouvost & Economides' technique and the modified Paccaloni's technique based on these accurate downhole pressure estimations. The essential parameters affecting BHP prediction through the neural network model appear in Figure 5, which serves as the most precise predictive model in this research. Each data instance in the dataset corresponds to SHAP values displayed on the horizontal axis throughout the graph for every parameter. The SHAP value measures the degree to which the feature value affects the predicted BHP model output relative to the standard model prediction. The right side values of the center indicate positive SHAP values, which reveal features that

contribute a positive effect on BHP prediction. The BHP prediction receives negative influence from features that fall to the left of the central position. The color scheme indicates values in two ways: red indicates larger values and blue identifies lower values. Developing the color spectrum starts from analyzing all the values found in the specific dataset feature. Among the studied parameters, CTP, AFR, and CTD demonstrate the strongest impact on projected BHP predictions, but the other features show a smaller influence.

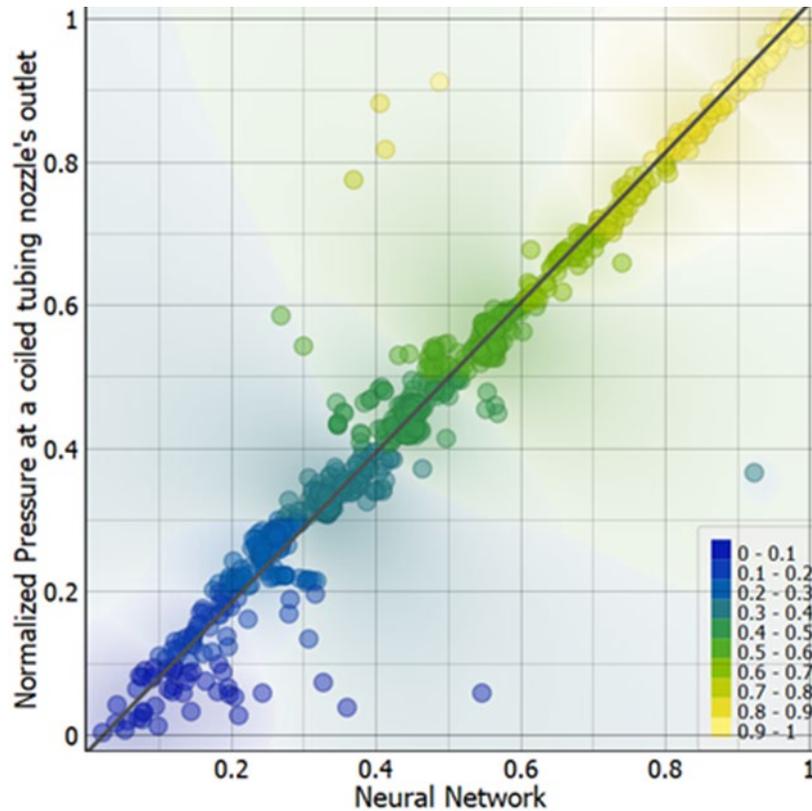


Figure 4. Predicted vs. actual normalized pressure using a neural network model.

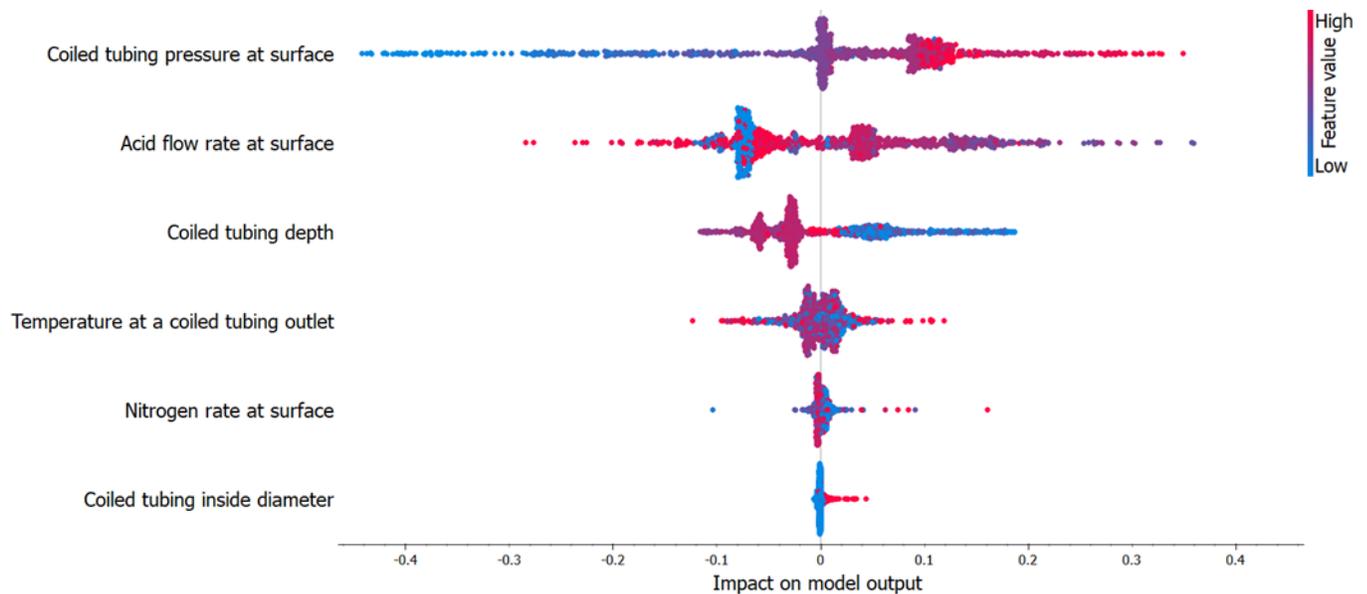


Figure 5. SHAP plot of the neural network model.

### 3.2. Model testing and validation

The evaluation of machine learning model precision utilizes two major methods. The evaluation of machine learning model performance through K-fold cross-validation ranks as an optimal method for assessing anticipated results from particular learning models. As a dependable measurement approach, it helps decrease model overfitting by specific datasets. The model requires splitting a dataset into k sections, which should hold similar proportional values. The procedure performs a k-time training sequence where k-1 folds serve for model training followed by a single fold earmarked for testing until every fold completes its cycle. The assessment of performance involves averaging performance metrics obtained from each of the k folds. Each iteration in the multi-stage cross-validation procedure uses different random divisions to produce a decreased variability of model performance due to data partitioning. Table 4 presents the results of the K-fold cross-validation, which operated with 10 folds for assessment.

The evaluation of machine learning models uses repeated random sampling as a statistical resampling technique. The protocol shares key features with k-fold cross-validation since both methods produce multiple training and testing partitions from the dataset. The model receives training on each partition through the training set before undergoing evaluation on the test set. The performance measurement emerges through averaging performance indicators from all simulation rounds. The method provides strong benefits when dealing with limited datasets or when users need to determine reliable model performance metrics. The assessment process becomes less influenced by random partitioning because smaller data sets and multiple data splits receive focused evaluation. The method strengthens the reliability of model performance evaluation and simultaneously prevents the model from overfitting to a particular training dataset. The repeated random sampling technique yielded its results through 10 separate trials, which appear in Table 5. The results include MSE, RMSE, MAE, and R<sup>2</sup> for every model in the analysis.

Table 4. Results of K-fold cross-validation procedure.

Model	MSE	RMSE	MAE	R <sup>2</sup>
Neural network	0.002	0.047	0.017	0.947
AdaBoost	0.002	0.048	0.016	0.943
Random forest	0.003	0.054	0.020	0.929
kNN	0.003	0.058	0.025	0.918
Gradient boosting	0.005	0.071	0.032	0.878
Tree	0.006	0.078	0.040	0.853
SVM	0.009	0.097	0.074	0.770
Linear regression	0.025	0.159	0.120	0.381
SGD	0.002	0.047	0.017	0.947

Table 5. Results of random sampling procedure.

Model	MSE	RMSE	MAE	R <sup>2</sup>
Neural network	0.002	0.044	0.017	0.951
AdaBoost	0.003	0.052	0.018	0.933
Random forest	0.003	0.054	0.021	0.926
kNN	0.003	0.057	0.026	0.919
Gradient boosting	0.005	0.070	0.033	0.878
Tree	0.006	0.079	0.042	0.845
SVM	0.008	0.090	0.067	0.800
Linear regression	0.025	0.158	0.119	0.378
SGD	0.026	0.161	0.120	0.352

### 3.3. Field application

Well-X operates as a water injection well within the Egyptian Western Desert. The drilled well reached a depth of 8,620 feet. A foamed acid treatment through the coiled tube proved

to be the most effective method to increase the injectivity index for the damaged carbonate reservoir Formation Y. A well test verified the skin factor value of 4 prior to acid treatment execution. A complete summary of parameters can be found in Table 6. The developed neural network model receives Well-X data for actual pressure prediction at coiled tubing nozzles during foamed acid treatment. Figure 06 displays accurate predictions of the neural network algorithm compared to bottomhole pressure readings from memory gauge devices. This should make it possible to continuously monitor skin factors during the acid treatment and make the optimization decision instantaneous while pumping.

As illustrated in Figure 07, the real-time reservoir skin measurements from acid pumping through the coiled tube were determined using modified Paccaloni's Technique and Prouvost & Economides' Technique. The initial skin values calculated through modified Prouvost & Economides' technique were 3.5, and values recorded through modified Paccaloni's technique reached 6.2. The steady-state calculation method of the modified Paccaloni's Technique generates excessive estimations of both the real initial skin compared to pre-stimulation skin measurements by well test (+4) and maintains this overestimation throughout the entire treatment period. The steady-state pressure drop maintains a higher value than the pressure drop occurring in infinite-acting transient fluid movement. The modified Prouvost & Economides' technique calculated a skin value (3.5), which closely matched the well test result of actual skin (4). Paccaloni's approach shows different skin effect calculations because it assumes steady-state pressure, while Prouvost's approach bases its assumptions on transient pressure behavior. The calculated skin values from both analysis techniques decreased during acid pumping operations while the skin fluctuations were caused by the alternate use of CTD and AFR mechanisms.

Table 6. Well-X input parameters.

Parameter	Unit	Value
Coiled tubing depth	ft	6582
Coiled tubing ID	inches	1.25
Nitrogen rate at surface	scf/m	500
Bottomhole temperature at memory gauge	F	155
Reservoir permeability	md	5
Reservoir thickness	ft	102
Reservoir porosity	fraction	0.12
Reservoir pressure	psi	1030

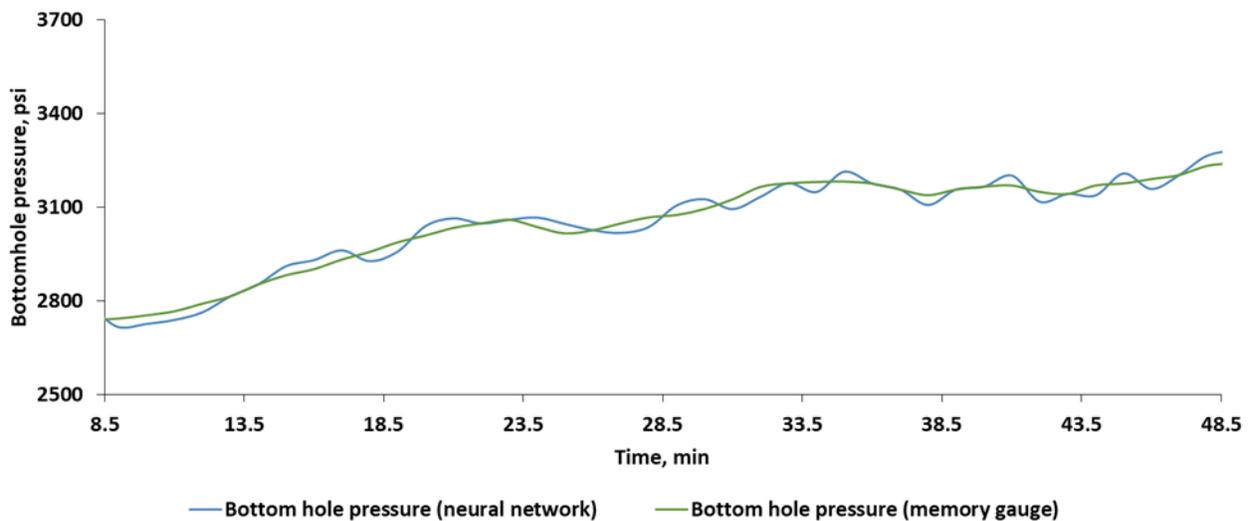


Figure 6. Bottomhole pressure for Well X.

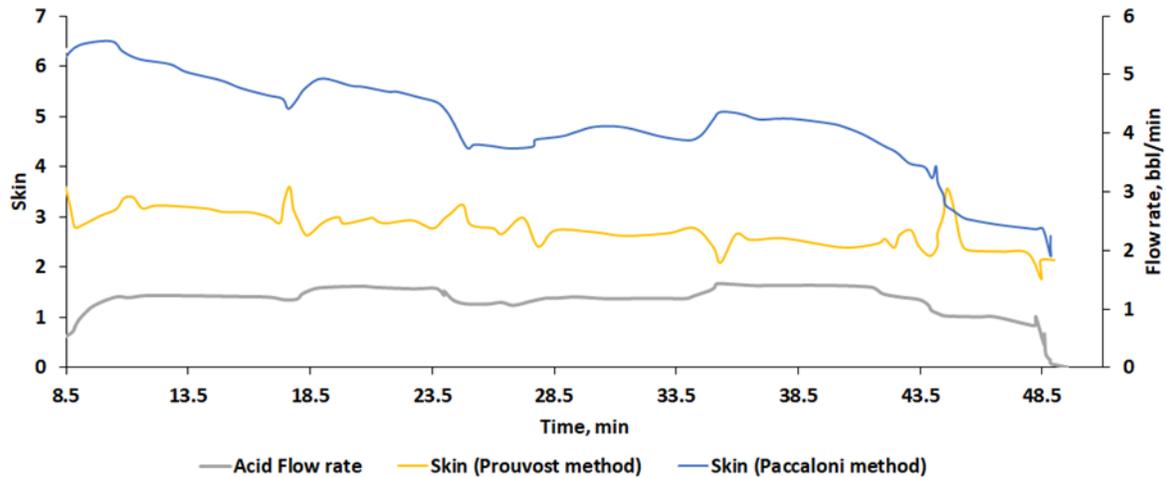


Figure 7. Real-time monitoring of foamed acid matrix acidizing in Well A.

#### 4. Conclusion

This research presents innovative machine learning techniques for predicting real-time bottom-hole pressure together with skin factor values in foamed acid matrix acidizing operations. The research incorporated 31 wells into a detailed dataset that allowed the neural network as well as the AdaBoost, Random Forest, and K-Nearest Neighbor (kNN) algorithms to achieve outstanding prediction accuracy through their low root mean square error (RMSE) results. The K-fold cross-validation process showed four optimum models to be robust with their respective RMSE results of 0.047, 0.048, 0.054, and 0.058, while  $R^2$  values reached 0.947, 0.943, 0.929, and 0.918. The developed models demonstrate reliability for real-time prediction in dynamic well stimulation environments because they produce exceptionally accurate results based on their low RMSE scores and high  $R^2$  values.

By offering real-time assessment of the effectiveness of stimulation, these machine learning models offer a transformative alternative to pre- and post-stimulation well testing, empirical multiphase flow correlations, mechanistic models, and unified models. This proves the capability of machine learning algorithms in optimizing the operations of acid treatment, in controlling the pump parameters with high precision, as well as improving the overall treatment efficiency. Moreover, real-time bottom hole pressure and skin factor predictions obtained from the above predictive models are in close agreement with reliable memory gauge and pre-stimulation well testing measurements during the field application on Well-X. The successful application of these models suggests that they are potentially suitable to replace downhole pressure gauges and thus reduce operating costs and speed up decision-making.

Using real-time data and machine learning systems will allow for the automatic improvement of how acidizing treatments are delivered. This work connects data-based knowledge with stimulation operations to create digital improvements for matrix acidizing treatments at a practical cost and effective results.

#### Nomenclature

$AFR$	Acid flow rate at surface
$B$	Formation volume factor, $rbbl/stb$
$BHP$	Pressure at a coiled tubing outlet
$BHT$	Temperature at a coiled tubing outlet
$C_t$	Total compressibility, $psi^{-1}$
$CTD$	Coiled tubing depth
$CTID$	Coiled tubing inside diameter
$CTP$	Coiled tubing pressure at surface
$DL$	Deep learning
$\Delta p$	Pressure difference between reservoir pressure and well flowing pressure, $psi$

$\Delta t$	Time difference, hour
DT	Decision Tree
$\Delta$	Porosity, fraction
$\gamma$	Euler constant
GB	Gradient boosting
$h$	Reservoir thickness, ft
KNN	K-Nearest Neighbor
LR	Linear Regression
$\mu$	Acid viscosity, cP
MAE	Mean absolute error
MAPE	Mean absolute percent error
ML	Machine learning
MSE	Mean square error
NFR	Nitrogen rate at surface
NN	Neural network
$P_i$	Initial reservoir pressure, psi
$P_{meas}$	Measured bottomhole injection pressure, psi
$P_{sim}$	Simulated pressure response, psi
$q_i$	Injection rate, bpm
$R^2$	Correlation coefficients
$r_b$	Acid bank radius, ft
$r_e$	Effective wellbore radius, ft
RF	Random Forest
RMSE	Root mean square error
$S$	Skin factor, dimensionless
SGD	Stochastic Gradient Descent
SHAP	Shapley additive explanations
$S_o$	Initial skin factor, dimensionless
SVMs	Support Vector Machines
$t$	Time, hour

## References

- [1] Pucknell JK, Clifford PJ. Calculation of Total Skin Factors; OnePetro, 1991. <https://doi.org/10.2118/23100-MS>
- [2] Yildiz T. Assessment of Total Skin Factor in Perforated Wells; OnePetro, 2003. <https://doi.org/10.2118/82249-MS>
- [3] Paccaloni G, Tambini M, Galoppini M. Key Factors for Enhanced Results of Matrix Stimulation Treatments; OnePetro, 1988. <https://doi.org/10.2118/17154-MS>
- [4] Restrepo A, Duarte J, Sánchez Y. A MultiParameter Methodology for Skin Factor Characterization: Applying Basic Statistics to Formation Damage Theory; OnePetro, 2007. <https://doi.org/10.2118/107674-MS>
- [5] Biu VT, Biu EO, Buduka S, Ajienska J, Onyekonwu M. Well and Reservoir Parameters Estimation [K and Skin] Using the Statistical Diagnostic Approach.Part II; OnePetro, 2010. <https://doi.org/10.2118/136963-MS>
- [6] Pichler T, Frick TP, Economides MJ, Nittmann J. Stochastic Modeling of Wormhole Growth in Carbonate Acidizing With Biased Randomness; OnePetro, 1992. <https://doi.org/10.2118/25004-MS>
- [7] da Motta EP, dos Santos JACM, Zhu D, Hill AD. Field Evaluation and Optimization of Matrix Acidizing Treatments; OnePetro, 1997. <https://doi.org/10.2118/37460-MS>
- [8] Qiu X, Zhao W, Chang F, Dyer S. Quantitative Modeling of Acid Wormholing in Carbonates-What Are the Gaps to Bridge; OnePetro, 2013. <https://doi.org/10.2118/164245-MS>
- [9] Prouvost LP, Economides MJ. Real-Time Evaluation of Matrix Acidizing Treatments. Journal of Petroleum Science and Engineering 1987; 1(2): 145–154. [https://doi.org/10.1016/0920-4105\(87\)90005-2](https://doi.org/10.1016/0920-4105(87)90005-2)
- [10] Zakaria ASS, Nasr-El-Din HAA, Ziauddin M. Predicting the Performance of the Acid-Stimulation Treatments in Carbonate Reservoirs With Nondestructive Tracer Tests. SPE Journal 2015, 20 (06), 1238–1253. <https://doi.org/10.2118/174084-PA>

- [11] Ahmed U, Watson JT. Use of Bottomhole Transient Rate Data To Enhance Transient Pressure Analysis; OnePetro, 1986. <https://doi.org/10.2118/15422-MS>
- [12] El-Khatib N. Evaluation of Formation Damage from Transient Pressure Analysis: Unsteady State Skin Factor; OnePetro, 2011. <https://doi.org/10.2118/140961-MS>
- [13] Akanni OO, Nasr-El-Din HA. The Accuracy of Carbonate Matrix-Acidizing Models in Predicting Optimum Injection and Wormhole Propagation Rates; OnePetro, 2015. <https://doi.org/10.2118/172575-MS>
- [14] Buijse MA. Understanding Wormholing Mechanisms Can Improve Acid Treatments in Carbonate Formations. SPE Production & Facilities 2000; 15 (03): 168–175. <https://doi.org/10.2118/65068-PA>
- [15] Gunaydin, D.; Mohaghegh, S. D.; Gaskari, R.; Aminian, K. Estimation of Skin Factor By Using Pressure Transient Testing Results of a Single-Rate Well Test; OnePetro, 2007. <https://doi.org/10.2118/111204-MS>
- [16] Olufisayo F, Anthony O. Production Optimization through Pressure Transient Analysis; OnePetro, 2019. <https://doi.org/10.2118/198787-MS>
- [17] Fredd CN, Miller MJ. Validation of Carbonate Matrix Stimulation Models; OnePetro, 2000. <https://doi.org/10.2118/58713-MS>
- [18] Furuï K, Burton RCC, Burkhead DWW, Abdelmalek NAA, Hill ADD, Zhu D, Nozaki M. A Comprehensive Model of High-Rate Matrix-Acid Stimulation for Long Horizontal Wells in Carbonate Reservoirs: Part II—Wellbore/Reservoir Coupled-Flow Modeling and Field Application. SPE Journal 2011; 17(01): 280–291. <https://doi.org/10.2118/155497-PA>
- [19] Heard SR, Economides MJ, Ehlig-Economides CA. Pressure and Rate Measurements to Validate Matrix Stimulation; OnePetro, 1990. <https://doi.org/10.2118/19409-MS>
- [20] Kuchuk F, Biryukov D. Pressure-Transient Behavior of Continuously and Discretely Fractured Reservoirs. SPE Reservoir Evaluation & Engineering 2014; 17 (01): 82–97. <https://doi.org/10.2118/158096-PA>
- [21] Thabet EA, Brady ME, Parsons CA, Byrne S, Voropaev SV, Lesko TM, Tardy PM, Cohen CE, Mchaweh A. Changing The Game In The Stimulation of Thick Carbonate Gas Reservoirs; OnePetro, 2009. <https://doi.org/10.2523/IPTC-13097-MS>
- [22] Zhu D. Treatment Evaluation and Real-Time Diagnosis. 2016. <https://doi.org/10.2118/9781613994269-07>
- [23] Zhu D, Hill AD, da Motta EP. On-Site Evaluation of Acidizing Treatment of a Gas Reservoir; OnePetro, 1998. <https://doi.org/10.2118/39421-MS>
- [24] Zhu D, Hill AD. Field Results Demonstrate Enhanced Matrix Acidizing Through Real-Time Monitoring; OnePetro, 1996. <https://doi.org/10.2118/35197-MS>
- [25] da Motta EP, dos Santos JACM, Zhu D, Hill AD. Field Evaluation and Optimization of Matrix Acidizing Treatments; OnePetro, 1997. <https://doi.org/10.2118/37460-MS>
- [26] Schiager BA, Frick TP. Real-Time Monitoring and Analysis of Matrix Stimulation Operations; OnePetro, 1996. <https://doi.org/10.2118/31104-MS>
- [27] MaGee J, Buijse MA, Pongratz R. Method for Effective Fluid Diversion When Performing a Matrix Acid Stimulation in Carbonate Formations; OnePetro, 1997. <https://doi.org/10.2118/37736-MS>
- [28] Al-Dhafeeri AM, Engler TW, Nasr-El-Din HA. Application of Two Methods to Evaluate Matrix Acidizing Using Real-Time Skin Effect in Saudi Arabia; OnePetro, 2002. <https://doi.org/10.2118/73703-MS>
- [29] Ma L, Johns RT, Zhu D, Hill AD. Fast Method for Real-Time Interpretation of Variable-Rate Wells with Changing Skin: Application to Matrix Acidizing; OnePetro, 2000. <https://doi.org/10.2118/64651-MS>
- [30] Rammay MH, Alnuaim S. Flow Regime Prediction Using Fuzzy Logic and Modification in Beggs and Brill Multiphase Correlation; OnePetro, 2015. <https://dx.doi.org/10.2523/IPTC-18267-MS>
- [31] Abdul Majeed GH, Abu Al-Soof NB, Alassal JR. An Improved Revision To The Hagedorn And Brown Liquid Holdup Correlation. Journal of Canadian Petroleum Technology 1989; 28 [06]. <https://doi.org/10.2118/89-06-02>
- [32] Opoku D, Al-Ghamdi A, Osei A. Novel Method to Estimate Bottom Hole Pressure in Multiphase Flow Using Quasi-Monte Carlo Method; OnePetro, 2020. <https://dx.doi.org/10.2523/IPTC-20080-MS>
- [33] Yahaya AU, Al Gahtani AA. Comparative Study Between Empirical Correlations & Mechanistic Models of Vertical Multiphase Flow; OnePetro, 2010. <https://dx.doi.org/10.2118/136931-MS>

- [34] Metin CO, Ozbayoglu ME. Two-Phase Fluid Flow Through Fully Eccentric Horizontal Annuli: A Mechanistic Approach; OnePetro, 2007. <https://dx.doi.org/10.2118/107076-MS>
- [35] Fan Y, Aljasser M. Unified Modeling of Pseudo-Slug and Churn Flows for Liquid Holdup and Pressure Gradient Predictions in Pipelines and Wellbores. SPE Production & Operations 2023; 38 (02): 350–366. <https://doi.org/10.2118/212863-PA>
- [36] Pauchon CL, Dhulesia H, Cirlot GB, Fabre J. TACITE: A Transient Tool for Multiphase Pipeline and Well Simulation; OnePetro, 1994. <https://dx.doi.org/10.2118/28545-MS>
- [37] Gomaa S, Mongy M, Emara R, Fahmy A, Attia A. Evaluating Medium Decision Tree Model, Support Vector Machine Rational Quadratic Gaussian Process Regression to Estimate the Total Organic Carbon of Shale Gas Reservoirs. Petroleum & Coal 2024; 6 (1): 122-131.
- [38] Thabet S, Zidan H, Elhadidy A, Taman A, Helmy A, Elnaggar H, Yehia T. Machine Learning Models to Predict Production Rate of Sucker Rod Pump Wells. In SPE Western Regional Meeting; SPE: Palo Alto, California, USA, 2024; p D011S004R008. <https://doi.org/10.2118/218857-MS>
- [39] Khan AM, Ugarte E, Kurniadi S, Al Shueili A, Wang WK. Physics-Informed Machine Learning for Hydraulic Fracturing—Part III: The Transfer Learning Validation. In International Petroleum Technology Conference; IPTC: Kuala Lumpur, Malaysia, 2025; p D022S005R007. <https://doi.org/10.2523/IPTC-24850-MS>
- [40] Alquraini AH, Al Sadah HH, AlBori MM, Al-Kadem MS. The Utilization of Machine Learning and Deep Learning to Predict Real-Time ESP Well Status. In SPE Artificial Lift Conference and Exhibition - Americas; SPE: The Woodlands, Texas, USA, 2024; p D021S004R004. <https://doi.org/10.2118/219541-MS>
- [41] Thabet SA, Elhadidy AA, Heikal M, Taman A, Yehia TA, Elnaggar H, Mahmoud O, Helmy A. Next-Gen Proppant Cleanout Operations: Machine Learning for Bottom-Hole Pressure Prediction. In Mediterranean Offshore Conference; SPE: Alexandria, Egypt, 2024; p D021S012R008. <https://doi.org/10.2118/223373-MS>
- [42] Kalule R, Abderrahmane HA, Ahmed S, Hassan AM, Alamer, W. Advancing Relative Permeability and Capillary Pressure Estimation in Porous Media Through Physics-Informed Machine Learning and Reinforcement Learning Techniques. In International Petroleum Technology Conference; IPTC: Dhahran, Saudi Arabia, 2024; p D031S143R003. <https://doi.org/10.2523/IPTC-23572-MS>
- [43] Ji L, Gordillo R, Xiao Y, Al Salmi S, Al Qasabi N. Rapid Field Development Assessment and Well Performance Evaluation in a Conventional Reservoir by Machine Learning and Data Analytics. In SPE Conference at Oman Petroleum & Energy Show; SPE: Muscat, Oman, 2025; p D021S030R001. <https://doi.org/10.2118/225038-MS>
- [44] Al Kiyumi SK, Al Balushi M. Real-Time Machine Learning for Advanced Beam-Pump Dynacard Surveillance and Predictive Analytics: Opportunities and Challenges. In SPE Conference at Oman Petroleum & Energy Show; SPE: Muscat, Oman, 2025; p D011S012R004. <https://doi.org/10.2118/224979-MS>
- [45] Thabet S, Elhadidy A, Elshielh M, Taman A, Helmy A, Elnaggar H, Yehia, T. Machine Learning Models to Predict Total Skin Factor in Perforated Wells. In SPE Western Regional Meeting; SPE: Palo Alto, California, USA, 2024; p D011S004R007. <https://doi.org/10.2118/218838-MS>
- [46] Mal A, Ødegård SI, Helgeland S, Zulkhifly Sinaga S, Svendsen M. Prediction of Stuck Pipe Incidents Using Models Powered by Deep Learning and Machine Learning. In IADC/SPE International Drilling Conference and Exhibition; SPE: Galveston, Texas, USA, 2022; p D022S005R002. <https://doi.org/10.2118/208778-MS>
- [47] Arukhe JO, Ghamdi KA, Anazi AF. Machine Learning Framework for Treating and Preventing Scale Formation. In Offshore Technology Conference; OTC: Houston, Texas, USA, 2025; p D041S056R002. <https://doi.org/10.4043/35618-MS>
- [48] Gaurav K. Developing a Machine Learning Application for Well Production Forecasting in Frequently Changing Operating Conditions. In Offshore Technology Conference; OTC: Houston, Texas, USA, 2025; p D022S059R003. <https://doi.org/10.4043/35680-MS>
- [49] Thabet SA, El-Hadydy AA, Gabry MA. Machine Learning Models to Predict Pressure at a Coiled Tubing Nozzle's Outlet During Nitrogen Lifting. In SPE/ICoTA Well Intervention Conference and Exhibition; SPE: The Woodlands, Texas, USA, 2024; p D021S013R003. <https://doi.org/10.2118/218294-MS>
- [50] Rogulina A, Zaytsev A, Ismailova L, Kovalev D, Katterbauer K, Marsala A. Similarity Learning for Well Logs Prediction Using Machine Learning Algorithms. In International Petroleum Technology Conference; IPTC: Riyadh, Saudi Arabia, 2022; p D032S158R005.

- [51] <https://doi.org/10.2523/IPTC-22067-MS>  
Nashed S, Lnu S, Guezei A, Ejehu O, Moghanloo R. Downhole Camera Runs Validate the Capability of Machine Learning Models to Accurately Predict Perforation Entry Hole Diameter. *Energies* 2024; 17(22), 5558. <https://doi.org/10.3390/en17225558>
- [52] Sheraz A, Khan S, Abdelhamid N, Manzoor A. Application of Machine Learning for Comprehensive Predictive Modelling of Drilling Key Performance Indicators Using Historical Drilling Data – Collaborative Case Study between Academia and Industry. In SPE Conference at Oman Petroleum & Energy Show; SPE: Muscat, Oman, 2025; p D021S030R003. <https://doi.org/10.2118/225014-MS>
- [53] Mwakipunda GC, Komba NA, Ayimadu ET, Yu L. Model Development for Brittleness Index Estimation and Depth Determination in Hydraulic Fracturing Operations in Shale Gas Reservoirs Using Machine Learning Algorithms. *SPE Journal* 2025, 1–22. <https://doi.org/10.2118/228285-PA>
- [54] Al-Sulaimi G, Al-amri M. Advanced Machine Learning for Accurate Reservoir Characterization: Predicting Fluid Saturations in Gharif Reservoir Using a Custom Ensemble Model. In SPE Conference at Oman Petroleum & Energy Show; SPE: Muscat, Oman, 2025; p D021S018R001. <https://doi.org/10.2118/224941-MS>
- [55] Nashed S, Moghanloo R. Replacing Gauges with Algorithms: Predicting Bottomhole Pressure in Hydraulic Fracturing Using Advanced Machine Learning. *Eng.*, 2025; 6(4): 73. <https://doi.org/10.3390/eng6040073>

---

*To whom correspondence should be addressed: Nermeen Hana, Ain Shams University, Cairo, Egypt,  
E-mail: [nermeen.hana.researcher@gmail.com](mailto:nermeen.hana.researcher@gmail.com) ORCID: 0009-0004-6487-0328*