RGNZ/MGPB

Predicting the weight and type of drilling mud by machine learning method

Rudarsko-geološko-naftni zbornik
(The Mining-Geology-Petroleum Engineering Bulletin)
DOI: 10.17794/rgn.2026.3.1

Original scientific paper

Amin Tohidi[1]*, Alireza Afradi[2]
[1] *Department of Mining Engineeing, Amirkabir University of Technology, Tehran, Iran.*
[2] *Department of Mining and Geology, QaS.C., Islamic Azad University, Qaemshahr, Iran.*

Abstract
Selecting the optimal drilling fluid, defined by its weight and chemical type, is critical for preventing costly wellbore instability and catastrophic accidents. Traditional methods often rely on trial-and-error, past experience or simplified models that fail to capture the complex rock-fluid interactions. While data mining offers a promising alternative, a research gap exists in simultaneously predicting both mud weight and type. This study introduces a novel machine learning framework that concurrently predicts these essential properties. Utilizing a comprehensive dataset extracted from 50 years of daily drilling reports across 20 oil wells, we trained and compared three nature-inspired algorithms: Ant Colony (ACO), Bee Colony (BCO), and Emperor Penguins Colony (EPC) optimization. The results demonstrate that all models achieved high predictive accuracy, with the Bee Colony Optimization (BCO) algorithm emerging as the most precise, yielding a correlation coefficient (R^2) of 0.9841 and a root-mean-square error (RMSE) of 0.0245. Furthermore, sensitivity analysis revealed that the Rate of Penetration (ROP) is the most influential parameter on mud properties, surpassing other drilling variables. A key practical finding was the consistent model consensus, with 79-87% confidence, that sea water-based mud with polymer and soltex additives (SW-PO-SX) is the optimal fluid for the studied field. This research provides a robust, data-driven solution that enables a systematic and proactive approach to drilling fluid selection, significantly enhancing operational safety and efficiency.

Keywords:
data mining, predictive model, mud weight, mud type, wellbore stability

1. Introduction

The problems of inappropriate drilling fluid are stuck pipe, wellbore instability, slow drilling rates, excess torque and drag, lost circulation, high temperature additive degradation, and corrosion (**Caenn et al., 2011**). The selection of the drilling fluid depends on numerous factor such as economic, and technical issues. Oil and synthetic based mud are expensive relative to water-based mud so used in challenging geological condition. Some of complex geotechnical conditions are salt and active shale rock. The excellent choice for active shale formation is oil-based mud (**Technology Roundtable: Drilling Fluids, 2017**).

Sometimes, the wrong choice of drilling fluid causes a catastrophic accident. For example, in the Rag-Sefid accident, drilling mud was lost into the natural fracture of the formation, and the reduction in the weight of mud caused the gas to be imported into the well and a blow-out took place (**Reuters, 2013**).

* Corresponding author: Amin Tohidi
e-mail address: a.tohidi@aut.ac.ir
Received: 27 April 2025. Accepted: 22 October 2025.
Available online: 14 May 2026

Also, to reach the oil and gas formation, drilling passes through various rocks, in this path some formations have natural fracture and bedding planes and some formations have creep, swelling and dissoluble potential so the stability of a wellbore is critical and makes drilling a challenging phase in the oil and gas industry. Wellbore instability, lost circulation, kick and blowout are examples of what may occur during drilling. Lost circulation is the migration of the drilling fluid into the pores and natural fractures of a rock and is estimated to cost over 2 billion USD annually. Drilling in active shale and salt formations can cause chemical instability, so accurate prediction of the type of mud is the most critical solution to dealing with these difficulties (**Ebadati & Najari, 2016; Gholami et al., 2021; Eppelle and Gerogiorgis., 2019; Nunez et al., 2021; Pao et al., 2016; Sabah et al., 2021**). Wellbore instability can be either of the mechanical or chemical type and can cause financial losses of 1–6 billion USD annually (**Albukhari et al., 2018; Tohidi et al. 2017**).

Mechanical wellbore instability occurs when the mud pressure act on the wellbore wall is too low or too high and causes shear or tensile failure of the rock, respectively (**Freij-Ayoub et al., 2003**). The mud weight can

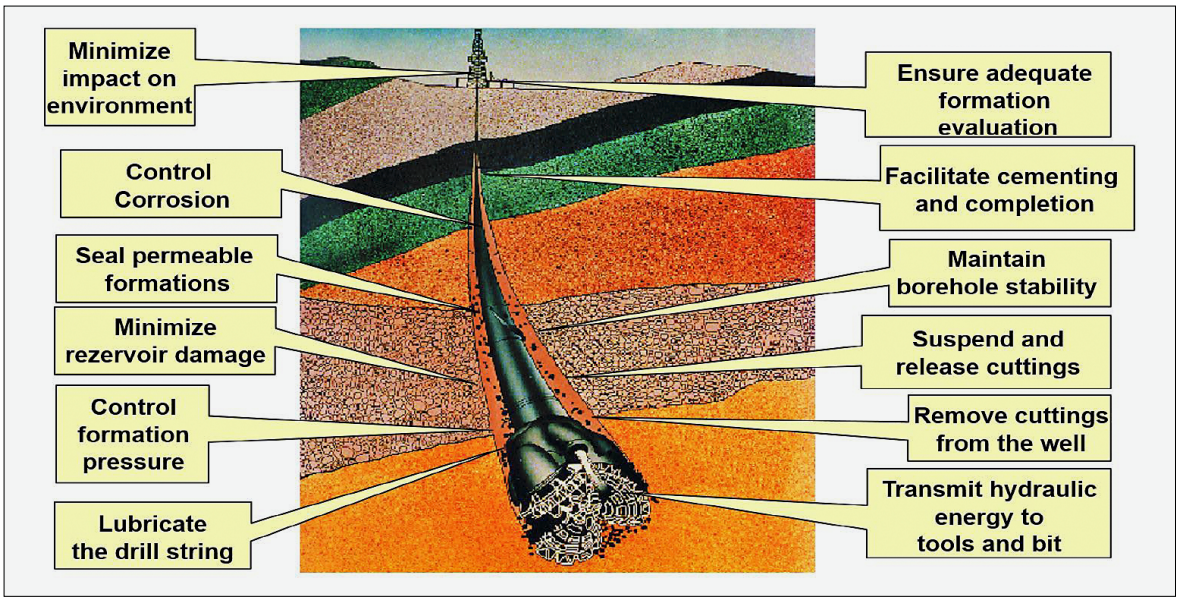


Figure 1. Multi-function of drilling fluid (**Mozie, n.d.**)

be lightened by inserting foam or gas into the mud and the mud can be weighted with an agent such as barite (**Aljubran et al., 2022**). Chemical wellbore instability results from an interaction between the drilling mud and formations such as active shale and salt. When the reason for wellbore stability is mechanical, the mud density is altered. When the reason for the problem is chemical, the mixture of drilling fluid such as clay, polymer, and surfactant changed to achieve the right solutions (**Allawi, 2023**). So, accurate prediction of drilling fluid weight and type is critical and requires simultaneous consideration during drilling operations (**Olson et al., 2005**).

As shown in **Figure 1**, drilling fluid or drilling mud has several advantages (**Aljubran et al., 2022; Deka, 2023**). The mechanical and chemical properties of the drilling fluid must be designed for a specific scenario (**DRILLING FLUID, 2023**).

Figure 2 shows that, for determining the maximum mud density of the fluid, the mud pressure must be below the tensile strength of the rock mass to prevent lost circulation. Two criteria control the minimum density of the mud weight. It must be greater than pore pressure to prevent kick or blowout from occurring and also should be greater than the shear strength of the rock so that wellbore instability does not occur (**Soroush, 2020**). Although the logic for estimating drilling fluid weight is simple, predicting the proper mud weight cannot be straightforward, because the number, dip, and strike of natural fracture control the strength behaviour of rock masses that cannot be estimated with deterministic approaches.

The chemical composition of the drilling fluid can be determined from the formation and interaction requirements. There are three main types of drilling fluid: oil, water and synthetically based fluid (**Aljubran et al., 2022**). The type of drilling fluid is based on the composition and concentration of the fluid components. In the current study, an algorithm is proposed that predicts both the weight and type of the mud in a specific oil field so that the most appropriate fluid can be selected. For clarification, at first, a schematic diagram of wellbore problems (that we extract data from it) during drilling was drawn.

Schematic diagram of problems that occurred during drilling in well 2 and 3 of oil field in south-west of Iran is shown in **Figure 3**.

According to **Figure 3**, when problems such as lost circulation, stuck pipe, and tight hole occur, the remediation actions involve adjusting the mud weight and/or its chemical composition. Mud weight is primarily adjusted to address mechanical wellbore instability (e.g. increasing weight to prevent shear failure or decreasing it to avoid tensile fracture and lost circulation). In contrast, the chemical composition is modified to combat chemical instability; for instance, switching to an inhibitive oil-based mud or adding salts like potassium chloride (KCl)

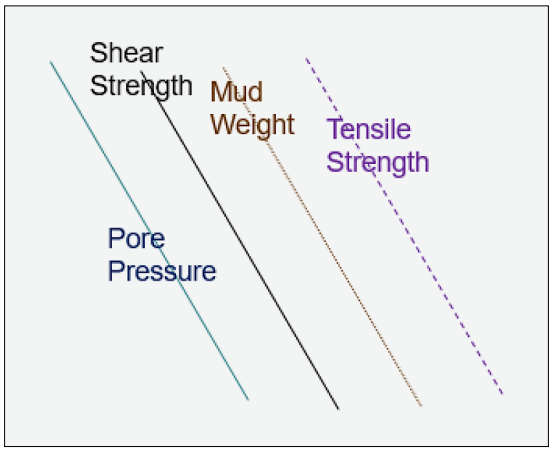


Figure 2. The logic for mud weight selection

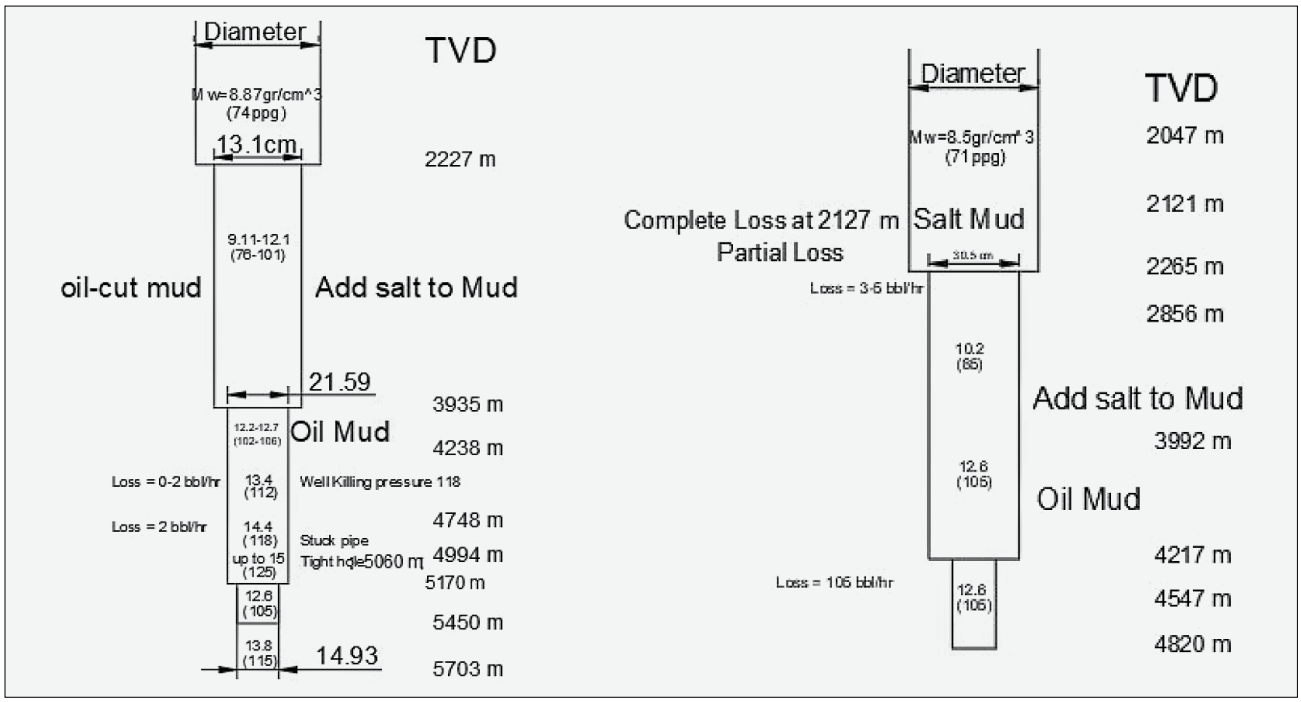


Figure 3. graphical representation of problems during drilling well 2 and well 3

to a water-based mud is a standard practice to prevent shale hydration and swelling that leads to stuck pipe and tight holes. If these remedial actions are unsuccessful in controlling the wellbore instability, a well control operation means killing the well may be required, followed by a workover to restore safe and efficient operations.

Unfortunately, these drilling problems occur for the next wells, in **Figure 4** the schematic drilling problems of well number 4 are shown as the same problems as well number 2 and 3. That means lost circulation and chemical problems of the well occurred during drilling. Thus, we need to find a systematic solution so that the mistake is not repeated. Data-mining techniques were used to learn from previous wells to minimize drilling problems. For this, the parameters such as depth, rate of penetration (ROP), drill string rotation (RPM), weight on a bit (WOB), and flow rate input to model and mud weight and type the output of prediction model.

All engineering efforts can be simplified into two terms, prediction and if the result of the prediction is not suitable the prevention applied, for prediction behaviour researchers have used different approaches. For example, to determine the appropriate drilling fluid properties with analytical methods, only the mechanical instabilities can predict when the stress induced around a wellbore is larger than the strength of the rock. Simplified assumptions used in this method like CHILE (Continues Homogeneous Isotropic Linear Elastic) rock are its main disadvantage (**Albukhari et al., 2018; Tohidi et al., 2017**). Numerical procedures for determining the mud weight depend on input parameters such as the in-situ stress, natural fracture orientation, friction angle, cohesion, etc. which cannot be determined precisely (**Beheshtian et al., 2022; Freij-Ayoub et al., 2003; Gowida et al., 2022; Zahiri et al., 2018**). Laboratory simulation is a miniature model of real problems usually conducted on synthetic specimens so it cannot describe the complexity existing in the field (**Agwu et al., 2018; Allawi, 2023**). Mud engineers generally select the weight and type of mud based on previous experience, but this knowledge can be lost if the engineer leaves the company (**Tatar et al, 2015; Ayoub et al., 2019**).

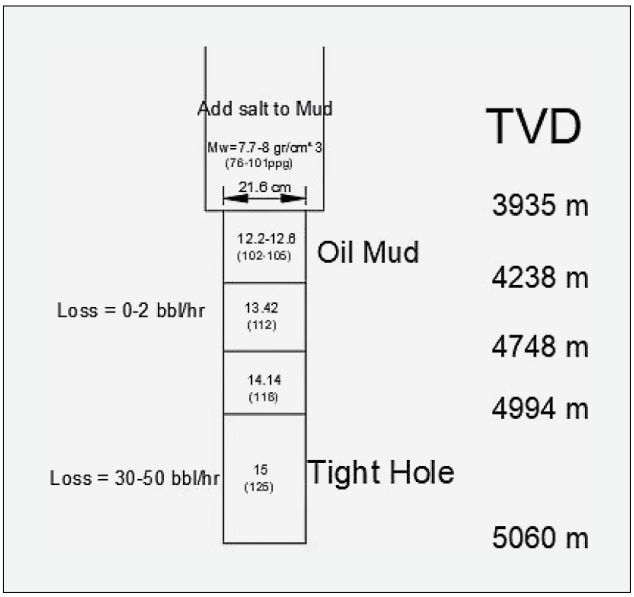


Figure 4. graphical representation of problems during drilling well 4

The complex nature of a reservoir, including a natural fracture, high pressure, high temperature, plastic behav-

Table 1. Descriptive statistics of database based on lithology

Lithology	TVD (m)	ROP (m/h)	Diameter (m)	RPM (1/min)	WOB (kN)	Flow Rate (kg/s)	Mud Density (kg/m³)
argillaceous limestone	2482.71	1.07	0.35	188.57	35.90	1237.38	1534.18
argillaceous limestone 80%	3594.00	1.40	0.31	244.00	35.59	1133.98	1774.23
calcareous claystone	2880.50	1.34	0.33	142.28	39.82	1168.71	1666.18
claystone	2204.00	1.98	0.40	166.66	46.71	1533.18	1630.18
dolomite	1766.02	1.51	0.48	104.00	49.82	1669.22	1450.10
limestone	3682.42	1.22	0.27	171.00	33.44	966.17	1774.23
marl	4002.50	0.84	0.26	210.00	33.36	1043.26	2025.07
shale	3986.00	0.49	0.22	60.00	20.02	907.18	2073.40
silty claystone	3626.00	0.49	0.31	244.00	44.48	1133.98	1774.23
Total	2892.22	1.34	0.33	143.15	39.73	1167.07	1676.44
min	17.00	0.03	0.15	10.00	0.00	222.26	1210.44
max	22121.00	10.56	0.61	650.00	122.33	1814.37	2253.33

iour and chemical composition, makes the precise prediction of the mud properties impossible by conventional methods. Hence, the use of data-mining techniques that are trained using the recorded well data is a promising way of predicting the mud properties. Data mining presents a comprehensive solution that is superior to the other methods. The current study used daily drilling reports to extract the mud properties and controllable and non-controllable parameters of 20 oil wells drilled over 50 years for machine learning to predict the mud properties of future wells.

2. Related Work

The efficacy of machine learning (ML) and metaheuristic optimization in solving complex geomechanical challenges is well-documented across various mining and rock engineering domains. For instance, sophisticated ML models have been successfully deployed to predict critical phenomena such as blast-induced ground vibration and flyrock intensity, demonstrating superior accuracy over traditional empirical methods (**Chen et al., 2025; Hasanipanah & Amineh, 2025**). Similarly, deep learning and optimized ensemble techniques have proven highly effective in assessing the intrinsic mechanical and shear strength properties of rock masses, tasks that are fundamental to geotechnical design (**Ding et al., 2025; Ding et al., 2024**). These studies consistently highlight that hybrid models, particularly those enhanced by metaheuristic algorithms, can capture the high non-linearity and complex interactions inherent in rock engineering systems. However, while these intelligent systems have been extensively validated for predicting rock behaviour and the consequences of engineering actions upon it, their application to the proactive optimization of operational inputs – specifically, the properties of drilling fluid – remains comparatively underexplored. Building upon this established precedent of ML success in geomechanics, the present study aims to bridge this gap by adapting and evaluating a suite of optimized metaheuristic algorithms for the simultaneous prediction of mud weight and type, a critical operational parameter in wellbore stability.

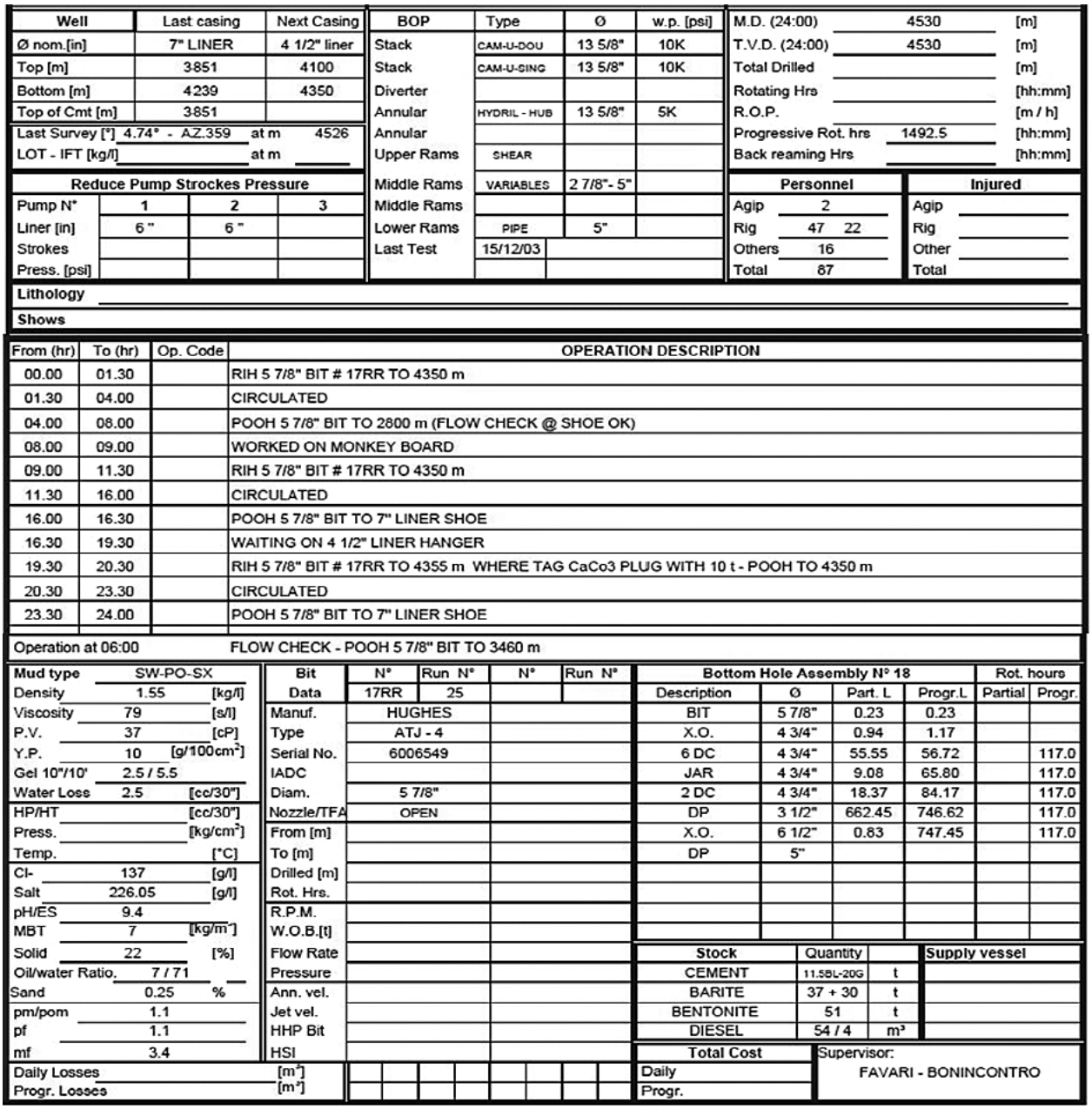
So the challenges of mud prediction include estimation of the mud weight and determination of the chemical properties of the mud. The first part can be named mud-weight window prediction.

2.1. Mud-weight prediction models

In order to determine the mud weight, the induced stress is first calculated and then compared with the failure criteria of the specific rock formation (Mohr-Coulomb, Hoke-Brown, etc.). The upper limit of the mud weight can be determined by substituting the stress induced in tensile failure. If this stress equals the shear strength of the rock, a lower mud weight limit should be estimated.

Researchers used a numerical method and sensitivity analyses to determine a safe mud-weight window (**Ayoub et al., 2019**). The results revealed that, for low-cohesion and friction angle formation, the mud weight window will become narrower. A decrease in the pore pressure and ratio between the maximum and minimum horizontal stresses will widen the mud weight window. This does not simulate the complex nature of the reservoir, such as natural fracture, bedding plane, high pressure and temperature of the reservoir. Additionally, the input parameters of numerical modelling such as the Young's modulus, cohesion and friction angle contain uncertainties.

Zahiri et al first derived the mud-weight window using an analytical approach for three wells and then use an ANN technique to derive a relation between the well-logging data and the mud-weight window. The major disadvantage was that ANN training using the analytical approach included simplified assumptions. It also did not consider chemical instability (**Zahiri et al., 2018**).



Well	Last casing	Next Casing
Ø nom.[in]	7" LINER	4 1/2" liner
Top [m]	3851	4100
Bottom [m]	4239	4350
Top of Cmt [m]	3851	
Last Survey [°] 4.74° - AZ.359	at m	4526
LOT - IFT [kg/l]	at m	

Reduce Pump Strockes Pressure			
Pump N°	1	2	3
Liner [in]	6 "	6 "	
Strokes			
Press. [psi]			

BOP	Type	Ø	w.p. [psi]
Stack	CAM-U-DOU	13 5/8"	10K
Stack	CAM-U-SING	13 5/8"	10K
Diverter			
Annular	HYDRIL - HUB	13 5/8"	5K
Annular			
Upper Rams	SHEAR		
Middle Rams	VARIABLES	2 7/8"- 5"	
Middle Rams			
Lower Rams	PIPE	5"	
Last Test	15/12/03		

M.D. (24:00)	4530	[m]
T.V.D. (24:00)	4530	[m]
Total Drilled		[m]
Rotating Hrs		[hh:mm]
R.O.P.		[m / h]
Progressive Rot. hrs	1492.5	[hh:mm]
Back reaming Hrs		[hh:mm]

Personnel		Injured	
Agip	2	Agip	
Rig	47 22	Rig	
Others	16	Other	
Total	87	Total	

Lithology

Shows

From (hr)	To (hr)	Op. Code	OPERATION DESCRIPTION
00.00	01.30		RIH 5 7/8" BIT # 17RR TO 4350 m
01.30	04.00		CIRCULATED
04.00	08.00		POOH 5 7/8" BIT TO 2800 m (FLOW CHECK @ SHOE OK)
08.00	09.00		WORKED ON MONKEY BOARD
09.00	11.30		RIH 5 7/8" BIT # 17RR TO 4350 m
11.30	16.00		CIRCULATED
16.00	16.30		POOH 5 7/8" BIT TO 7" LINER SHOE
16.30	19.30		WAITING ON 4 1/2" LINER HANGER
19.30	20.30		RIH 5 7/8" BIT # 17RR TO 4355 m WHERE TAG CaCo3 PLUG WITH 10 t - POOH TO 4350 m
20.30	23.30		CIRCULATED
23.30	24.00		POOH 5 7/8" BIT TO 7" LINER SHOE

Operation at 06:00 FLOW CHECK - POOH 5 7/8" BIT TO 3460 m

Mud type	SW-PO-SX	
Density	1.55	[kg/l]
Viscosity	79	[s/l]
P.V.	37	[cP]
Y.P.	10	[g/100cm²]
Gel 10"/10'	2.5 / 5.5	
Water Loss	2.5	[cc/30"]
HP/HT		[cc/30"]
Press.		[kg/cm²]
Temp.		[°C]
Cl-	137	[g/l]
Salt	226.05	[g/l]
pH/ES	9.4	
MBT	7	[kg/m³]
Solid	22	[%]
Oil/water Ratio.	7 / 71	
Sand	0.25	%
pm/pom	1.1	
pf	1.1	
mf	3.4	

Bit Data	N°	Run N°	N°	Run N°
	17RR	25		
Manuf.	HUGHES			
Type	ATJ - 4			
Serial No.	6006549			
IADC				
Diam.	5 7/8"			
Nozzle/TFA	OPEN			
From [m]				
To [m]				
Drilled [m]				
Rot. Hrs.				
R.P.M.				
W.O.B.[t]				
Flow Rate				
Pressure				
Ann. vel.				
Jet vel.				
HHP Bit				
HSI				

Bottom Hole Assembly N° 18				Rot. hours	
Description	Ø	Part. L	Progr.L	Partial	Progr.
BIT	5 7/8"	0.23	0.23		
X.O.	4 3/4"	0.94	1.17		
6 DC	4 3/4"	55.55	56.72		117.0
JAR	4 3/4"	9.08	65.80		117.0
2 DC	4 3/4"	18.37	84.17		117.0
DP	3 1/2"	662.45	746.62		117.0
X.O.	6 1/2"	0.83	747.45		117.0
DP	5"				

Stock	Quantity		Supply vessel
CEMENT	11.5BL-20G	t	
BARITE	37 + 30	t	
BENTONITE	51	t	
DIESEL	54 / 4	m³	

Daily Losses	[m³]
Progr. Losses	[m³]

Total Cost	Supervisor:
Daily	FAVARI - BONINCONTRO
Progr.	

Figure 5. Daily drilling report for well 19

Agwu et al. used an ANN model to forecast the density of oil-based mud in high temperature/high pressure (HTHP) wells. They summarized previous empirical and AI approaches to determining the mud density in HTHP wells (**Agwu et al., 2020**). Sabah et al. used machine learning to predict the amount of lost circulation. This parameter was essential for selecting an appropriate lost circulation material to cope with. The data was extracted from 2820 daily drilling reports from 305 oil wells and 18 input parameters were used to predict the amount of circulation lost into the formation (**Sabah et al., 2021**).

Beheshtian et al. gathered 3389 daily drilling reports from three gas wells to predict a safe mud-weight window using several machine learning algorithms (**Beheshtian et al., 2022**). Gowida et al. used the ANN approach to predict the minimum and maximum mud weights (**Gowida et al., 2022**). Phan et al. found that a model was required to predict the real-time mud weight. They reported that analytical and numerical models for this purpose are time consuming and that the AI based method was the best solution for real-time mud-weight prediction (**Phan et al., 2022**). As mentioned earlier prediction drilling fluid is divided into two-part mud weight and mud type. In the next section, the mud-type prediction works were reviewed.

2.2. Mud type prediction models

A greater mud-weight will reduce the tangential stress and increase the normal stress on the wellbore; however,

Table 2. Descriptive statistics of database based on MUD type

Mud Type	TVD (m)	ROP (m/h)	Diameter (m)	RPM (1/min)	WOB (kN)	Flow Rate (kg/s)	Mud Density (kg/m³)
FW-GE	176.46	2.20	0.60	90.71	18.93	1392.87	1339.83
FW-GE SPUD	65.00	1.55	0.61	130.00	6.67	1270.06	1318.44
FW-GE-PO	258.00	3.35	0.61	130.00	20.02	1587.57	1390.44
FW-GEL	201.00	2.14	0.61	100.00	24.47	1723.65	1330.44
FW-GY	2669.11	0.76	0.33	151.87	48.73	1207.73	1600.99
FW-KC	1440.37	2.17	0.42	117.37	39.57	1616.95	1462.99
FW-KC-PC	1491.52	2.38	0.42	137.16	50.28	1605.00	1484.55
FW-LS	4476.00	0.69	0.21	150.00	19.28	718.18	1983.66
FW-PO	2905.80	3.14	0.31	220.00	13.88	1319.95	1618.20
FW-S-PO	1539.67	1.68	0.21	96.67	37.81	801.31	1378.44
FW-SW-GE	1070.10	1.86	0.61	50.00	46.71	1769.01	1246.44
SW	3136.95	0.20	0.31	110.00	6.67	861.83	1744.11
SW-GE	1514.76	2.24	0.43	150.04	63.66	1597.40	1509.33
SW-GE-PO	2114.45	1.53	0.39	162.59	44.18	1480.65	1536.66
SW-LS-LI	4333.72	0.62	0.17	106.41	17.64	490.73	1920.44
SW-PO	2706.99	1.54	0.34	148.78	41.34	1273.12	1647.99
SW-PO-LS	3473.40	0.43	0.31	136.21	37.31	1084.39	1642.77
SW-PO-SX	4030.80	0.87	0.24	145.85	34.28	844.10	1860.00
PO-LS	2492.00	0.16	0.31	129.40	9.34	900.00	1476.66
Total	2892.23	1.34	0.33	143.16	39.73	1167.52	1676.44
min	17.00	0.03	0.15	10.00	0.00	222.26	1210.44
max	22121.00	10.56	0.61	650.00	122.33	1814.37	2253.33

FW = fresh water; SW = sea water; PO = polymer; KC = potassium chloride; GE = gel; LI = lime mud; SX = soltex; LS = lingosulfonate; GY = gypsum base; PC = partially hydrolyzed polyacrylamide.

in some situations this has an adverse effect on wellbore stability. When mechanical solutions are not appropriate, a chemical analysis of the drilling fluid should be carried out. Most chemical analyses are based on an experimental approach by measuring parameters such as the cation exchange capacity (CEC) and x-ray diffraction (XRD) to determine the active clay mineral content. The choice of drilling fluid is based on the composition of the formation. For example, shale swells slightly when sodium silicate is added to polymer mud. However, the addition of potassium chloride to the polymer mud produced significant swelling, indicating that the mud consisted of various chemical constituents which interacted with the formation, especially shale (**Allawi, 2023**).

Swelling shale has low strength and tends toward time-dependent wellbore instability. Traditionally, oil-based mud has been employed in such formations; however, because of its negative environmental impact, other types of mud, such as polymer/KCL/lime/gypsum have been used instead. Few works have used AI to estimate mud type, but some researchers have predicted the rheological properties of mud and the density of its constituents (**Agwu et al., 2020**).

Tatar et al. used pressure, temperature and concentration as inputs to predict the brine density of mud (**Tatar et al, 2015**). Aljubran et al. determined the mechanical-thermal/physico-chemical interaction on wellbore instability. They reported that cooler temperatures and higher salt concentrations in the mud improved the wellbore stability in a shale formation (**Aljubran et al., 2022**).

Due to the lack of research that comprehensively solves drilling fluid prediction, this work used a data mining approach to find mud's physical and chemical properties. In the following, the details of the research are explained.

3. Methods

A sample daily drilling report is shown in **Figure 5**. This report records mud type, bit data, bottom hole assembly, and pressure. Such reports are the most reliable data because of gathering during daily drilling and on the well location.

3.1. Data Analysis

The initial determination of the dataset's descriptive statistics involved a comprehensive examination of each

parameter through mathematical calculations. It was ascertained that the highest correlation with mud density across all parameters was observed when the relationship between them followed a power and linear equation. The selection of this equation in the present research was made to encompass all three algorithms. **Tables 1 and 2** display the descriptive statistics of the database.

The general form of the predictive equation used in this investigation is given by **Equation 1**. In this model, the output (predicted mud density) is a function of six standard drilling parameters, each multiplied by a specific weight coefficient (W~i~) that is optimized by the algorithms. The coefficients determine the relative influence and scaling of each input parameter.

$$\begin{aligned} &\mathrm{TVD} \times \mathrm{W1} + \mathrm{ROP} \times \mathrm{W2} + \mathrm{W3} \times \mathrm{diam} + \\ &+ \mathrm{RPM} \times \mathrm{W4} + \mathrm{W5} \times \mathrm{WOB} + \mathrm{flow\ rate} + \mathrm{W6} \end{aligned} \quad (1)$$

Where:
TVD is the True Vertical Depth (m),
ROP is the Rate of Penetration (ft/h),
diam is the drill bit diameter (inches)
RPM is the drill string Revolutions Per Minute (1/min),
WOB is the Weight On Bit (kip),
flow rate is the mud flow rate (lb/s),
W1, W2, W3, W4, W5, W6 are the weight coefficients optimized by the ACO, BCO, and EPC algorithms.

3.2. Evaluation criteria

This investigation evaluated the accuracy and efficiency of the predictive models using the coefficient of determination (R^2) and the root mean-square error (RMSE), as stipulated in **Equations 2** and **3**. These metrics are well-established in the literature for assessing the performance of machine learning models in geotechnical and mining engineering applications, providing a standardized and interpretable measure of predictive power (**Hasanipanah et al., 2022; Hasanipanah et al., 2015**). The ideal values for R^2 and RMSE are one and zero, respectively. Utilizing distribution diagrams and comparative graphs of observational versus computational data, the results were analyzed and compared.

3.2.1. Coefficient of determination

The coefficient of determination measures the explanatory power of a model by quantifying the percentage of variance in the dependent variable that can be accounted for by the independent variables. Specifically, it represents the total variation in the dependent variable as the sum of the variation explained by the regression and the variation not explained by the regression. This coefficient provides a probabilistic estimate of the correlation between two sets of data in the future, based on a defined mathematical model that conforms to existing data (**Allawi & Al-Jawad, 2021**).

The R^2 is utilized as a key measure for evaluating the precision of the regression model in depicting variables, with a higher value denoting a more suitable fit. It demonstrates the concordance between the recorded and anticipated values, which can be assessed using parsing and fitting techniques. The coefficient of determination delineates a segment of the overall variance in the recorded values that can be explained by simulated values, ranging from zero to one, where an ideal value of one signifies impeccable alignment between the simulated and recorded values (**Phan et al., 2022**). Its robustness and intuitive interpretation have made it a standard metric in intelligent systems developed for geotechnical prediction, such as forecasting rock mass properties and environmental hazards (**Hasanipanah et al., 2022; Hasanipanah et al., 2015**).

Equation 2 defines the relationship between the computational value $x\ i$ and the observational value $y\ i$ for time-step i. The variables N, $\bar{x}$, and $\bar{y}$ denote the total number of time-steps and the average computational and observational values, respectively.

$$R^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \ \sum_{i=1}^{N}(y_i - \bar{y})^2}} \quad (2)$$

3.2.2. Root mean square error (RMSE)

The RMSE is a function of the fit or objective function and is defined as the square root of the mean squared error value. The RMSE serves as a measure of the absolute error between predicted and observable values. The range of this statistical index is from zero to infinity, with lower values indicating better simulation performance. The optimal value of the RMSE is zero. The mathematical expression for the RMSE is:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (3)$$

3.3. Ant Colony Optimization (ACO)

Ant colony optimization can be modeled as P= (S, Ω, f), where S is the set of all finite sets of discrete decision variables, and Ω represents the constraints between these variables and a target function f: $S \longrightarrow R0+$, subject to minimization or maximization (**Dorigo & Blum, 2005**). Ant colony optimization incrementally generates solutions by considering the likelihood of solution components, with probabilities determined by pheromone levels (**Dorigo & Blum, 2005**). When applied to hybrid optimization problems, a set of solution components is defined based on the problem formulation. During solution construction, ants select ci from $N(s^p)$ using **Equation 4** at each step.

$$p(c_{ij}|s^p) = {c_{ij}}^p \cdot \frac{\eta(cij)^\beta}{\sum c_{il \in N(s^p)} \tau ij^\alpha . \eta\,(cij)^\beta}, \ \forall c_{ij} \in N(s^p) \quad (4)$$

In pheromone-based optimization, τij represents the pheromone level on edge cij. A weight function, $\eta(0)$, updates edges $cij \in N(s^p)$ at each iteration, producing values termed "apocalyptic information." Positive parameters α and β govern the relationship between pheromone and heuristic information. A Gaussian kernel, used for sampling, is defined as the sum of weighted one-dimensional Gaussian functions, $Gi(x)$.

$$G^i(x) = \sum_{l=1}^{k} \omega_l g_l^i(x) = \sum_{l=1}^{k} \omega_l \frac{1}{\sigma_l^i \sqrt{2\pi}} e^{-\frac{\left(x-\mu_l^i\right)^2}{2\sigma_l^{i^2}}} \quad (5)$$

Multi-core probability density functions determine a single probability density function based on the problem's dimension (i = 1,…, n). Each $Gi(x)$ is represented using three parametric vectors: ω (weights), μi (means), and δi (standard deviations), along with one-dimensional Gaussian functions, g_l^i (x). These vectors have cardinality equivalent to the number of Gaussian functions (denoted by k) in the Gaussian kernel.

$$| \omega | = | \mu i | = | \delta i | = k \quad (6)$$

Ant Colony Optimization (ACO) is governed by several control parameters. α controls the influence of pheromone trails, with higher values favoring exploitation. β controls the influence of heuristic information, with higher values favoring exploration. ρ represents the pheromone evaporation rate, preventing premature convergence; higher values increase exploration. The number of ants affects search thoroughness, with more ants increasing computational cost. Q is a constant related to pheromone deposition. These parameters are typically tuned empirically, using literature-based values, or through computationally expensive meta-optimization.

3.4. Bee Colony Optimization (BCO)

The freak algorithm utilizes a freak colony with three groups hired, hunt, and watch notions. Each food source represents a implicit result, and its quencher volume indicates result quality (**Karaboga, 2005**). Originally, a population of results, corresponding to food source locales, is generated.' SN' denotes the number of hired/ hunt notions. Each result (j = 1.2.3.. SN) is a D- dimensional vector, where' D' is the number of optimization parameters. Search notions elect food sources grounded on quality; the probability of opting a source is calculated using **Equation 7**.

$$P_{i=\frac{fit_i}{\sum_{N-1}^{SN} fit_N}} \quad (7)$$

In the context of statistical analysis, the variable $fiti$ represents the fitness value of the i^{th} observation.

The process of selecting a novel food source (Vij) is determined by **Equation 8**, which is contingent upon the preceding food source (Xij). This approach is adopted within an academic context.

$$Vij = Xij + (Xij - Xkj) \quad (8)$$

Random pointers j and k, where j and k are rudiments of the set{ 1.2.. SN}, are named. It's important to note that k is named aimlessly, but must be distinct from j. In the environment of the freak algorithm, a food source that fails to recover after a specified replication is supposed an abandoned food source. In similar cases, the notions cover the situation in agreement with **Equation 9** and latterly replace the abandoned food source with a new one named at arbitrary (**Karaboga & Basturk, 2008**).

$$X_i^j = X_{min}^j + \text{rand}[0.1](X_{max}^j - X_{min}^j) \quad (9)$$

The variable j denotes the quantity of optimization variables.

In BCO, key control parameters include colony size (SN), representing the number of food sources or solutions, and Limit, which dictates when an unproductive food source is replaced by a scout bee to escape local minima. The maximum cycle number (MCN) serves as the stopping criterion, defining the maximum number of iterations.

In the ABC algorithm, foundational literature (**Karaboga, 2005**) typically suggests setting the Limit parameter to SN * D, where D represents the number of dimensions. Here, D=6 (TVD, ROP, Diam, RPM, WOB, Flowrate). We systematically varied SN and Limit to optimize performance for this specific problem. The omission of this detail regarding the Limit parameter is significant, as it plays a crucial role in balancing exploration and exploitation within the ABC algorithm.

3.5. Emperor Penguins Colony (EPC)

The foundational premise of the Emperor Penguin algorithm posits that penguins are distributed across the examined environment. Consequently, at the outset, prior to the introduction of more complex procedures, it is essential to assess the relative positions of each penguin in relation to one another. It is important to note that a penguin typically gravitates towards another that presents a lower absorption cost. Thus, the second critical measurement involves determining the absorption cost for each penguin, which is influenced by both heat intensity and distance. Throughout the execution of this algorithm, each phase of absorption entails the evaluation of a new potential solution to the problem at hand, accompanied by an update of the heat intensity. Following this, all potential solutions are organized, and the optimal one is identified. Additionally, a damping ratio is calculated for each penguin, which is influenced by various factors, including the individual penguin's movement, heat radiation, and heat absorption. A lower damping ratio indicates a more favorable solution (**Harifi et al., 2019**).

Emperor Penguin Colony (EPC) is a metaheuristic algorithm that mimics the huddling behaviour of emperor penguins seeking warmth.

Table 3. ACO results for six models

Model	Training Dataset	Testing Dataset	Coefficient of weighting						Model evaluation	
			W1 TVD	W2 ROP	W3 diameter	W4 RPM	W5 WOB	W6 flow rate	R^2	RMSE
1	P2-P3-P4	P1	-0.0078	0.0432	-0.4567	0.1324	-2.3456	0.8765	0.6589	0.1245
2	P1-P3-P4	P2	-0.0089	0.0456	-0.5678	0.1234	-2.8765	0.7654	0.7890	0.1434
3	P1-P2-P4	P3	-0.0086	0.0567	-0.4433	0.1259	-2.1234	0.6890	0.8896	0.1278
4	P1-P2-P3	P4	-0.0085	0.0642	- 0.4567	0.1347	-2.0987	0.8976	0.7644	0.1239
5	Entire training dataset.	P1	-0.0064	0.0778	-0.6555	0.1447	-1.3364	0.7651	0.9176	0.0987
		P2	-0.0064	0.0778	-0.6555	0.1447	-1.3364	0.7651	0.9034	0.0876
		P3	-0.0064	0.0778	-0.6555	0.1447	-1.3364	0.7651	0.9078	0.0912
		P4	-0.0064	0.0778	-0.6555	0.1447	-1.3364	0.7651	0.9025	0.0789
6	80% of all dataset	20% of all dataset	-0.0012	0.047	-0.5824	0.2234	-1.1345	0.5612	0.9649	0.0364

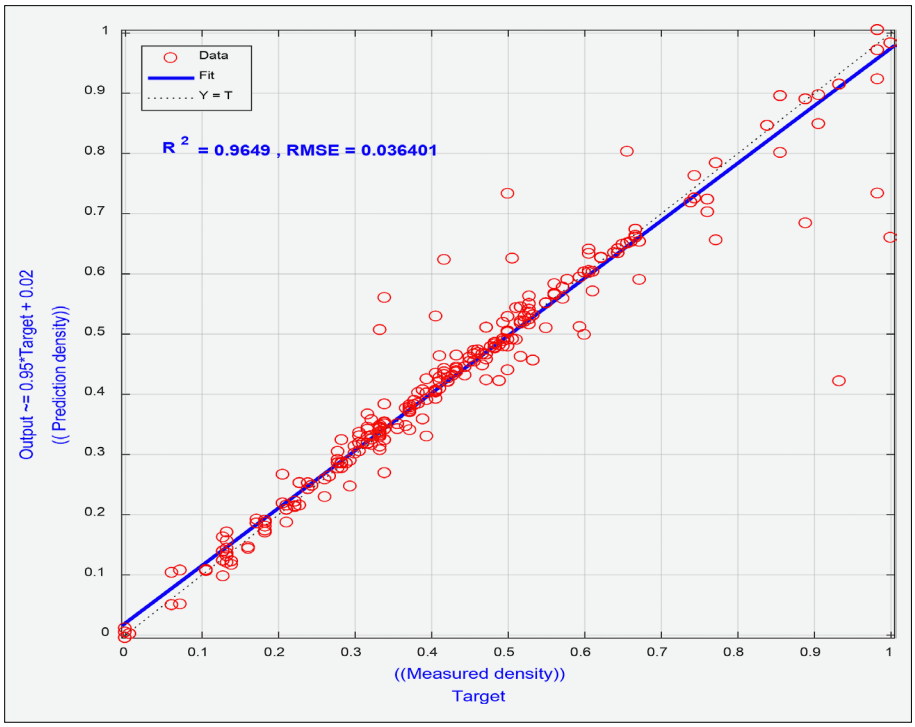


Figure 6. Distribution of top model by ACO

Key control parameters include:

- Population Size (N): the number of penguins in the colony.
- Heat Absorption Cost Function Parameters (A, B): these parameters govern penguin movement and huddling behaviour based on metaphorical temperature and distance calculations.
- Damping Ratio (R): simulates reduced heat loss during huddling, influencing convergence.
- Maximum Iteration Number: the stopping criterion.

Parameter Value Selection based on the algorithm's recent development, the original paper (**Harifi et al., 2019**) is often consulted for default values. Sensitivity analysis may also be performed to assess the impact of parameter variations on model output for specific applications like drilling fluid prediction.

4. Results

This section presents the comprehensive results of applying the three nature-inspired optimization algorithms – Ant Colony Optimization (ACO), Bee Colony Optimization (BCO), and Emperor Penguins Colony (EPC) – to predict drilling mud weight and identify the optimal mud type. The performance of each algorithm is evaluated based on the coefficient of determination (R^2) and root mean square error (RMSE), as defined in the methodology. Furthermore, the final weighting coefficients (W1 to W6) derived from the optimal model for each algorithm are presented, revealing the relative influence of the input parameters (TVD, ROP, Diameter, RPM, WOB, Flow Rate) on the predicted mud density. The results are systematically organized to first detail the predictive performance and derived equations for each al-

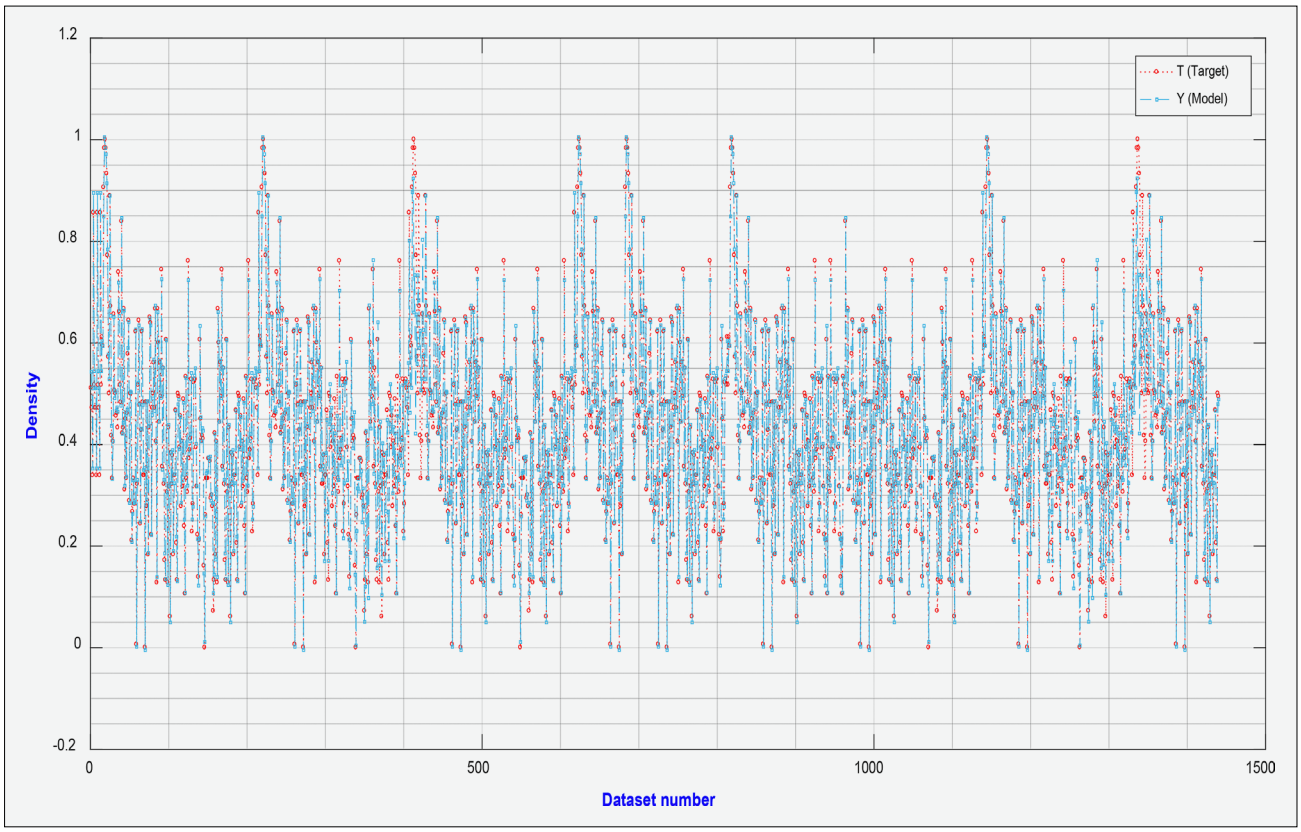


Figure 7. Matching of top model by ACO

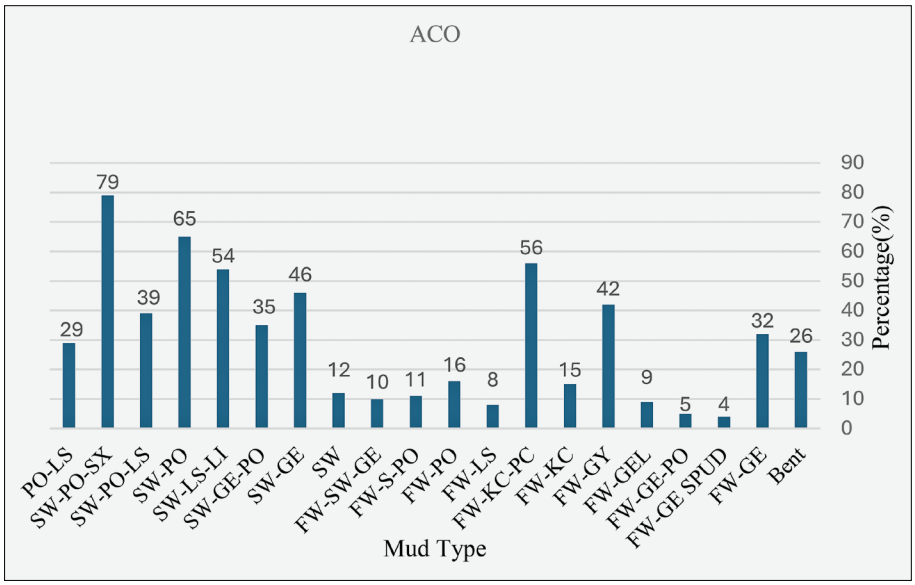


Figure 8. Best mud type by ACO

gorithm individually, followed by a comparative analysis that highlights the most effective model and its practical implications for optimal fluid selection in the studied field.

4.1. Ant colony optimization Result (ACO)

Models 1-4 used a four-fold cross-validation approach, training on three steps and testing on the remaining one. Model 5 trained on the entire dataset and was then tested on datasets P1-P4. Model 6 trained on 80% of the data and tested on the remaining 20%. Model 6 demonstrated sufficient accuracy, achieving an R2 of 0.9649 and RMSE of 0.0364. **Table 3** presents R2, RMSE, and coefficient weighting across all models. **Figures 6** and **7** illustrate the distribution and correlation between measured, target, and predicted values for the top-performing model. **Figure 8** identifies the optimal mud type according to ACO.

Table 4. BCO Algorithm Results: Six Models

Model	Training Dataset	Testing Dataset	Coefficient of weighting						Model evaluation	
			W1 TVD	W2 ROP	W3 diameter	W4 RPM	W5 WOB	W6 flow rate	R^2	RMSE
1	P2-P3-P4	P1	-0.0023	0.0876	-0.3321	0.4567	-3.7654	0.6543	0.9086	0.3421
2	P1-P3-P4	P2	-0.0014	0.0789	-0.3346	0.3257	-2.4532	0.8976	0.8765	0.4456
3	P1-P2-P4	P3	-0.0015	0.0456	-0.5567	0.5678	-2.5643	0.6532	0.7864	0.3245
4	P1-P2-P3	P4	-0.0054	0.0897	- 0.7864	0.3245	-2.9543	0.9986	0.8324	0.4478
5	Trained on 100% of dataset.	P1	-0.0022	0.0332	-0.4567	0.2223	-1.1123	0.4532	0.9132	0.1265
		P2	-0.0022	0.0332	-0.4567	0.2223	-1.1123	0.4532	0.9543	0.1870
		P3	-0.0022	0.0332	-0.4567	0.2223	-1.1123	0.4532	0.9325	0.1032
		P4	-0.0022	0.0322	-0.4567	0.2223	-1.1123	0.4532	0.9143	0.0989
6	80% of all dataset	20% of all dataset	-0.0019	0.0664	-0.2365	0.2987	-1.9433	0.2137	0.9841	0.0245

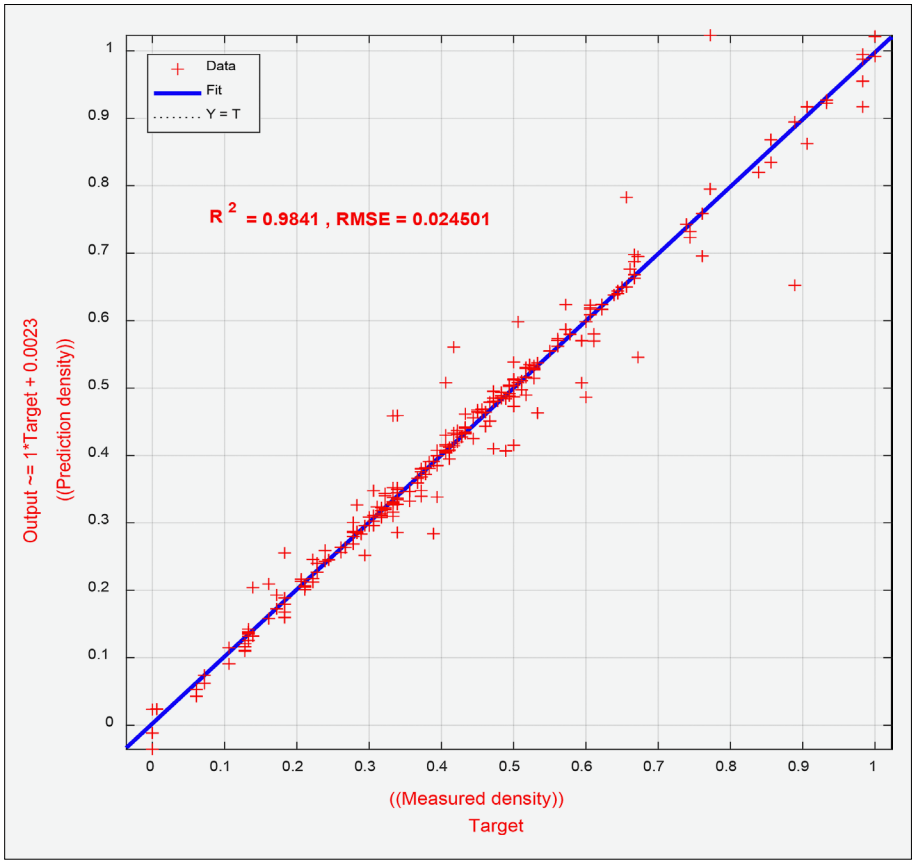


Figure 9. Distribution of top model by BCO

The equation obtained using ACO was:

$$-0.0012 \times TVD + ROP^{0.047} - 0.5824 \times diam + RPM^{0.2234} - 1.1345 \times WOB + flowrate^{0.5612} \quad (10)$$

4.2. Bee colony optimization results (BCO)

BCO algorithm Models 1-4 used a four-step partition: each step tested the targets while the remaining steps trained the sets. Model 5 trained on the entire dataset, then each step (P1-P4) was used for testing. Model 6 trained on 80% of the dataset and tested on the remaining 20%. The results showed Model 6 to be sufficiently accurate, yielding an R2 of 0.9841 and RMSE of 0.245 for the best prediction. **Table 4** presents the R2, RMSE, and weighting coefficients for all BCO models. **Figures 9** and **10** display the distribution and matching charts, respectively, of measured and predicted values for the top model. **Figure 11** identifies the best mud type according to the BCO.

The equation obtained using BCO was:

$$-0.0019 \times TVD + ROP^{0.0664} - 0.2365 \times diam + RPM^{0.2987} - 1.9433 \times WOB + flowrate^{0.2137} \quad (11)$$

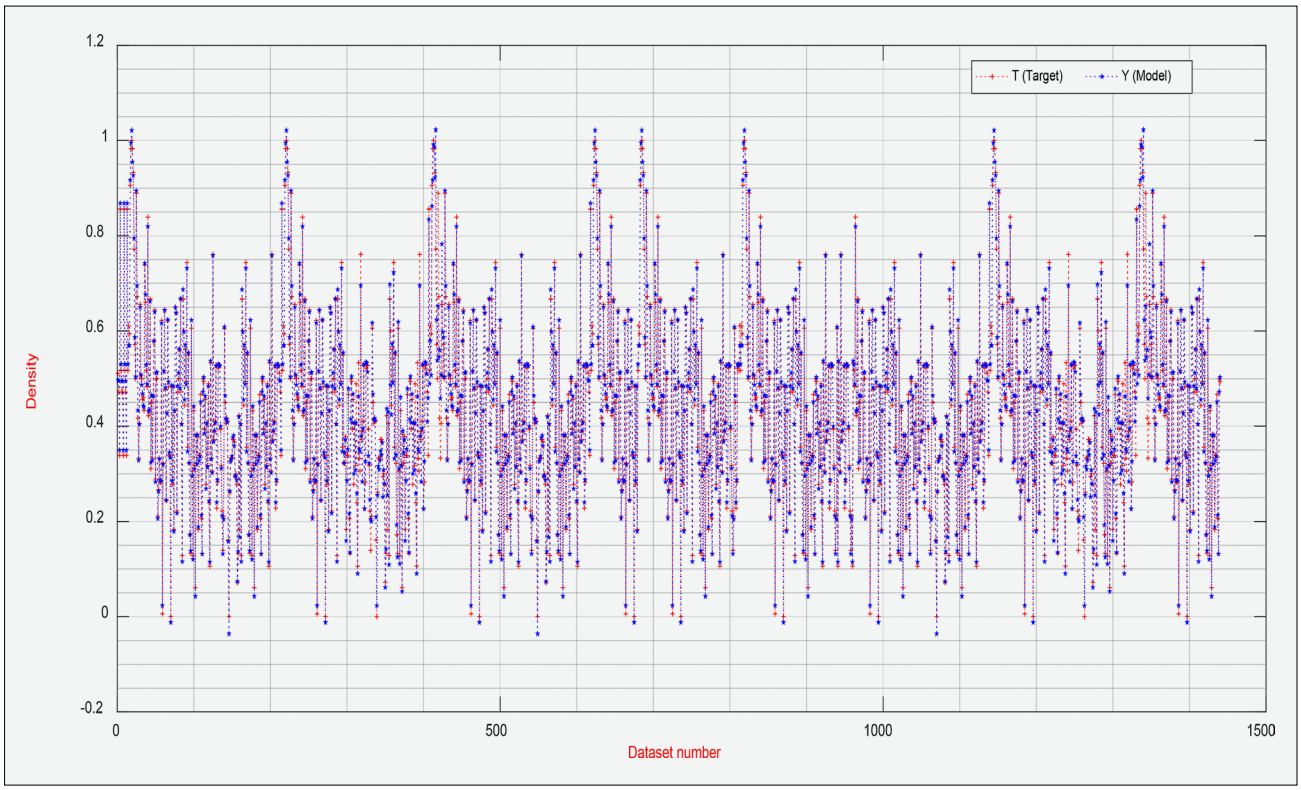


Figure 10. Matching of top model by BCO

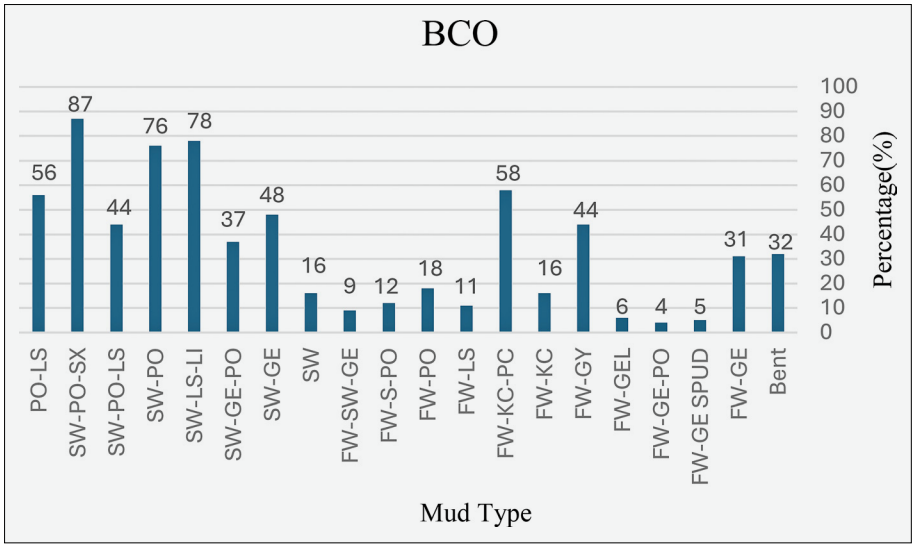


Figure 11. Best mud type by BCO

4.3. Results of Emperor Penguins Colon (EPC)

Models 1- 4 of the EPC algorithm used a four-step partitioning, employing one step for testing and the remaining three for training. Model 5 trained on the full dataset and also tested each step (P1-P4) across all models. Model 6 employed 80 of the dataset for training and 20 for testing. The results demonstrated that Model 6 and its associated equation displayed high delicacy, achieving an R2 of 0.9707 and an RMSE of 0.0332 in the stylish model. **Table 5** presents the R2, RMSE, and weighting portions for all EPC algorithm vaticination models. **Figures 12** and **13** illustrate the distribution and correlation, independently, between measured and prognosticated target values for the top- performing model. **Figure 14** identifies the optimal slush type grounded on the EPC.

The equation using EPC was:

$$-0.0056 \times TVD + ROP^{0.432} - 0.2223 \times diam + \\ + RPM^{0.3254} - 1.3456 \times WOB + flowrate^{0.3317} \quad (12)$$

An investigation was conducted on the modelling process and the outcomes are presented in **Table 6** and

Table 5. EPC Six models' algorithm results

Model	Training Dataset	Testing Dataset	Coefficient of weighting						Model evaluation	
			W1 TVD	W2 ROP	W3 diameter	W4 RPM	W5 WOB	W6 flow rate	R^2	RMSE
1	P2-P3-P4	P1	-0.0055	0.0666	-0.7787	0.4567	-4.2246	0.4321	0.8765	0.1789
2	P1-P3-P4	P2	-0.0065	0.0876	-0.8876	0.6543	-4.5678	0.3456	0.8790	0.1897
3	P1-P2-P4	P3	-0.0067	0.0654	-0.7654	0.3245	-2.5678	0.7896	0.7890	0.1765
4	P1-P2-P3	P4	-0.0076	0.0897	- 0.6543	0.3456	-2.7654	0.6542	0.8765	0.1123
5	Entire dataset used for model training.	P1	-0.0044	0.0556	-0.2345	0.3345	-2.1346	0.4567	0.9234	0.0779
		P2	-0.0044	0.0556	-0.2345	0.3345	-2.1346	0.4567	0.9345	0.0987
		P3	-0.0044	0.0556	-0.2345	0.3345	-2.1346	0.4567	0.9214	0.0543
		P4	-0.0044	0.0556	-0.2345	0.3345	-2.1346	0.4567	0.9123	0.0635
6	80% of all dataset	20% of all dataset	-0.0056	0.0432	-0.2223	0.3254	-1.3456	0.3317	0.9707	0.0332

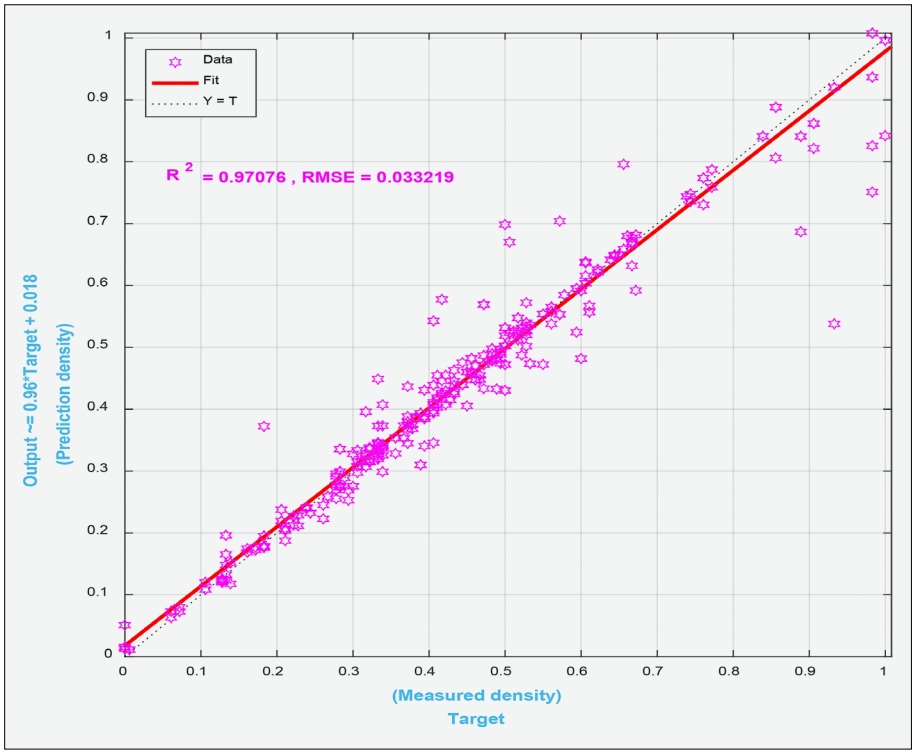


Figure 12. Distribution of top model by EPC

Figures 15 and **16** The results (predicted values) from the use of ACO, BCO and EPC were compared with the measured values (targets). These predicted values also were compared with the obtained values.

5. Discussion

The findings of this study underscore the transformative potential of machine learning in revolutionizing drilling fluid engineering. By successfully predicting both mud weight and type – a dual challenge often addressed in isolation – our framework provides a holistic, data-driven solution to a complex geomechanical problem. The following discussion elaborates on the performance of the algorithms, the critical influence of input parameters, the practical significance of the optimal mud type, and the broader implications for the industry.

5.1. Comparative Algorithm Performance and Robustness

The superior performance of the Bee Colony Optimization (BCO) algorithm (R^2 = 0.9841, RMSE = 0.0245) over Ant Colony Optimization (ACO) and Emperor Penguins Colony (EPC) can be attributed to its intrinsic search mechanisms. BCO employs a balanced combination of employed, onlooker, and scout bees, which allows for robust exploitation of promising solutions while simultaneously exploring new regions of the search space to avoid local optima. This is particularly advantageous for the high-dimensional, non-linear relationship

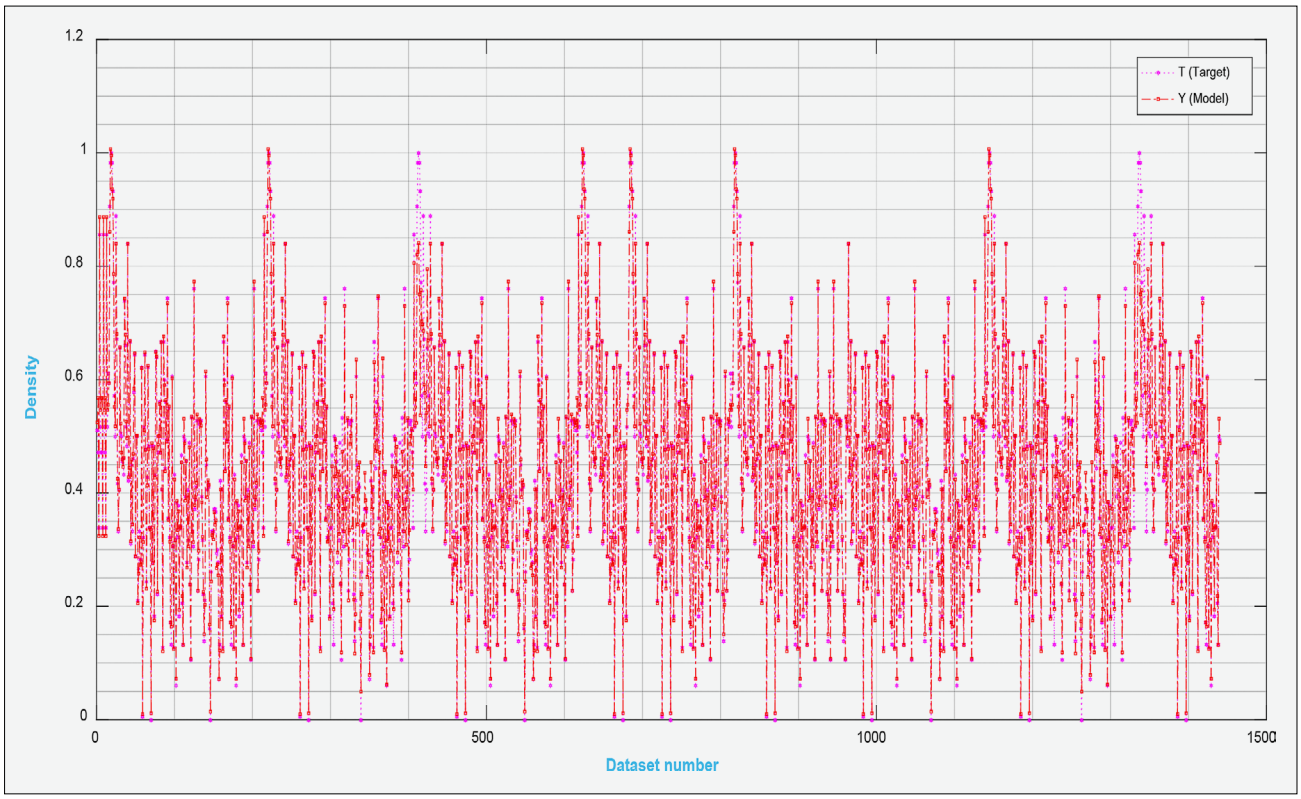


Figure 13. Matching top models by EPC

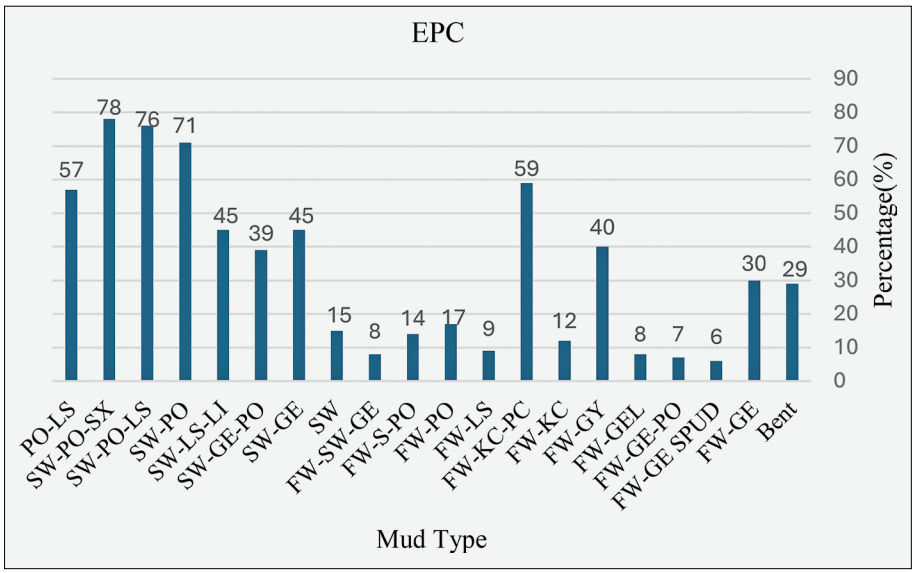


Figure 14. Best mud type by EPC

between drilling parameters and mud properties, where the risk of convergence to suboptimal solutions is high.

While ACO and EPC also delivered highly accurate results ($R^2 > 0.96$), their slightly lower performance may stem from their specific convergence behaviours. ACO's reliance on pheromone trails can sometimes lead to premature stagnation if the evaporation rate (ρ) is not optimally tuned. EPC, inspired by huddling behaviour, is effective for continuous optimization but might be less agile than BCO in discretely navigating the complex weight space for each parameter in our predictive equation. The consistent high performance across all three nature-inspired algorithms, however, validates the fundamental premise that metaheuristic-driven ML models are exceptionally well-suited for capturing the intricate rock-fluid interactions that elude traditional analytical methods.

5.2. Sensitivity Analysis and Parameter Influence

A critical step in validating any data-driven model is to interpret its inner workings. To this end, we conducted

Table 6. Comparison of models

Model	R^2	RMSE	Best Mud Type	Best Mud Type (%)
ACO	0.96	0.03	SW-PO-SX	79
BCO	0.98	0.02	SW-PO-SX	87
EPC	0.97	0.03	SW-PO-SX	78

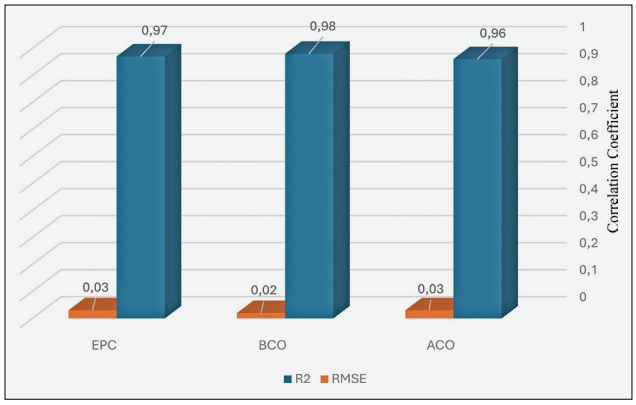


Figure 15: Comparison of models

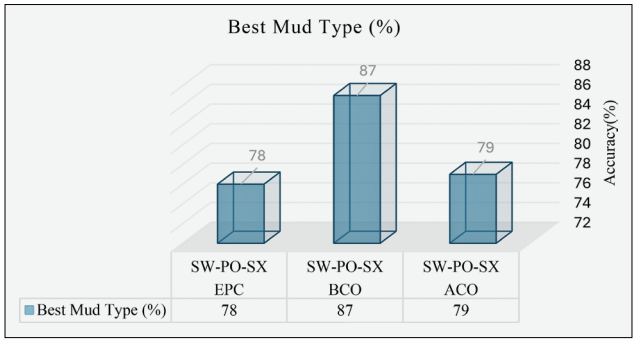


Figure 16. Best mud type from models

a comprehensive sensitivity analysis based on the final weight coefficients (Wi) from the optimal BCO model (Equation 11). The relative importance of each input parameter was quantified and is presented in **Figure 18**.

According to **Figure 18**, ROP (W2): 100% (Normalized Importance): as indicated by its high positive weight (0.0664), ROP emerged as the most influential parameter. This is mechanistically sound; ROP is a direct manifestation of the drill bit's interaction wth the formation. A high ROP often indicates softer, more porous, or fractured rock, which may require a different mud weight to stabilize and a specific chemical type to inhibit interaction. The strong correlation suggests that ROP acts as a real-time proxy for in-situ rock strength and lithology.

Weight on Bit (WOB) (W5): ~65% Importance - the significant negative weight (-1.9433) of WOB indicates an inverse relationship with mud density. Higher WOB typically leads to higher ROP and generates more cuttings, potentially necessitating mud system adjustments. Its high importance underscores its role as a primary controlled variable affecting the drilling process's mechanical efficiency.

Diameter (W3): ~35% Importance - the negative coefficient (-0.2365) suggests that larger boreholes may require slightly lower mud weights, possibly due to the reduced hoop stress and different cuttings load.

Flow Rate (W6): ~30% Importance - flow rate is crucial for hole cleaning. Its positive influence (0.2137) aligns with the need for adequate hydraulic energy to lift cuttings, which influences the effective mud density and rheology downhole.

RPM (W4): ~25% Importance - while important, RPM's lower influence suggests its effect might be partially captured by its correlation with ROP.

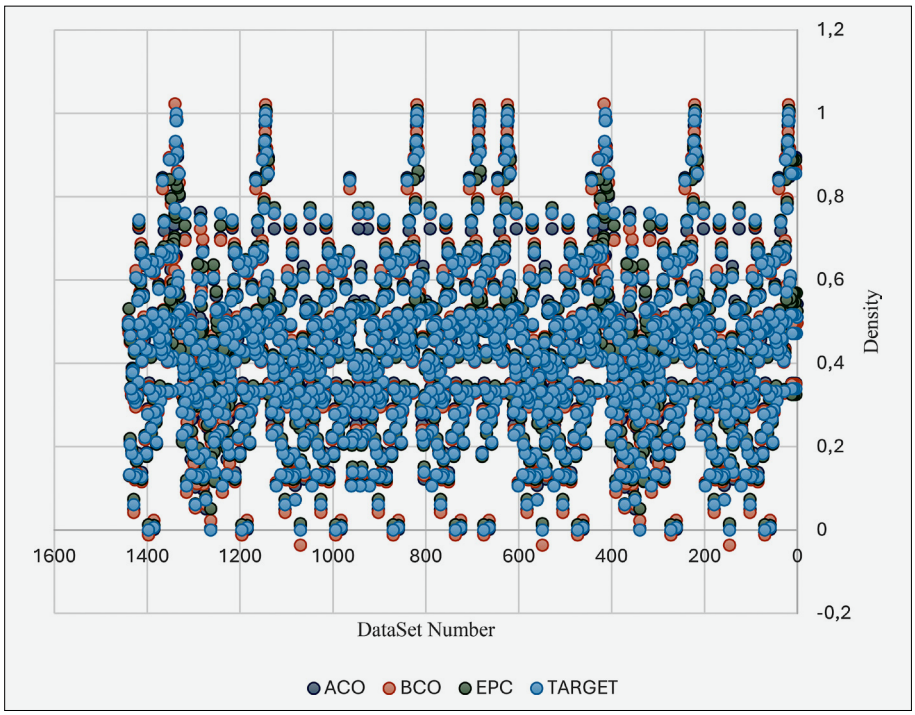


Figure 17. Comparison of results from ACO, BCO and EPC and target

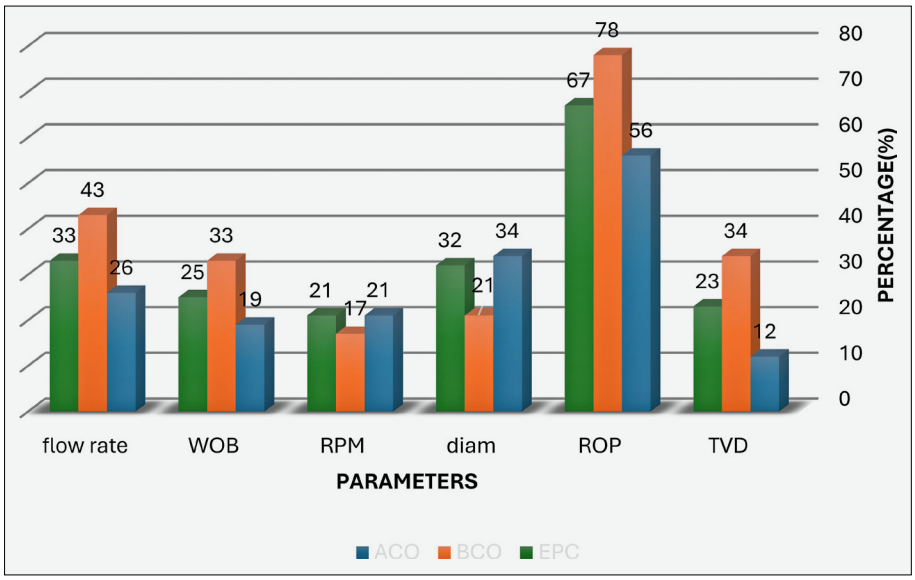


Figure 18. Parameter sensitivity analysis diagram

TVD (W1): ~3% Importance - surprisingly, TVD had the least direct influence in this model. This could be because the critical changes in mud properties are more directly linked to lithological changes and drilling mechanics at a specific depth rather than the depth itself, and this lithological information is indirectly captured by parameters like ROP and WOB.

This sensitivity analysis moves beyond mere prediction, offering valuable interpretability. It informs mud engineers that, for proactive mud management, real-time monitoring of ROP and WOB is paramount. The ROP model among all models is shown in **Figure 19**.

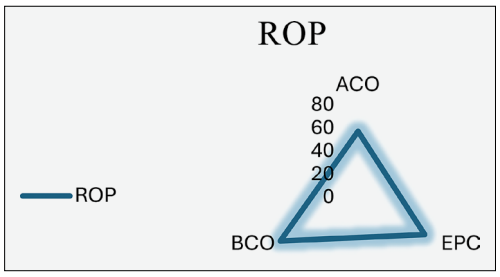


Figure 19. ROP in all models

Moreover, a 3D modelling approach illustrated the non-linear thresholds where ROP alterations necessitate immediate changes in mud composition. This finding corroborates the premise that real-time data integration with ML models enhances decision-making processes as opposed to relying solely on deterministic models, which may not account for the complexities of subsurface conditions, as shown in **Figure 20**.

5.3. *Practical Implications and Model Consensus on Mud Type*

A key practical outcome of this research is the strong, cross-model consensus identifying Sea Water-based mud

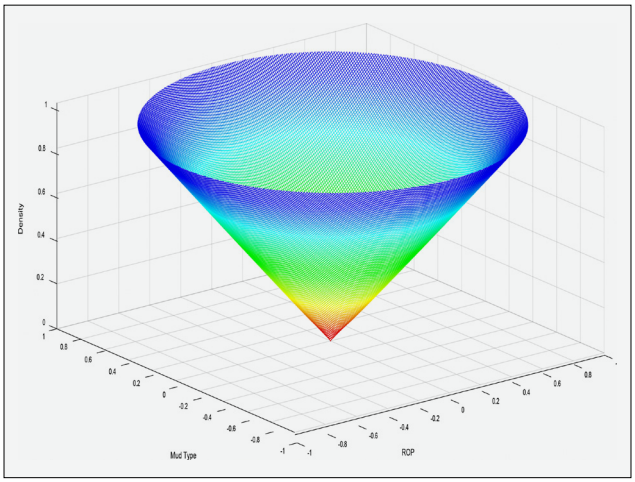


Figure 20. ROP in all models

with Polymer and Soltex (SW-PO-SX) as the optimal fluid for the studied field, with a confidence level of 79-87%. This recommendation is not a black-box output but a data-driven conclusion derived from 50 years of operational experience encoded in the daily drilling reports.

The SW-PO-SX system offers a balanced solution:

- Chemical Stability: the polymer and Soltex additives provide excellent shale inhibition, controlling clay swelling and dispersion in the active shale formations encountered, thereby mitigating chemical wellbore instability.
- Mechanical Stability: the model has learned the appropriate density range for this mud type to maintain wellbore pressure within the safe window.
- Environmental and Economic Advantage: as a water-based mud, SW-PO-SX presents a lower environmental footprint and is more economical than synthetic or oil-based alternatives, aligning with the industry's push towards more sustainable operations.

The high confidence level across three different algorithms significantly de-risks the fluid selection process for future wells in this field, transforming it from an experience-based art to a systematic, repeatable science.

6. Conclusions

This study successfully developed and validated a novel machine learning framework for the simultaneous prediction of drilling mud weight and mud type, a critical dual objective for ensuring wellbore stability and operational safety. By leveraging a comprehensive dataset extracted from 50 years of daily drilling reports across 20 oil wells, three nature-inspired optimization algorithms were employed and rigorously compared.

The key findings conclusively demonstrate that:

1. Data-driven prediction is a viable and superior paradigyym: machine learning models can effectively capture the complex, non-linear relationships between standard drilling parameters (TVD, ROP, RPM, WOB, diameter, flow rate) and the required mud properties, overcoming the simplifications inherent in traditional analytical and numerical methods.

2. Bee Colony Optimization (BCO) is the most effective algorithm: among the models tested, BCO emerged as the most precise predictor, achieving a correlation coefficient (R^2) of 0.9841 and a root-mean-square error (RMSE) of 0.0245 for mud density, highlighting its superior capability in navigating the complex solution space of this problem.

3. Rate of Penetration (ROP) is the dominant parameter: sensitivity analysis revealed that ROP is the most influential input variable, underscoring its role as a real-time indicator of formation characteristics and its critical importance for proactive mud management.

4. A consensus optimal mud type exists: all models converged with high confidence (78-87%) on Sea Water-based mud with Polymer and Soltex additives (SW-PO-SX) as the optimal fluid for the studied field, providing a robust, data-backed recommendation that reduces reliance on subjective experience.

6.1. Limitations of the Study

Despite the promising results, this research is subject to several limitations:

- Data Dependency and Quality: the model's performance is inherently tied to the quality, consistency, and comprehensiveness of the historical Daily Drilling Reports (DDRs). Inconsistencies in data recording, missing values, or human error in the original reports could propagate into the model.
- Feature Set Scope: the current model utilizes a set of common drilling parameters. It does not incorporate real-time mud logging data (e.g. gas units, cuttings morphology), lithology-specific logs (e.g. gamma ray, sonic), or precise downhole pressure measurements, which could further enhance predictive accuracy.
- Geological Specificity: the model was trained and validated on data from a specific oil field in South-West Iran. Its direct applicability to other fields with different geological settings, stress regimes, and formation fluids may be limited without retraining or transfer learning.
- Algorithmic Scope: the study focused on three nature-inspired metaheuristics. While they performed excellently, other powerful machine learning families (e.g. ensemble methods like Gradient Boosting or deep learning networks) were not explored for comparative benchmarking.

6.2. Suggested Improvements

To address the above limitations and strengthen the current model, the following immediate improvements are suggested:

1. Data Pre-Processing Enhancement: implement more advanced data imputation and cleansing techniques to handle missing or anomalous data points in the DDRs. Incorporating a domain-expert review to label data quality could also be beneficial.

2. Feature Engineering and Expansion: expand the input feature set to include mud rheological properties (e.g. plastic viscosity, yield point), lithology codes from mud logs, and estimated pore pressure and fracture gradient curves where available.

3. Hybrid Model Development: create a hybrid model that combines the global search capability of BCO with the strong predictive performance of an ensemble method (e.g. using BCO to optimize the hyperparameters of an Extreme Gradient Boosting (XGBoost) model).

6.3. Future Research Directions

This work opens several avenues for future research:

- Real-Time Integration and Deployment: the most critical future direction is the development of a real-time advisory system. This would involve integrating the trained model with a live data stream from the rig, enabling dynamic mud recommendations as drilling progresses.
- Generalizability and Transfer Learning: investigate the model's transferability to other oil and gas fields. Research could focus on developing a foundational model that can be fine-tuned with a smaller dataset from a new field, reducing the data requirement for deployment.
- Multi-Objective Optimization for Economics and Environment: expand the optimization objective beyond technical success to include economic and environmental factors. A future model could simultaneously optimize wellbore stability, drilling cost (e.g. mud cost, non-productive time), and the envi-

ronmental impact score, recommending the most sustainable and economical fluid.

- Causal Analysis: move beyond correlation to causality by exploring how the model's predictions change with specific geological events (e.g. entering a salt section, encountering a fault), enhancing interpretability and engineer trust.

In conclusion, this research provides a robust, data-driven foundation for systematic drilling fluid selection. By acknowledging its limitations and pursuing the outlined future directions, the industry can move closer to fully automated, adaptive, and optimized drilling fluid management, significantly enhancing safety, efficiency, and sustainability in hydrocarbon exploration.

7. References

Tatar, A., Halali, MA.,Mohamadi, AH. (2016). On the Estimation of the Density of Brine with an Extensive Range of Different Salts Compositions and Concentrations. *Journal of Thermodynamics & Catalysis*, *7*(2). https://doi.org/10.4172/2157-7544.1000167

Agwu, O. E., Akpabio, J. U., Alabi, S. B., & Dosunmu, A. (2018). Artificial intelligence techniques and their applications in drilling fluid engineering: A review. *Journal of Petroleum Science and Engineering*, *167*(February), 300–315. https://doi.org/10.1016/j.petrol.2018.04.019

Agwu, O. E., Akpabio, J. U., & Dosunmu, A. (2020). Artificial neural network model for predicting the density of oil-based muds in high-temperature, high-pressure wells. *Journal of Petroleum Exploration and Production Technology*, *10*(3), 1081–1095. https://doi.org/10.1007/s13202-019-00802-6

Albukhari, T. M., Beshish, G. K., Abouzbeda, M. M., Madi, A., & others. (2018). Geomechanical Wellbore Stability Analysis for the Reservoir Section in JNC186 Oil Field. *1st International Conference on Advances in Rock Mechanics-TuniRock 2018*.

Aljubran, M. J., Aramco, S., Arabia, S., Hussain,), Albahrani, I., Ramasamy, J., & Magana-Mora, A. (2022). *Drilling Fluid Properties Prediction: A Machine Learning Approach to Automate Laboratory Experiments*. 0–20.

Allawi, R. H. (2023). Chemical and mechanical model to analysis wellbore stability. *Petroleum Science and Technology*, 1–23.

Allawi, R. H., & Al-Jawad, M. S. (2021). Wellbore instability management using geomechanical modeling and wellbore stability analysis for Zubair shale formation in Southern Iraq.

Journal of Petroleum Exploration and Production Technology, *11*(11), 4047–4062. https://doi.org/10.1007/s13202-021-01279-y

Ayoub, D., Masoud, C. S., & Anthony, W. D. (2019). Wellbore stability analysis to determine the safe mud weight window for sandstone layers. *Petroleum Exploration and Development*, *46*(5), 1031–1038. https://doi.org/10.1016/S1876-3804(19)60260-0

Beheshtian, S., Rajabi, M., Davoodi, S., Wood, D. A., Ghorbani, H., Mohamadian, N., Alvar, M. A., & Band, S. S. (2022). Robust computational approach to determine the safe mud weight window using well-log data from a large gas reservoir. *Marine and Petroleum Geology*, *142*(May), 105772. https://doi.org/10.1016/j.marpetgeo.2022.105772

Chen, L., Fissha, Y., Hasanipanah, M., Ghodhbani, R., Dehghani, H., & Khatti, J. (2025). Accurate Prediction of Blast-Induced Ground Vibration Intensity Using Optimized Machine Learning Models. Defence Technology.

Caenn, R., Darley, H. C. H., & Gray, G. R. (2011). *Composition and properties of drilling and completion fluids*. Gulf professional publishing.

Deka, B. (2023). Chapter-9 drilling fluids'functions. *Theory of drilling fluids*, 23.

Dorigo, M., & Blum, C. (2005). Ant colony optimization theory: A survey. Theoretical Computer Science, 344(2–3), 243–278. https://doi.org/10.1016/j.tcs.2005.05.020

DRILLING FLUID. (n.d.). Retrieved December 17, 2023, from viking-intl.com/Viking-intl/index.php/drilling-fluid/

Ding, X., Hasanipanah, M., & Rezaei, M. (2025). Assessment of mechanical properties of rock using deep learning approaches. Measurement, 250, 117180.

Ding, X., Hasanipanah, M., Rouhani, M. M., & Nguyen, T. (2025). Hybrid catboost models optimized with metaheuristics for predicting shear strength in rock joints. Bulletin of Engineering Geology and the Environment, 84(3), 150.

Ding, Xiaohua, Maryam Amiri, and Mahdi Hasanipanah. "Enhancing shear strength predictions of rocks using a hierarchical ensemble model." Scientific Reports 14, no. 1 (2024): 20268.

Ebadati, N., & Najari, M. (2016). Calculating the drilling mud weight window and geo mechanical properties of Darian limestone formation in Reshadat oil field. *Journal of Geotechnical Geology*, *12*(1), 39.

Epelle, E. I., & Gerogiorgis, D. I. (2020). A review of technological advances and open challenges for oil and gas drilling systems engineering. AIChE Journal, 66(4), e16842.

Freij-Ayoub, R., Tan, C. P., Choi, S. K., & others. (2003). SPE/IADC 85344 Simulation of Time-Dependent Wellbore Stability in Shales Using A Coupled Mechanical-Thermal-Physico-Chemical Model. *SPE/IADC Middle East Drilling Technology Conference and Exhibition*.

Gholami, R., Raza, A., Rabiei, M., Fakhari, N., Balasubramaniam, P., Rasouli, V., & Nagarajan, R. (2021). An approach to improve wellbore stability in active shale formations using nanomaterials. *Petroleum*, *7*(1), 24–32. https://doi.org/10.1016/j.petlm.2020.01.001

Gowida, A., Ibrahim, A. F., & Elkatatny, S. (2022). A hybrid data-driven solution to facilitate safe mud window prediction. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-20195-7

Hasanipanah, M., & Amnieh, H. B. (2025). Enhancing Ground Vibration Prediction in Mine Blasting: A Committee Machine Intelligent System Optimized with Metaheuristic Algorithms: Hasanipanah and Amnieh. Natural Resources Research, 1-27.

Hasanipanah, M., Jamei, M., Mohammed, A. S., Amar, M. N., Hocine, O., & Khedher, K. M. (2022). Intelligent prediction of rock mass deformation modulus through three optimized cascaded forward neural network models. Earth Science Informatics, 15(3), 1659-1669.

Hasanipanah, M., Monjezi, M., Shahnazar, A., Armaghani, D. J., & Farazmand, A. (2015). Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement, 75, 289-297.

Harifi, S., Khalilian, M., Mohammadzadeh, J., & Ebrahimnejad, S. (2019). Emperor Penguins Colony: a new metaheuristic algorithm for optimization. Evolutionary Intelligence, 12(2), 211–226. https://doi.org/10.1007/s12065-019-00212-x

Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization (Vol. 200). Technical report-tr06, Erciyes University, Computer Engineering Department.

Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. Applied soft computing, 8 (1), 687–697. https://doi.org/10.1016/j.asoc.2007.05.007

Mozie, K. N. (n.d.). *Characterization of Ultrasonic Waves in Various Drilling Fluids*.

Nunez, Y. J., Sameer, M., Ruiz, F., Al Mutawa, A. A., Al Shamisi, E. D., Hamdy, I., Al Hendi, M., Al Dhaheri, K. H., & Torres, J. (2021). Deviated Drilling Through Salt Dome and Horizontalization Across Extremely Heterogeneous Formations: A Case Study in North Abu Dhabi. In *SPE/IADC Middle East Drilling Technology Conference and Exhibition*. https://doi.org/10.2118/202149-MS

Olson, J. E., Srinivasan, S., Zhang, J., & Chenevert, M. E. (2005). The impact of shale properties on wellbore stability. *Faculty of the Graduate School*, *Doctor of*.

Pao, W., Faris, M., Ali, B., & Pao, W. K. (2016). *Mud Weight Prediction for Offshore Drilling*. *December*.

Phan, D. T., Liu, C., AlTammar, M. J., Han, Y., & Abousleiman, Y. N. (2022). Application of Artificial Intelligence To Predict Time-Dependent Mud-Weight Windows in Real Time. *SPE Journal*, *27*(1), 39–59. https://doi.org/10.2118/206748-PA

Reuters. (2013). *Iran says oil well blowout controlled, no injuries*. https://www.reuters.com/article/uk-iran-blowout-idUKBRE93G09S20130417/

Sabah, M., Mehrad, M., Ashrafi, S. B., Wood, D. A., & Fathi, S. (2021). Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *Journal of Petroleum Science and Engineering*, *198*, 108125.

Soroush, Dr. H. (2020). *Field Development Geomechanics* [Video recording]. *Technology Roundtable: Drilling Fluids*. (2017, October).

Tohidi, A., Fahimifar, A., & Rasouli, V. (n.d.). *The effect of Non-Darcy Flow on Induced Stresses Around a Wellbore in an anisotropic in-situ stress field*.

Tohidi, A., Fahimifar, A., & Rasouli, V. (2017). Analytical Solution to Study Depletion/Injection Rate on Induced Wellbore Stresses in an Anisotropic Stress Field. *Geotechnical and Geological Engineering*. https://doi.org/10.1007/s10706-017-0429-z

Tatar, A., Naseri, S., Bahadori, M., Rozyn, J., Lee, M., Kashiwao, T., & Bahadori, A. (2015). Evaluation of different artificial intelligent models to predict reservoir formation water density. Petroleum Science and Technology, 33(20), 1749-1756.

Zahiri, J., Abdideh, M., & Golab, E. G. (2018). Determination of safe mud weight window based on well logging data using Determination of safe mud weight window based on well logging data using artificial intelligence Javad Zahiri, Mohammad Abdideh & Elias Ghaleh Golab. *Geosystem Engineering*, *00*(00), 1–13. https://doi.org/10.1080/12269328.2018.1504697

SAŽETAK

Predviđanje težine i vrste isplake metodom strojnoga učenja

Odabir optimalnoga fluida za bušenje, definiranoga njegovom težinom i kemijskom vrstom, ključan je za sprječavanje nestabilnosti kanala bušotine i skupih katastrofalnih nesreća. Tradicionalne metode često se oslanjaju na metodu pokušaja i pogreške, prošla iskustva ili pojednostavnjene modele koji ne uspijevaju opisati složene interakcije stijena i fluida. Iako obrada velikoga broja podataka nudi obećavajuću alternativu, nedostaju istraživanja o istodobnome predviđanju težine i vrste isplake. Ovim istraživanjem predstavlja se novi okvir temeljem na strojnome učenju koji istovremeno predviđa ove važne faktore. Koristeći se sveobuhvatnim skupom podataka dobivenim iz 50 godina dnevnih izvješća o bušenju na 20 naftnih bušotina, napravljena su i uspoređena tri algoritma inspirirana prirodom: optimizacija *Ant Colony* (ACO), *Bee Colony* (BCO) i *Emperor Penguins Colony* (EPC). Rezultati pokazuju visoku prediktivnu točnost svih modela, pri čemu se algoritam optimizacije *Bee Colony* (BCO) pokazao kao najprecizinji, dajući koeficijent korelacije (R2) od 0,9841 i srednju kvadratnu pogrešku (engl. *root-mean-square error*, RMSE) od 0,0245. Nadalje, analiza osjetljivosti pokazala je da je na svojstva isplake mehanička brzina bušenja (ROP) najutjecajniji parametar, nadmašujući ostale varijable bušenja. Ključni je rezultat istraživanja dosljedno podudaranje modela sa 79–87 % pouzdanosti, da je isplaka na bazi morske vode s dodatkom polimera i *soltex* aditivima (SW-PO-SX) idealan bušaći fluid za pročavano polje. Ovo istraživanje pruža robusno rješenje temeljeno na podatcima koje omogućuje sustavan i proaktivan pristup odabiru bušaćega fluida znatno povećavajući sigurnost i učinkovitost rada.

Ključne riječi:
grupiranje podataka, proaktivni model, težina isplake, vrsta isplake, stabilnost kanala bušotine

Author's contribution

Amin Tohidi (Assistance Professor): conceptualization, investigation, original draft and writing. **Alireza Afradi** (PhD): software and Data Analysises, review & editing.
All authors have read and agreed to the published version of the manuscript.