https://doi.org/10.48048/tis.2023.7036

A Case Study for Evaluating Effective Geomechanical Parameters and the Influence of the Biot Coefficient on the Stability of a Wellbore Wall in the Asmari Formation using Machine Learning

Farzad Fahool[1,*], Reza Shirinabadi[2,3] and Parviz Moarefvand[1]

[1]Department of Mining Engineering, Amirkabir University of Technology, Iran
[2]Department of Petroleum and Mining Engineering, South Tehran Branch, Islamic Azad University, Iran
[3]Research Center for Modeling and Optimization in Science and Engineering, South Tehran Branch, Islamic Azad University, Iran

([*]Corresponding author's e-mail: ffahool@aut.ac.ir)

Received: 30 April 2023, Revised: 2 June 2023, Accepted: 9 June 2023, Published: 3 September 2023

Abstract

In oil extraction projects, knowledge of reservoir geomechanics is essential for estimating the stability of wellbore walls drilled at great depths. In this regard, mechanics of porous media according to the definition of Biot's coefficient instead of Terzaghi effective stress can provide a more accurate estimate compared to other analyses. Additionally, using artificial intelligence and machine learning algorithms such as XGBoost and optimizing it with algorithms like Bayesian, along with using SHAP algorithm as an interpretable AI model, can provide us with deeper insights into available data. In this research, 1st geomechanical data of a well in Asmari formation in southwest Iran was obtained through well logs and operational reports and then analyzed by machine learning. Also, a 1-way coupled reservoir rock-fluid model was built to investigate the volume of fractured rock around the well. The interpretation of machine learning results helps us better understand the parameters affecting instability in this well's wall. Moreover, finite element model results indicate that assuming a value equal to 1 for Biot coefficient (or Terzaghi effective stress) leads to incorrect results and overestimates the volume of fractured rock around this well up to 13 times more than actual values. Therefore, any proper analysis regarding wellbore wall stability and evaluating effective stresses requires accurate knowledge about real and existing values of this coefficient due to simultaneous behavior of stress and pore pressure.

Keywords: Oil well stability, Machine learning, XGBoost algorithm, Bayesian optimization, SHAP algorithm, Porous elasticity, Stress-pore pressure coupling, Mohr-Coulomb shear failure, Biot coefficient

Introduction

In many geological processes and drilling problems in oil wells, the in-situ stresses and pore pressure, beside the other parameters play an important role, including determining the migration paths of hydrocarbons and the stability of wellbore walls. In fields facing hydrocarbon migration, these stresses and their coupling with pore pressure and changes in effective stress on the drilled wellbore walls can be so significant that they cause wellbore collapse. As a result, controlling the amount and direction of in-situ stresses and pore pressure beside the other parameters, is one of the main factors in well drilling, extraction and completion, especially in reservoir production simulation operations [1,2]. Therefore, having complete information about them during oil well drilling (especially in areas with active tectonics where natural fractures exist) is essential. Experts in well logging and drilling have proposed relationships to determine minimum and maximum horizontal stresses in various parts of the world [3,4]. In most of these relationships, minimum and maximum horizontal stresses are calculated by well logging (e.g., examining compressional and shear wave velocity) while minimum horizontal stress can be obtained directly from formation leak-off test data (LOT), hydraulic fracturing (HF), micro fracturing and mud weight lost circulation for specific points in the wellbore [4,5]. However, no specific relationship has been proposed for determining maximum horizontal stress, which is usually calculated using Anderson's fracture theory or poroelasticity relationships [5,6]. In the past years a lot of great researches have been done to evaluate the stability of an oil well, and it should be mentioned that most of them are based on well-logging geological models. In these researches correlation between instability of the well and mechanical parameters of the reservoir rock and its fluid content have been investigated based on analytical approaches [7,8].

Machine learning is a novel computational approach that is widely used in various fields of engineering. Appling this method to the available geomechanical data of an oil well can be a precise and useful manner for evaluating the effective parameters on the stability of an oil well wall in the petroleum industry [9,10]. But among of the regression methods, the XGBoost algorithm has strong advantages in finding the correlation between parameters and it can be optimized with many algorithms such Bayesian. Also, this method has the ability to be interpreted by an interpretable artificial intelligence like SHAP algorithm.

Since changes in stress and pore pressure in a porous medium such as oil reservoirs have simultaneous effects on each other, using poroelasticity relationships can provide a more accurate solution for geomechanical analysis of reservoirs. This is because in each section of the formation, porosity values, saturation degree and naturally Biot coefficients may vary; therefore, effective stress and pore pressure may also have different values that can lead to variable conditions regarding wall stability and potential collapse compared to rock mass strength. As will be seen in the following discussion on the sensitivity of wellbore stability to the value of the Biot coefficient, determining this behavior is crucial for analyzing a highly permeable environment and addressing engineering issues and necessary cost savings in geomechanical projects, such as oil well stability. Geomechanics of hydrocarbon reservoirs is actually an investigation of rock behavior based on elasticity theory under conditions where the environment must be considered porous and the temperature variable to some extent. This concept is called thermoporoelasticity in geomechanics [11].

Various factors such as in-situ stresses, formation rock strength, formation fluid pressure, drilling mud pressure and wellbore trajectory are involved in wellbore stability. Among these factors, wellbore trajectory and drilling mud pressure range are under our control and can be changed to prevent problems [12,13]. Mechanical processes in porous media (such as rocks) consist of 2 basic parts: Fluid flow and rock displacement. In reservoir simulations, changes in fluid pressure resulting from injection or production disrupt the state of stress in rocks [14]. Changes in stress state drive rock displacement and conversely affect fluid flow processes. To describe the mutual effect between flow and deformation, the concept of effective stress (total stress on rock minus fluid pressure) was introduced by Terzaghi in 1923. This concept was later corrected by Biot in several papers [15]. Mathematical models for fluid flow in elastic porous media have been developed based on specific assumptions. The 2 fundamental principles used in these cases are the conservation of mass and momentum balance, for both fluid and rock phases. The theory of poroelasticity forms the basis for simulating and estimating stress and strain distributions in the oil and gas fields [15]. It is worth mentioning that a poroelasticity-based model can be solved in 2 ways: One is the non-coupled or uncoupled method, and the other is the coupled modeling approach. In the coupled method, poroelastic and fluid dynamic equations are simultaneously used to determine unknown values of pore pressure and medium displacements, and the analysis is based on time. Whereas in the non-coupled method, after obtaining the fluid pressure distribution inside or around the reservoir by a flow simulator, stress and strain distributions are separately solved using a poroelastic model. In the non-coupled method, pore pressure gradients are used to determine strains, and analysis will not be based on time [16,17].

This article discusses the evaluation of effective parameters on the stability of a wellbore wall using real data and machine learning in the 1st s. In the 2nd s, the concepts of pore pressure-stress interaction and effective stress in the shear failure of fluid-containing rock are investigated using the relationships of poroelasticity at various values of the Biot coefficient. Initial values of in-situ stresses, pore pressure, elastic parameters of rock and fluid surrounding a horizontal well in Asmari formation located in southwest Iran (Mansouri oil field) were obtained from technical reports and well logs using Interactive Petrophysics software (IP). Subsequently, these data were analyzed using machine learning and XGBoost regression algorithm whose hyper parameters were optimized by an optimization algorithm such as Bayesian. Then, the output of the optimized XGBoost model was examined by SHAP algorithm as an interpretable artificial intelligence model to identify the most influential parameters affecting wellbore stability using real data. Additionally, a 3D model in Comsol Multiphysics finite element software was built for a 1-way coupled reservoir rock-fluid system under stationary conditions to investigate the volume of fractured rock around the well using Mohr-Coulomb failure criterion at different values for the Biot coefficient. Finally, by comparing caliper and bit size logs with values calculated by Comsol Multiphysics software, research was conducted on the relationship between wellbore stability and volume of fractured rock at a specific length of the well at various amounts of the Biot coefficient.

Materials and methods

Data collection and analysis using machine learning

The IP software interprets well log data using available equations and functions, as well as information from various logs such as resistivity, radioactivity, sonic, caliper and correction curves, and combines them to provide a comprehensive interpretation of the logs. The software can use estimated models to calculate properties such as porosity, water saturation, shale volume, and other properties defined by the user for the different zones. It also has various modules for evaluating geomechanical parameters of the reservoir rock and the fluid within it, including Geomechanics, Rock Mechanics, Rock Physics and Pore Pressure Calculation. By using these modules and making necessary corrections along with other field information, all necessary geomechanical parameters and boundary conditions can be extracted for conducting a poroelastic analysis or for use in machine learning. In this paper, the necessary data for machine learning and evaluating parameters affecting wellbore instability, as well as constructing a poroelastic model for an oil well in its horizontal section under 2 different stress regimes (normal and reverse); have been obtained as shown in **Figure 1** and **Table 1**. It should be noted that these parameter values were collected using well logging data analyzed by IP software, as well as geological reports, drilling operations and the designed mud window for this well.

The well was originally drilled vertically at (E: 1956108.00, N: 902229.00 in Lambert projection system, Nahrwan UAE, Zone: Iraq) located in the central part of Mansouri structure. The well was spudded on 24 March 1970, drilling operation ended on 16 June 1970 and the well was drilled to the total depth of 3,003 m (RKB = 17 m) in to upper Sarvak and have been producing from Ghar: 2,334 - 2,341.5 m, Asmari (Dolomite): 2,435 - 2,443 m and Sarvak: 2,939 - 2,982.5 m. This well originally was planned to be a re-entry well from main well with multi-lateral horizontal legs (3 legged), but actually it was drilled in 2 legs (i.e. main leg and leg#1) in Sarvak and directed to the south of the original well. The horizontal section spudded on 1,382, 11, 16 (31 January 2004) in order to turn the vertical main well to a horizontal multi-lateral oil producer in the central part of the Mansouri structure. Stratigraphic section from Jahrum to Sarvak formation (Eocene to middle Cretaceous) were drilled in this well, which mainly consists of limestone deposited in a shallow marine environment which base on the fluctuation of the sea level some intervals are interbedded with shale, argillaceous limestone with chert nodules. A brief lithological description is the same as follows: A - Jahrum Formation (Eocene) from: 2,651.5 - 2,748.4 m with Interbedded chalky limestone, with dolomite, brown, fine grain, vuggy, in the lower section it consists of limestone light brown, with fossil and nodules and lenses of chert, B-Pabdeh Formation (Paleocene) from: 2,748.4 - 2,947 m with Limestone, grey, brown, very argillaceous with nodules and lenses of chert, C-Gurpi Formation (Upper Cretaceous) from: 2,947 - 3,011 m with Limestone gray to light gray, soft, in parts argillaceous with inter-beds of shale and marl and D-Sarvak (Middle Cretaceous) from: 3,011 - 3,560 m (Main Leg), 3,223 - 3,526 m (Leg-1) with Limestone light brown, fossil-ferrous with Rudists and benthonic foraminifera, dolomite light brown, crystalline, vuggy.

In **Figure 1**, the 1st column shows a section of the well length in the horizontal part of it, while the 2nd column contains 3 logs of gamma ray, compressional wave velocity and shear wave velocity. The 3rd column displays relevant logs for in-situ stress fields such as vertical, minimum and maximum horizontal stresses. The 4th column includes logs related to mechanical properties of rocks such as static Young's modulus, Poisson's ratio, uniaxial compressive strength, internal friction coefficient and bulk density of rock. The 5th column discusses parameters related to rock-fluid interaction including permeability, Biot coefficient, pore pressure, porosity and degree of saturation. The 6th column presents the difference between the sizes of drill bit and caliper log to indicate unstable areas in the wellbore wall. The 7th column describes the volume content of reservoir fluid in rocks while the last column indicates the lithology of the studied well section.

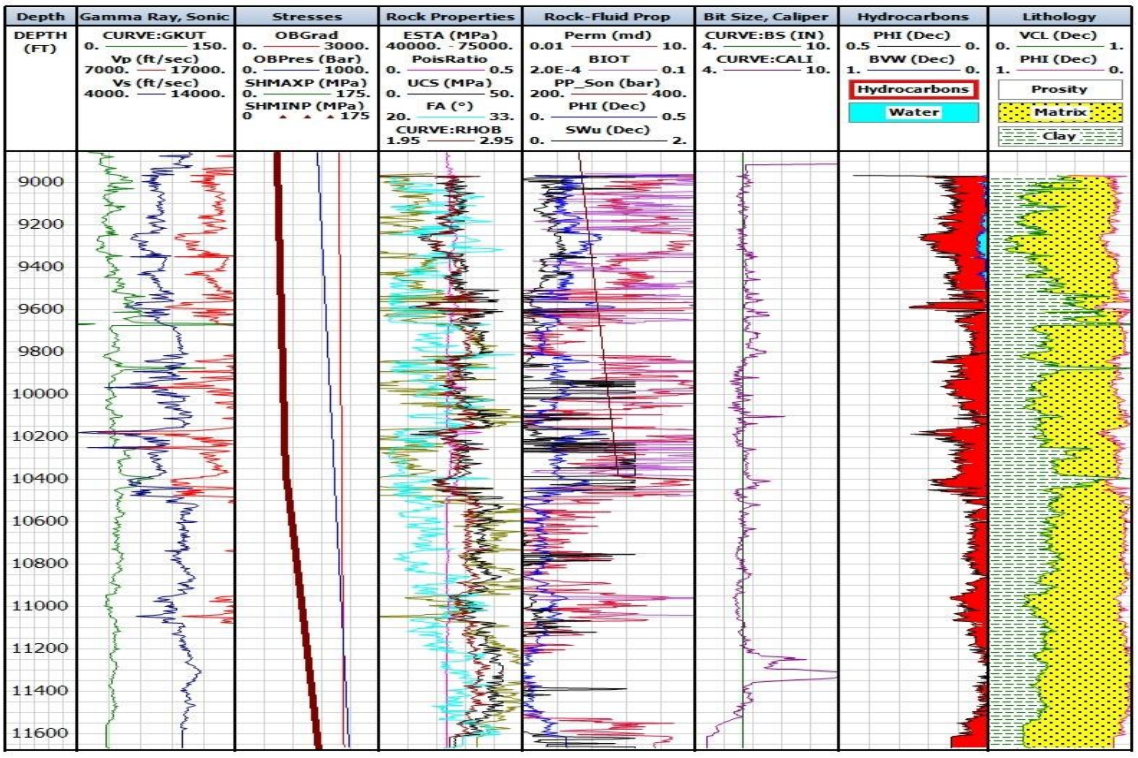


Figure 1 Geomechanical parameters of rock and reservoir fluid in the horizontal section of well. The depth column indicates the length of the well and it is not true vertical depth or TVD.

Table 1 Description of the geomechanical parameters of rock and reservoir fluid in the horizontal section of the well in 2 different stress regimes.

Symbol	Unit	Min		Max		Average		Standard deviation	
		Normal	Reverse	Normal	Reverse	Normal	Reverse	Normal	Reverse
Sigma V (Vertical Stress)	MPa	57.02	69.26	69.25	76.67	63.11	72.95	3.53	2.140
Sigma H (Max Horizontal Stress)	MPa	50.6	69.26	69.24	100.9	57.08	85.26	4.42	9.18
Sigma h (Min Horizontal Stress)	MPa	50.6	69.26	69.24	100.9	57.08	85.26	4.42	9.18
E (Elastic Modulus)	Gpa	7.68	27.74	86.13	85.78	50.6	67.83	12.45	7.66
PoisRatio (Poisson's Ratio)	1	0.2325	0.2322	0.3645	0.2805	0.2477	0.2367	0.0131	0.0036
UCS (Uniaxial Compression Strength)	MPa	9.71	24.64	42.99	42.99	29.36	36.01	6.28	4.478
FA (Coulomb Friction Angle)	deg	20.87	21.3	33.81	31.11	24.30	26.07	2.53	2.08
RHO (Density of Rock)	Kg/m^3	1.70	2.422	2.74	2.808	2.47	2.57	0.1024	0.060
Perm (Permeability)	m^2	0	0	1E - 12	6.36E - 15	5E - 15	0.3E - 15	50.49E - 15	0.76E - 15
PHI (Porosity)	1	0.0001	0.0001	0.50	0.14	0.1039	0.051	0.0520	0.0337
BIOT (Dynamic Biot Coefficient (α))	1	0.0004	0.0002	0.7283	0.3891	0.1616	0.0811	0.1284	0.0906
PP (Pore Pressure)	Bar	264.8	4,282.5	4,273	25,294	538.44	14,256	775.81	6,095
ERROR (Bit Size-Caliper)	in	–1.816	–5	0.698	1.526	–0.079	–0.078	0.2735	0.9975

Data analysis using machine learning (Pearson correlation matrix, XGBoost algorithm optimized by Bayesian algorithm, and interpretation of results using SHAP algorithm)

In recent decades, the use of modern computational methods has been increasingly popular. Machine learning and deep learning are among these modern methods in geomechanics, and research in this field includes references to articles [18-20] in the area of stability of oil well walls. The aim of this section of the article is to present a machine learning-based method using the XGBoost algorithm to determine the parameters that affect the stability of the horizontal part of an oil well based on a series of real data obtained in the previous section. Therefore, after calculating the Pearson correlation coefficients on the data, the XGBoost algorithm is briefly examined, and in order to achieve optimal results from the fitting by machine learning, the Bayesian optimization algorithm is used to adjust the input hyper parameters of the XGBoost algorithm, which will be briefly explained. Finally, the final output of the optimized XGBoost algorithm is presented using an interpretable artificial intelligence method or the SHAP algorithm.

Spearman correlation matrix

The range of geomechanical parameters considered in these analyses is provided in **Table 1**. Additionally, this table presents the range of differences between 2 log, bit size and caliper (as a characteristic for instability and wall failure of the well) that will be considered as ERROR hereafter. In data science, the relationship between 2 quantitative variables is measured by calculating the correlation coefficient. Two common types of these methods are Pearson and Spearman correlation, both of which have values between 1 and –1 for their coefficients. Thus, if the correlation coefficient is close to or equal to 1, there is a strong and positive relationship between the 2 variables. Since the Pearson correlation coefficient is calculated based on the mean and variance, it may be biased against outliers and not accurately reflect the degree of correlation. In such cases, the Spearman rank correlation coefficient is used. The Spearman rank correlation coefficient, like the Pearson coefficient, indicates how much a variable tends to follow other variables [21]. The Spearman correlation coefficient of the parameters listed in **Table 1** in 2 different sections of the well, which are related to 2 different stress regimes, is presented in **Figure 2**, using approximately 73,000 real data of the well. The last row of these figures shows the correlation coefficients of ERROR values with other reservoir parameters. Based on these values, it can be said that in both stress regimes, an increase in parameters such as pore pressure, Biot coefficient, porosity, permeability, Poisson's ratio and in-situ stress is directly related to instability in the well wall and an increase in them causes fractures and changes in shapes. On the other hand, an increase in parameters such as elasticity modulus, uniaxial compressive strength and density of the intact rock has an inverse relationship with the level of instability in the wellbore walls. An increase in these parameters leads to a decrease in the occurrence of fractures and deformations.

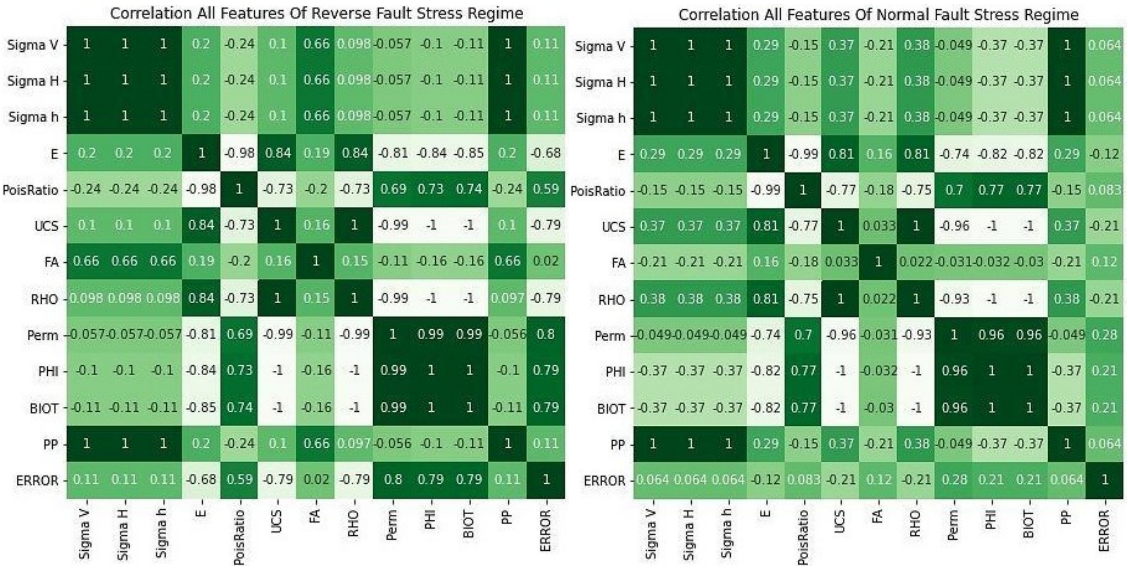


Figure 2 The Spearman correlation coefficients of geomechanical parameters of the wellbore. The ERROR parameter represents the difference between 2 logs (bit size and caliper) as an indicator of instability and failure in the wellbore walls.

XGBoost algorithm

In recent years, various machine learning models have been introduced, among which the XGBoost algorithm was proposed by Chen and Gastrin [22]. This method is a gradient boosting algorithm based on decision tree methods and has been widely used in various engineering fields due to its regularization and parallel processing, which has shown very good performance. The main parameters of this algorithm can be determined by other optimization algorithms (such as the Bayesian algorithm used in this article), and it allows the user to customize the objective function. Additionally, it requires less feature analysis. Since XGBoost implements parallel processing, it has a low computational cost, and it has low bias and variance. This algorithm also has the ability to deal with outliers and manage them. Based on the availability of labeled data, machine learning can be classified into supervised, unsupervised and boosted learning. However, most of these models do not provide any insight into the correlation between the available parameters and the importance of the studied variables.

The XGBoost algorithm cannot be classified as an interpretable AI model. A comprehensive interpretable AI model should be able to evaluate the correlation between each registered variable and the output class, demonstrates the correlation between them, measure the effectiveness of their importance, and rank them according to their importance [23]. Therefore, in this article, the SHAP algorithm has been used to interpret the results of the XGBoost algorithm. SHAP, as a major breakthrough in interpretable AI, can provide all necessary evaluations for different machine learning models and improve transparency and reliability [24]. Based on the game theory approach, SHAP can provide a fair output for each interpretable AI model and explain the importance of each feature in the output classification [25]. This algorithm can be combined with a boosted gradient model (such as XGBoost) to create a desirable model for various data evaluations. The steps of the analysis using machine learning algorithms in this article are illustrated in **Figure 3**(**A**).

The XGBoost algorithm used in this article is an enhanced gradient boosting method that provides a boosted parallel tree to solve regression and classification problems. Assuming we have a dataset $D = \{(x_i, y_i)\}: i = 1, \ldots, n\}$ with n samples and m features$(x_i \in R^m, y_i \in R)$ and R represents real numbers, the proposed tree model in this method uses an additional function z to approximate the model output, as follows [26]:

$$y_p = \phi(x_i) = \sum_{z=1}^{Z} f_z(x_i), f_z \in F \quad (1)$$

In this equation y_p is the target y_i, F is the problem space consisting of fitted decision trees, which is defined as follows [26]:

$$\mathrm{F} = \{ f(x) = w_{q(x)}\}(q: R^m \to T, w \in R^T) \quad (2)$$

In the above equation, q represents the structure of the decision tree, w is the weight of each leaf in the decision tree, and T represents the number of leaves in the tree. In this equation, f is a function that depends on the values of q, and the weights w are dependent on each of the decision trees. To optimize the set of decision trees and reduce the error, the task of the algorithm is to minimize the objective function as follows [26]:

$$L(t) = \sum_{i=1}^{n} l\left(y_e, y_p^{(t-l)} + f_t(x_i)\right) + (f_t) \quad (3)$$

In the above equation y_e is the prediction y_i and l is the loss function used to calculate the difference between y_e and y_p, and t is the iteration number for minimizing the error. The function (f_t) is also a penalty function used to reduce the complexity of the model fit. In this article, the XGBoost algorithm is implemented using the free library written for various programming languages and used in the Python programming language [27].

Optimizing XGBoost algorithm

In machine learning algorithms, there are 2 types of variables usually referred to as parameters and hyper parameters. Parameters are variables that are obtained during the minimization of the objective function using the training dataset. One example of such a variable is w in Eq. (2). On the other hand, hyper parameters are variables that must be set by the user before executing the algorithm. By changing hyper parameters, the performance of the machine learning model can be improved. Various methods have been proposed for determining hyper parameters, including grid search, random search and the use

of optimization methods. Bayesian optimization is one of the optimization methods that have been used in various research studies, such as [28-30], to adjust hyper parameters of machine learning algorithms. In this article, the free Bayes-opt library, which is written for the Python programming language, has been used for Bayesian optimization [31].

The input and output data of each problem have different units and ranges of variation. Therefore, to achieve satisfactory results in machine learning algorithms, the dataset must be scaled to a specific range. In this case, the dataset has been scaled to the range of 0 to 1. Different performance metrics have been proposed for evaluating the performance of various machine learning models, including the root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE) and linear correlation coefficient (R2) [32]. In this article, 4 main metrics, MSE, RMSE, MAE and R2, have been used to evaluate the performance of the model.

One of the problems with machine learning algorithms is the possibility of overfitting. In this case, as shown in **Figure 3**(**B**), the dataset should be randomly divided into 2 sets of training and testing. Before optimization, important hyper parameters of the algorithm must be identified and the range of their variations for optimization should be specified. The XGBoost algorithm has a total of 35 hyper parameters [27]. In this study, 7 categories of the most important hyper parameters of this algorithm have been selected and their default values are given in **Table 2**. Usually, during hyper parameter optimization, cross-validation is used to increase accuracy. In this method, as shown in **Figure 3**(**B**) for 5 sets, the dataset required for optimization is divided into *k* sets and the algorithm is run *k* times. In each iteration, one of the *k* selected sets is assumed to be the test data and the remaining sets are considered as training data. In each iteration, the optimization algorithm is first trained with *k-1* sets and then the algorithm performance metric (for example, MSE) is calculated using one of the remaining sets. Finally, the output performance metric of the algorithm in each optimization iteration is the average of *k* algorithm runs. This is done to reduce the effect of data selection in optimization iterations. In this article, *k* is considered to be equal to 5.

Table 2 Comparison of final values of optimized and unoptimized performance metrics along with hyper parameters adjusted by the Bayesian algorithm.

		Method			
		XGBR Base		**XGBR final**	
		Normal	**Reverse**	**Normal**	**Reverse**
	MSE	0.01030	0.01174	0.00209	0.00017
	RMSE	0.1015	0.1084	0.0458	0.0133
	MAE	0.0701	0.0729	0.0285	0.0087
	R2	0.1195	0.4543	0.821	0.9917
Hyper Parameters	**learning_rate**	0.1	0.1	0.110982	0.053912
	n_estimators	10	10	500	128
	max_depth	2	2	15	15
	eta	0.3	0.3	0.35631	0.01
	subsample	0.5	0.5	0.65838	0.736262
	colsample_bytree	0.2	0.2	0.582746	0.648616
	min_child_weight	5	5	11	14

Results and discussion

SHAP algorithm

Interpretability of a model is a key issue for machine learning methods. In this regard, SHAP is a game-theoretic based framework that provides a unified approach for interpreting the predictions of machine learning models and determines the importance of each feature in the input variables. SHAP is a method for interpreting the predictions of machine learning models using SHAP values proposed by

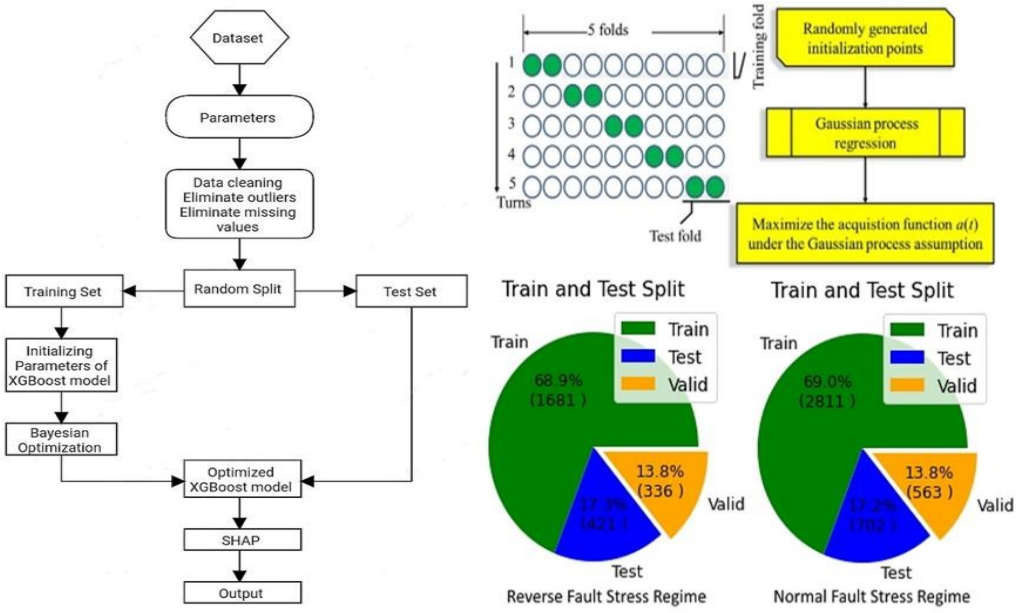
Lundberg and Lee. The SHAP values for each feature are the average marginal contribution of all permutations of that feature, indicating its impact on the output generated. In this interpretive model, *g(z)* is a linear add-on of the input features and is expressed according to the following formula [33]:

$$f(x) = g(z') = \varphi_0 + \sum_{i=1}^{M} \varphi_i z_i' \quad (4)$$

where z' is the simple input and M represents the number of features, φ_0 is a constant and z_i' is a binary variable indicating whether feature I is observed or not. Given the model f, the values of φ_i for each input feature can be computed according to the following formula, where S is a set of non-0 indices in z' and N denotes the set of all input features [34]:

$$\varphi_i = \sum_{S \in N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

In the plots shown in **Figure 4**, the results of this process on well data are displayed for both normal and reverse fault stress regimes. Based on the processing steps shown in **Figure 3**(**A**), the final output of the SHAP algorithm, which is an interpretation of the XGBoost algorithm optimized for well data, indicates the impact of each of the parameters in **Table 1** on the amount of ERROR (the difference between the values of 2 log, bit size and caliper measurements), which also represents the amount of well borehole failure and instability at different depths. As shown in **Figures 4**(**A**) - **4**(**B**), in the normal fault stress regime, parameters such as vertical stress, maximum and minimum horizontal stresses, pore pressure, etc., have the maximum impact on well borehole instability, and the values shown in each bar chart (**Figures 4**(**A**) - **4**(**C**)) indicate the degree of influence of the corresponding parameter on well borehole deformation in inches (all positive or negative values are shown as positive). Also, in the chart on the right (**Figures 4**(**B**) - **4**(**D**)), which is known as the "beeswarm" chart, the horizontal axis shows the output values of the SHAP algorithm, where higher values represent a greater impact of that parameter on the ERROR or deformation of the wellbore wall, similar to a bar chart. The difference is that the red color indicates the impact of higher values of that parameter on the SHAP value, and the blue color indicates the impact of smaller values of that parameter on the SHAP value. The charts in **Figure 4**(**C**) - **4**(**D**), which are for the inverse fault stress regime, are also interpretable, similar to the previous charts. Additionally, based on the results shown in **Figure 2**, which represents the correlation coefficients of the well data, it can be concluded that there is a strong and significant similarity between these correlation coefficients and the results obtained from machine learning on this data.



(A) (B)

Figure 3 Steps of machine learning (A) and random division of the dataset and optimization using 5-fold cross-validation (B).

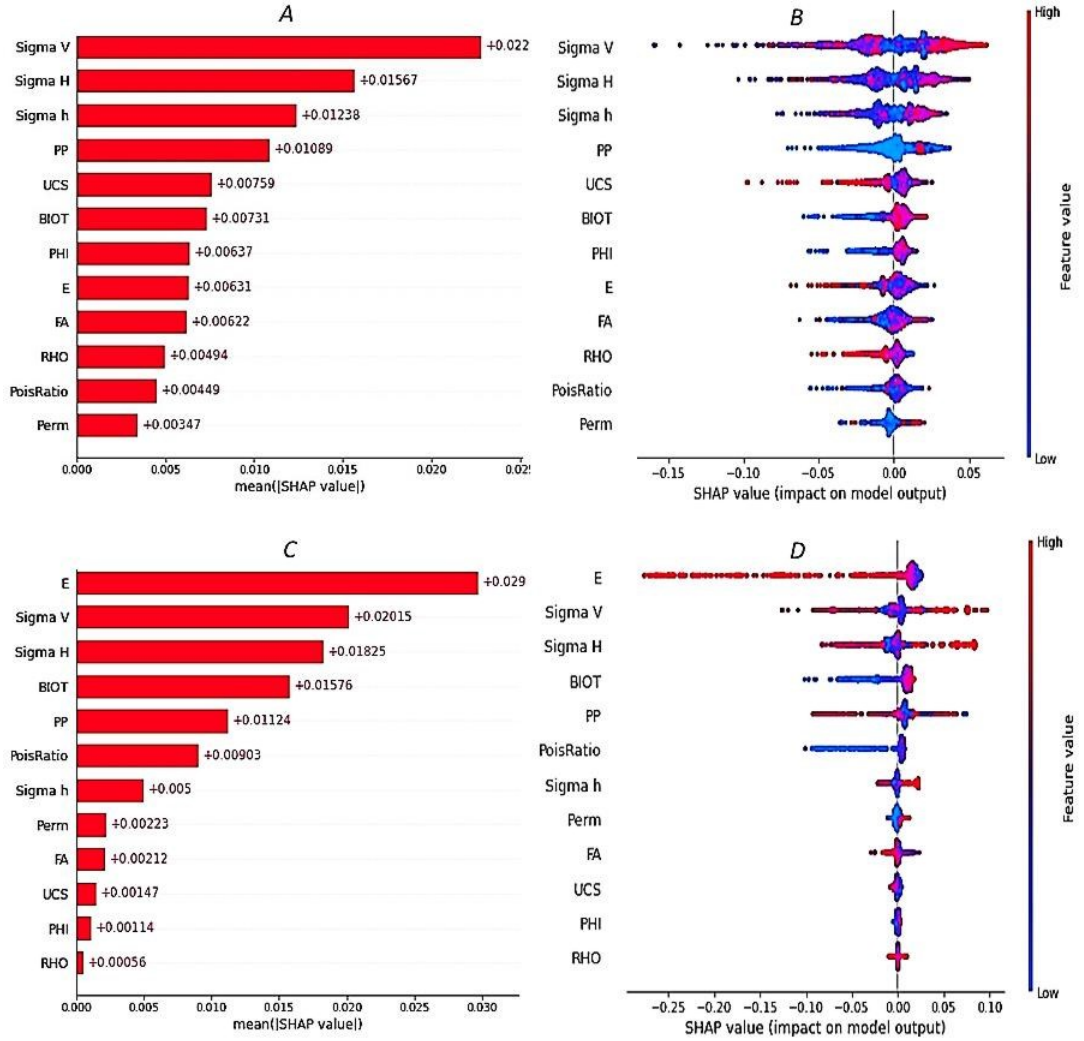


Figure 4 SHAP algorithm values for parameters affecting the instability of the well wall: Normal fault (A) and (B); Reverse fault (C) and (D).

Porous elasticity relationships and the Biot coefficient (coupling of pore pressure-horizontal stress interaction and effective stress in rock failure)

Pore pressure-horizontal stress interaction

The Biot coefficient is usually assumed to be 1, but it can vary in 3 cases. Firstly, changes in pore pressure. Secondly, changes in the Biot coefficient due to changes in pore pressure (this occurs when changes in pore pressure cause changes in the volume of rock cavities and their porosity). And thirdly, in cases where changes in pore pressure itself can cause changes in the stress state of the well drilling range, known as pore pressure-stress coupling or PSC. In this case, the interaction between pore pressure and stress (minimum horizontal stress) is defined according to the following relationship as $\Delta\sigma_h / \Delta P$ and α is the Biot coefficient when ϑ represents the Poisson's ratio [14]:

$$\frac{\Delta\sigma_h}{\Delta P} = \alpha \frac{1-2\vartheta}{1-\vartheta} \quad (6)$$

Therefore, pore pressures not only affects effective stress, but also total stress. In a stress regime such as the normal fault stress regime, σ_v or the vertical stress is not affected by the pore pressure-effective stress interaction. Thus, increasing P decreases the value of σ_{veff}. However, taking into account the interaction between P and σ_h, increasing pore pressure leads to an increase in σ_h. Therefore, the value of σ_{heff} does not decrease with an increase in P, but rather changes due to the value of PSC, or the pore pressure-effective stress interaction. In a stress regime such as the reverse fault stress regime, σ_v is also not affected by the pore pressure-effective stress interaction, but due to the interaction between P and σ_H,

the value of σ_H decreases with an increase in *P*, although this decrease is less than the value of σ_{veff} [14]. These cases are shown in **Figures 5(B) - 5(D)** for 2 different stress regimes and in the cases of reservoir depletion and injection. Additionally, using real well data, the values of PSC were compared for the assumed case of $\alpha = 1$ and the case where the value of the Biot coefficient has its real value (**Figures 5(A) - 5(C)**). Considering the mentioned results and what is observable in **Figure 5**, it can be concluded that assuming the Biot coefficient equal to 1 for estimating in-situ stress changes during the use of a reservoir, is an incorrect assumption.

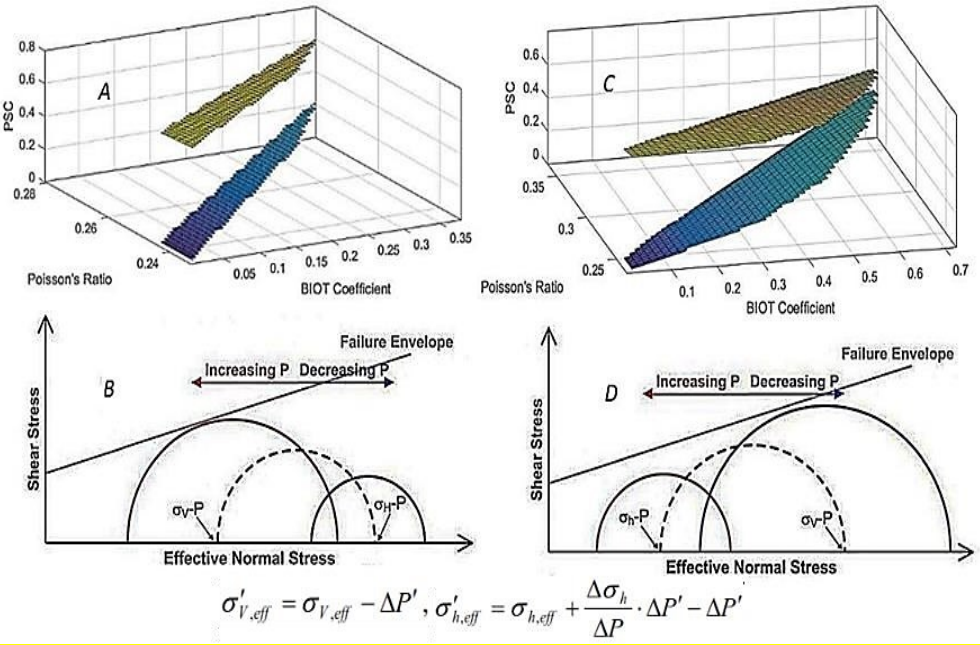


Figure 5 The interaction between effective stress and pore pressure during reservoir depletion and injection (reverse fault (A) and (B); normal fault (C) and (D)) along with Calculation of PSC values (Brown color assuming $\alpha = 1$ and blue color using the actual value of the Biot coefficient obtained from well data in (A) and (C)).

The effect of effective stress on shear failure of fluid-containing rocks (rock-fluid interaction)

The principle of effective stress, introduced by Terzaghi in the 1920s, plays a role in fundamental issues related to the deformation and failure of rocks, soil and other porous materials containing fluids and subjected to pore pressure, such as concrete. This principle allows us to predict the mechanical behavior of saturated porous media in the presence of pore pressure. Terzaghi stated that only a fraction of the total stress or (σ), which he called effective stress (σ'), is responsible for deformation and failure in soils and rocks, and is greater than the pore water pressure or (P), based on his empirical data. The proposed formula by Terzaghi for effective stress is given as a 1-dimensional equation in Eq. (7) [35]:

$$\sigma' = \sigma - P \tag{7}$$

In general, the stresses applied to a saturated porous medium are distributed between the solid rock structure and the pore fluid. The 1st part of these stresses is responsible for the deformation of the solid structure and is therefore called effective stress. In rock mechanics, tensile stresses are considered negative and compressive stresses are positive. Considering the existence of positive pore pressure in the porous rock medium, the equation for effective stress, taking into account the coefficient of Biot, is presented as Eq. (8):

$$\sigma_{ij} = \sigma'_{ij} - \alpha P \delta_{ij} \tag{8}$$

In Eq. (8), σ_{ij} and σ'_{ij} are the components of total stress and effective stress, respectively, and P is the pore pressure. The symbol δ_{ij} refers to the Kronecker's delta, which is defined as $\delta_{ij} = 1$ for elements with $i = j$ and $\delta_{ij} = 0$ for elements with $i \neq j$. The parameter α is also known as the Biot

coefficient, which as mentioned in the studies of Terzaghi, was considered to be 1, which is usually valid for soils. However, as will be mentioned later, it is not accurate for rocks with low porosity. Biot defined the α coefficient as Eq. (9) in 1941:

$$\alpha = \frac{K}{H} \tag{9}$$

where in K represents the drained bulk modulus for porous materials like rocks, while $1/H$ is the coefficient of poroelastic expansion which was introduced by him. This coefficient describes the changes in the volume of the bulk modulus due to changes in pore pressure under constant stress conditions. For soft soils, the value of α is considered to be 1. Soils are typically soft and have a highly compressible structure, while their solid particles have less compressibility. However, these properties are not always true for all types of rocks. As a result, the value of this coefficient will be completely different for different types of rocks [36].

As mentioned, the Biot coefficient is often assumed to be 1, which reduces Eq. (8) to (7) and, therefore, according to Eq. (8), effective stress can vary in 3 different states. In the theory of poroelasticity, mechanical processes of the solid part of the body and hydraulic processes of the fluid part present in the pores of the body are interrelated and interact simultaneously. This enables simultaneous modeling of these processes in a porous environment. Employing this type of modeling is generally essential in all processes where fluid flow through a porous medium is accompanied by deformation. Pore pressure, as a key variable, causes fluid flow and solid part movements and simultaneously affects various fluid parameters with changes in the solid part of the body. High fluid flow velocity inside rock pores may lead to rock deformation, such as in proximity to injection wells or oil depletion wells. This model takes into account fluid flow according to Darcy's theory and linear elastic behavior for the solid part of the body to simulate reservoir mechanical behavior better [37]. A poroelastic model is defined in the Comsol Multiphysics software using Eq. (10):

$$\begin{aligned} &-\nabla \cdot \sigma = F_V . \sigma = s \\ &s - s_0 = C : (\varepsilon - \varepsilon_0 - \varepsilon_{inel}) - \alpha p I \\ &\varepsilon = \frac{1}{2}((\nabla u)^T + \nabla u) \\ &\rho S \frac{\partial p}{\partial t} + \nabla \cdot (\rho q) = Q - \rho \alpha \frac{\partial \varepsilon_V}{\partial t} \\ &q = -\frac{k}{\mu} \nabla p \end{aligned} \tag{10}$$

where σ and s are induced stress tensors, s_0 is the in-situ stress field, I is the unit matrix, F_V is the vector of volumetric forces, ε is the strain tensor, u is the deformation change vector, P is the pore pressure, C is the elastic (stiffness) matrix, and α is the Biot coefficient. The subscript *0* indicates initial values and *inel* indicates the plastic strain. As observed in the relevant equation regarding fluid flow, changes are time-dependent. In this model, mechanical and hydraulic processes are combined. The hydraulic model is described by Eq. (11):

$$\begin{aligned} &\rho S \frac{\partial p}{\partial t} + \nabla \cdot (\rho q) = Q - \rho \alpha \frac{\partial \varepsilon_V}{\partial t} \\ &q = -\frac{k}{\mu} \nabla p \end{aligned} \tag{11}$$

where q represents the fluid velocity, k represents the permeability of the rock, μ represents the dynamic viscosity of the fluid, S represents the storage parameter, Q represents the fluid flow rate, and ε_V represents the volumetric strain. As the equations above indicate, the Biot coefficient appears as a weighting factor in the stress-strain relationship to estimate the effect of fluid pressure on induced stress. Additionally, this coefficient is mentioned as one of the time-dependent hydraulic model parameters in Eq. (11). In the stationary state, the interaction between the rock and fluid is 1-sided because the time derivatives in the flow equation are eliminated, but in the time-dependent state, the mechanical and hydraulic interactions are mutual. This means that the effect of fluid pressure on the stress-strain relationship and the resulting deformation of the solid body are significant, and conversely, the flow equation is affected by the volumetric strain of the solid body [37].

The Mohr-Coulomb 3-dimensional criterion for investigating shear failure in the wellbore wall using the concept of effective stress by the Biot theory

There are many theoretical and empirical criteria for analyzing rock failure and soil behavior in 3-dimensional stress states. However, the most important 3-dimensional criteria for shear analysis in 3 dimensions are the Mohr-Coulomb, the Lade criterion and the Drucker-Prager criterion. These criteria present different relationships between rock failure and the 3 principal stress values.

It should be noted that the 2-dimensional Mohr-Coulomb failure criterion is independent of the intermediate principal stress, or σ_2, and depends only on the minimum and maximum principal stresses, or σ_1 and σ_3, as well as the magnitude of pore pressure and the value of the Biot coefficient. In the Drucker-Prager criterion, we observe a significant increase in rock strength with an increase in intermediate principal stress, or σ_2, and the Lade criterion also predicts a gradual increase in rock strength with an increase in intermediate principal stress. Therefore, removing the intermediate stress value, or σ_2, in the 2-dimensional Mohr-Coulomb criterion will result in errors and incorrect estimates of rock strength in a 3-axis stress field. It is worth mentioning that the 3-dimensional Mohr-Coulomb criterion predicts a linear relationship between the intermediate stress and rock strength and can be considered as a linearized form of the Lade criterion, which is also very simple to use in numerical modeling [38,39]. Therefore, in this study, we used the Mohr-Coulomb 3-dimensional criterion as a failure function, similar to Eq. (12), to evaluate shear failure in the wall of a horizontal oil well under in-situ stresses and pore pressure. In Eqs. (13) and (14), *So* is equal to the Coulomb cohesion and φ is the Coulomb friction angle. It is necessary to mention that the values of σ_2, σ_3 and σ_1 in Eq. (12) are calculated in a poroelastic medium under pore pressure using the 2nd equation in Eq. (10) in the Comsol Multiphysics software. As a result, the effect of pore pressure is taken into account in them by the software itself, considering the value of the Biot coefficient. Additionally, in Eq. (12), the condition "fail = 0" indicates the moment of rock failure initiation, while the condition "fail < 0" represents sudden rock failure. Therefore, positive values of the failure function indicate stability and reduced risk of failure, while negative values indicate a higher risk of failure [40].

$$fail = \left[\sigma_3 - Q\sigma_1 + N\left(1 + \frac{(\sigma_2 - \sigma_1)}{(\sigma_3 - \sigma_1)}\right)\right] < 0 \quad (12)$$

$$Q = \left[\frac{1 + sin(\varphi)}{1 - sin(\varphi)}\right] \quad (13)$$

$$N = \left[\frac{2Socos(\varphi)}{1 - sin(\varphi)}\right] \quad (14)$$

As mentioned earlier, in the time-dependent solution, the solid-fluid interaction is bidirectional, while in the solution independent of time, it is unidirectional. Since the assumptions related to this paper are specific to the stability analysis of the drilling process and do not include time-dependent processes such as production from the well and injection into the well, stationary and time-independent solutions have been used for the related analysis. In this study, we will focus on the relationships between hydraulic and mechanical environments. In this research, a 3-dimensional hydro-mechanical finite element model with elastic deformation (in the elastic range) and Darcy flow is used. The mentioned model calculates the distribution of stress, strain and pore pressure (resulting from well pressure (Pw), reservoir pressure (P0), and in-situ stresses (σ_v, σ_h and σ_H)) throughout a cubic model domain with 1 m dimensions. The modeling results provide the distribution of strain and stress (including pore pressure) around a 1 m section of the well and are used to evaluate areas beyond the rock failure conditions according to Eq. (12). For fluid flow, a series of constant boundary conditions are considered, and free flow conditions are maintained on all external boundaries. It is also assumed that the distance between the boundaries is sufficiently far from the pumping regime, and the boundaries remain unchanged. For flow boundaries, the change in fluid pressure from the well to the reservoir edge is specified, and the well is the only outlet for fluid. Regarding the mechanical model, free deformation is possible inside the network, but all side boundaries are restricted according to the displacement conditions of Eq. (15) using the Roller operator. Here, we present the results for the static case, as these results represent the maximum final response of the earth system to changes in the stress regime. The modeled geometry consists of a 1 m cube with a horizontally drilled well of 6 inches in diameter, which is part of a larger network. The fluid only exits through the well and displacements are limited along the reservoir. However, the walls of the well deform freely. In Eq. (15), *n* is the normal vector in the direction of the boundary.

$$\begin{array}{ll} p = p_0 & \partial\Omega \ Reservior \\ p = p_w & \partial\Omega \ Well \\ n.u = 0 & \partial\Omega \ Reservior \\ Free & \partial\Omega \ Well \end{array} \quad (15)$$

Shear failure analysis results in wellbore wall using the concept of Biot effective stress

In this study, the processing was carried out in 2 stages. In the 1st stage, the parameters of **Table 1** were applied along with the boundary conditions of Eq. (15) for a selected depth of the well and the stability of the well was investigated. In the 2nd stage, using the parametric sweep feature of the software, the value of the Biot coefficient among the parameters in **Table 1** for the selected depth was varied in the range of 0.1 to 1 with a step size of + 0.1. It should be noted that in the 2nd processing stage, only changes in the Biot coefficient were considered, and other parameters in **Table 1** remained constant to investigate the stability of the well. Since the stable and time-independent state has been chosen for the analysis, the parameters related to the fluid are independent of solid stress-strain changes and, therefore, are not affected by changes in the Biot coefficient. In **Figure 6**, the total displacement is calculated at different Biot coefficients based on the distance from the wellbore wall. As it can be seen from this figure, changes in the Biot coefficient have a significant effect on the displacement and the total displacement increases with its increase. In **Figure 7**, using the software's volume integral operator, the volume of the crushed rock around the wellbore was calculated according to the failure function Eq. (12) (fail function) at different Biot coefficients. As it is clear, assuming a value of 1 for the Biot coefficient results in a much higher amount of sand production than the actual state. In the case where, according to the real values, the well is stable and there is no sand production. The result that the comparison of the values of the caliper and bit Size charts also makes clear for us in this part of the well. **Figure 8** also includes the effective parameters affecting the fail function in a section of the wellbore length under normal fault stress regime based on real data, along with the value of the ERROR parameter for a more accurate investigation. This serves as confirmation of the results presented in previous sections.

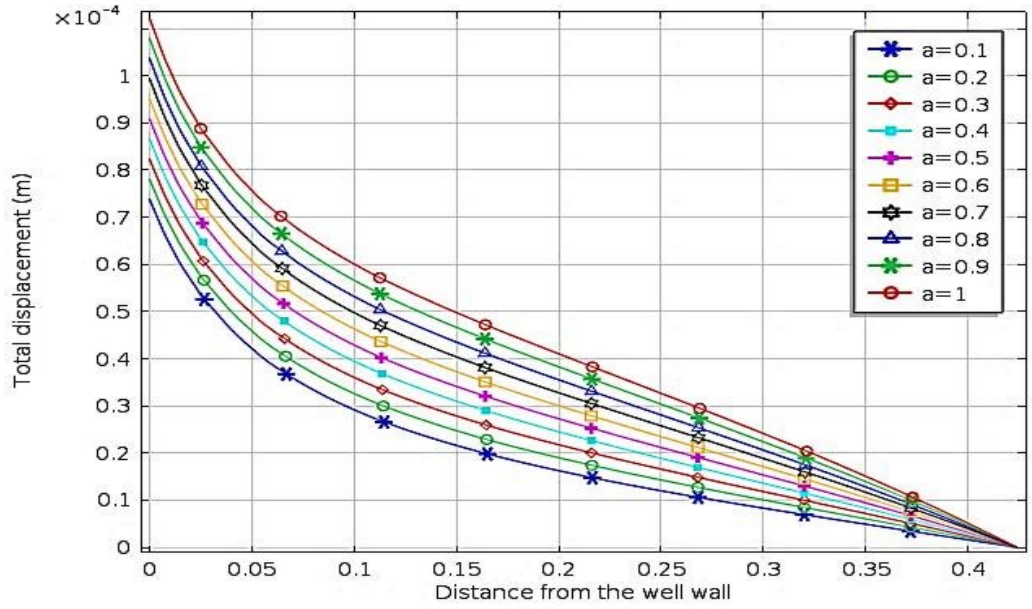


Figure 6 The total displacement as a function of distance from the wellbore wall at different values of the Biot coefficient.

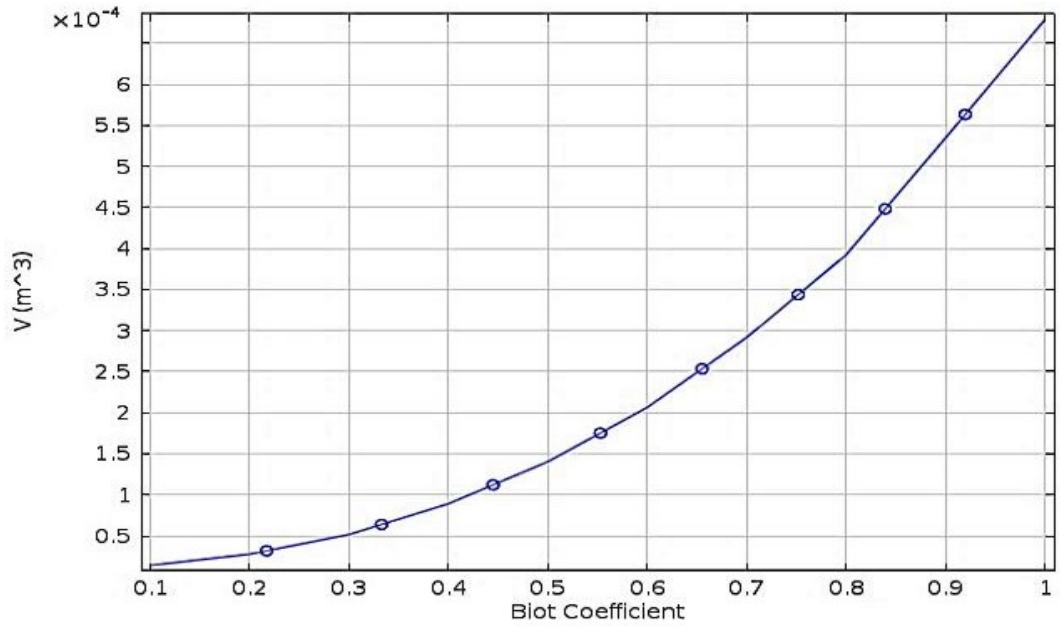


Figure 7 The volume of crushed rock around the wellbore at various values of the Biot coefficient, according to the fail function.

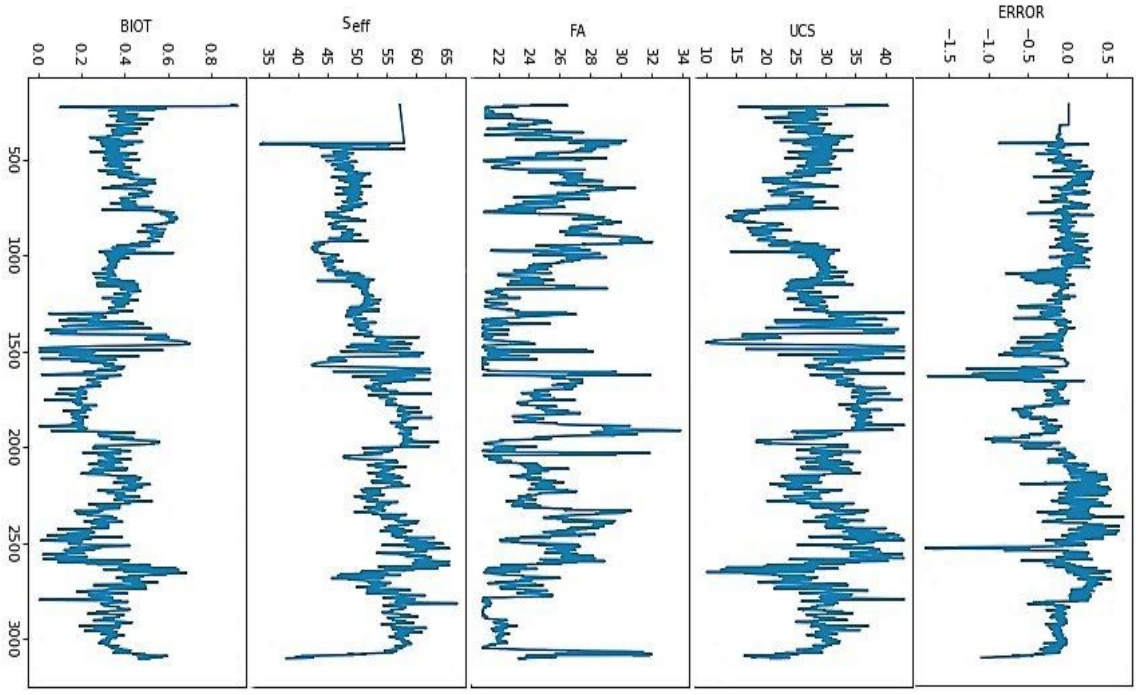


Figure 8 Effective parameters on the fail function in a section of the wellbore length based on real data alongside the ERROR parameter value.

Conclusions

To achieve the best results in the 1st section of the article, the XGBoost algorithm, which is a gradient boosting regression method, was used. In addition, 7 categories of hyper parameters in this algorithm were optimized using the Bayesian optimization method. Finally, to interpret the results, an interpretable artificial intelligence model, or the SHAP algorithm, was used. Considering the 4 main criteria of MSE, RMSE, MAE and R2 for evaluating the quality of regression modeling by the XGBoost

algorithm, it should be noted that the values of these criteria, after optimization by the Bayesian method, have changed from values such as 0.01030, 0.1015, 0.0701 and 0.1195 to values such as 0.00209, 0.0458, 0.0285 and 0.821 (in the stress regime similar to normal faults of the data), respectively. Also, the SHAP algorithm's resulting graphs, with the optimized output visualization of the XGBoost algorithm, have reliable results in determining the effective parameters on the stability of the wellbore wall in 2 stress regimes of normal and reverse faults. Based on these results, in the normal fault stress regime, parameters such as vertical stress, maximum horizontal stress, minimum horizontal stress, pore pressure, uniaxial compressive strength, Biot coefficient, porosity, elastic modulus, internal friction coefficient of rock, rock density, Poisson's ratio and fluid permeability have the most significant impact on instability.

In the reverse stress regime, parameters such as elastic modulus, normal stress, maximum horizontal stress, Biot coefficient, pore pressure, Poisson's ratio, minimum horizontal stress, permeability; internal friction coefficient, uniaxial compressive strength, porosity and ultimately rock density have the greatest impact on the instability of wellbore walls. From these results, it can be inferred that in the initial section of the wellbore, the ERROR parameter (difference between caliper log and bit size) is more related to in-situ stresses, pore pressure, rock strength, and Biot coefficient, and wall failures are of the type of fracture and sand production. However, at greater depths and under the reverse stress regime, the ERROR parameter is more affected by plastic deformation of the wellbore wall (although fracture-induced failure is also observable in this area). It should be noted that simultaneous examination of Pearson correlation coefficient and SHAP algorithm plots is also highly useful for a better understanding of the geomechanics of this reservoir. It is also essential to mention that the results obtained are only applicable to the studied well in southwest Iran and in the Asmari formation.

Furthermore, since the changes in pore pressure and horizontal stress in a porous medium such as oil reservoirs are interrelated and affect each other, the difference between assuming a value of 1 for the Biot coefficient compared to the actual value was investigated in order to calculate the PSC coefficient (pore pressure-stress interaction) and it was determined that using the poroelasticity equations can provide a more accurate solution for analyzing the geomechanics of the reservoir. Next, the effect of effective stress on the shear failure of fluid-containing rocks (rock-fluid pore pressure interaction) was discussed, and as was seen in the sensitivity analysis of the model to the Biot coefficient, assuming a value of 1 for the Biot coefficient, which results in the effective stress equation being similar to the Terzaghi effective stress equation, is an incorrect assumption for analyzing the stability of an oil well for shear fractures according to the Mohr-Coulomb 3D criterion. Assuming a value of 1 for the Biot coefficient compared to the actual value increased the volume of crushed rock resulting from shear failure (using the Mohr-Coulomb 3D criterion) by up to 13 times, and also increased the overall displacement in the wellbore wall by up to 36 %. Therefore, any accurate analysis regarding the stability of this well's wall using the concept of shear failure requires precise knowledge of the actual and non-ideal value of this coefficient.

Acknowledgements

The authors would like to acknowledge the Geology and Petrophysics Department of Iranian Offshore Oil Company (IOOC), Iran, for their cooperation in preparation and supply the specific data for the well at Mansouri oil field.

References

[1] D Moos and M Zoback. Utilization of observations of well bore failure to constrain the orientation and magnitude of crustal stresses: Application to continental, Deep Sea Drilling Project, and Ocean Drilling Program boreholes. *J. Geophys. Res. Solid Earth* 1990; **95**, 9305-25.

[2] R Sanaee, SR Shadizadeh and MA Riahi. Determination of the stress profile in a deep borehole in a naturally fractured reservoir. *Int. J. Rock Mech. Min. Sci.* 2010; **47**, 599-605.

[3] BC Haimson and FH Cornet. ISRM suggested methods for rock stress estimation - part 3: Hydraulic fracturing (HF) and/or hydraulic testing of pre-existing fractures (HTPF). *Int. J. Rock Mech. Min. Sci.* 2003; **40**, 1011-20.

[4] E Fjaer, RM Holt, P Hordrud, AM Raaen and R Risnes. *Developments in petroleum science.* Elsevier Press, Hungary, 2008, p. 295-305.

[5] M Zoback, CA Barton, M Brudy, DA Castillo, T Finkbeiner, BR Grollimund, DB Moos, P Peska, CD Ward and DJ Wiprut. Determination of stress orientation and magnitude in deep wells. *Int. J. Rock Mech. Min. Sci.* 2003; **40**, 1049-76.

[6] M Zoback. *Reservoir geomechanics*. Cambridge university press, Cambridge, 2010, p. 206-34.

[7] M Ezati, M Azizzadeh, MA Riahi, V Fattahpour and J Honarmand. Wellbore stability analysis using integrated geomechanical modeling: A case study from the Sarvak reservoir in one of the SW Iranian oil fields. *Arabian J. Geosci.* 2020; **13**, 149.
[8] HS Almalikee and FM Al-Najm. Wellbore stability analysis and application to optimize high-angle wells design in Rumaila oil field, Iraq. *Model. Earth Syst. Environ.* 2019; **5**, 1059-69.
[9] D Raljević, JP Vuković, V Smrečki, LM Pajc, P Novak, T Hrenar, T Jednačak, L Konjević, B Pinević and T Gašparac. Machine learning approach for predicting crude oil stability based on NMR spectroscopy. *Fuel* 2021; **305**, 121561.
[10] L Qiang. Prediction of ultimate bearing capacity of oil and gas wellbore based on multi-modal data analysis in the context of machine learning. *Int. J. Inform. Tech. Syst. Approach* 2023; **16**, 1-13.
[11] CM Sayers, C Sayers, LD Boer, D Lee, P Hooyman and R Lawrence. Predicting reservoir compaction and casing deformation in deepwater turbidites using a 3D mechanical earth model. *In*: Proceedings of the International Oil Conference and Exhibition in Mexico, Cancun, Mexico. 2006.
[12] X Chen, CP Tan and CM Haberfield. Guidelines for efficient wellbore stability analysis. *Int. J. Rock Mech. Min. Sci.* 1997; **34**, 50.e1-50.e14.
[13] L Qiuguo, X Zhang, K Al-Ghammari, L Mohsin, F Jiroudi and AA Rawahi. 3-D Geomechanical modeling and wellbore stability analysis in Abu Butabul Field. *In*: Proceedings of the Abu Dhabi International Petroleum Conference and Exhibition, Abu Dhabi, United Arab Emirates. 2012, p. 312-9.
[14] JB Altmann, BIR Müller, TM Müller, O Heidbach, MRP Tingay and A Weißhardt. Pore pressure stress coupling in 3D and consequences for reservoir stress states and fault reactivation. *Geothermics* 2014; **52**, 195-205.
[15] K Lipnikov. 2002, Numerical methods for the Biot model in poroelasticity. Ph. D. Dissertation University of Houston, Texas.
[16] G Gambolati and RA Freeze. Mathematical simulation of the subsidence of Venice: 1. Theory. *Water Resour. Res.* 1973; **9**, 721-33.
[17] RW Lewis and B Schrefler. A fully coupled consolidation model of the subsidence of Venice. *Water Resour. Res.* 1978; **14**, 223-30.
[18] A Lin, M Alali, S Almasmoom and R Samuel. Wellbore instability prediction using adaptive analytics and empirical mode decomposition. *In*: Proceedings of the IADC/SPE Drilling Conference and Exhibition, Fort Worth, Texas. 2018.
[19] M Sabah, M Mehrad, SB Ashrafi, DA Wood and S Fathi. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J. Petrol. Sci. Eng.* 2021; **198**, 108125.
[20] H AlBahrani and N Morita. Risk-controlled wellbore stability criterion based on a machine-learning-assisted finite-element model. *SPE Drill. Completion* 2022; **37**, 38-66.
[21] D Makowski, MS Ben-Shachar, I Patil and D Lüdecke. Methods and algorithms for correlation analysis in R. *J. Open Source Software* 2020; **5**, 2306.
[22] T Chen and C Guestrin. XGBoost: A scalable tree boosting system. *In*: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, California. 2016, p. 785-94.
[23] R Dazeley, P Vamplew, C Foale, C Young, S Aryal and F Cruz. Levels of explainable artificial intelligence for human-aligned conversational explanations. *Artif. Intell.* 2021; **299**, 103525.
[24] SC Chelgani. Estimation of gross calorific value based on coal analysis using an explainable artificial intelligence. *Mach. Learn. Appl.* 2021; **6**, 100116.
[25] S Mangalathu, SH Hwang and JS Jeon. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* 2020; **219**, 110927.
[26] VH Truong, G Papazafeiropoulos, QV Vu, VT Pham and Z Kong. Predicting the patch load resistance of stiffened plate girders using machine learning algorithms. *Ocean Eng.* 2021; **240**, 109886.
[27] XGoost.readthedocs.io, Available at: https://xgboost.readthedocs.org/, accessed January 2023.
[28] J Guo, L Yang, R Bie, J Yu, Y Gao, Y Shen and A Kos. An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Comput. Network.* 2019; **151**, 166-80.
[29] W Zhang, C Wu, H Zhong, Y Li and L Wang. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci. Front.* 2021; **12**, 469-77.

[30] Y Xia, C Liu, Y Li and N Liu. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst. Appl.* 2017; **78**, 225-41.
[31] Bayesian optimization, Available at: https://github.com/fmfn/BayesianOptimization, accessed January 2023.
[32] M Steurer, RJ Hill and N Pfeifer. Metrics for evaluating the performance of machine learning based automated valuation models. *J. Property Res.* 2021; **38**, 99-129.
[33] SC Chelgani, H Nasiri and M Alidokht. Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A "conscious-lab" development. *Int. J. Min. Sci. Tech.* 2021; **31**, 1135-44.
[34] SC Chelgani, H Nasiri and A Tohry. Modeling of particle sizes for industrial HPGR products by a unique explainable AI tool-A "Conscious Lab" development. *Adv. Powder Tech.* 2021; **32**, 4141-8.
[35] V Guerriero and S Mazzoli. Theory of effective stress in soil and rock and implications for fracturing processes: A review. *Geosciences* 2021; **11**, 119.
[36] A Merxhani. An introduction to linear poroelasticity. *Geophysics* 2016, https://doi.org/10.48550/arXiv.1607.04274
[37] E Holzbecher. Poroelasticity benchmarking for FEM on analytical solutions. *In*: Proceedings of the COMSOL Conference, Rotterdam, Netherlands. 2013.
[38] JF Labuz and A Zang. Mohr-Coulomb failure criterion. *Rock Mech. Rock Eng.* 2012; **45**, 975-9.
[39] JB Altmann. 2010, Poroelastic effects in reservoir modelling. Ph. D. Dissertation. Karlsruher Institut für Technologie, Karlsruhe, Germany.
[40] R Suarez-Rivera, JW Martin and BJ Begnaud. Numerical analysis of open-hole multilateral completions minimizes the risk of costly junction failures. *In*: Proceedings of the Rio Oil and Gas Expo and Conference, Rio de Janeiro, Brazil. 2004.