Journal of Petroleum Exploration and Production Technology
https://doi.org/10.1007/s13202-024-01769-9

ORIGINAL PAPER-PRODUCTION ENGINEERING

Machine learning classification approaches to optimize ROP and TOB using drilling and geomechanical parameters in a carbonate reservoir

Mohammad Reza Delavar[1] · **Ahmad Ramezanzadeh**[1]

Received: 21 September 2022 / Accepted: 12 February 2024

Abstract

Drilling optimization has been broadly developed in terms of influential parameters. The assessment time and the effects of both geomechanical and drilling parameters were vital challenges of investigations. Drilling factors are applied force or rotation of drilling agents such as weight on bit (WOB), and geomechanical features represent mechanical indexes of rocks including unconfined compressive strength (UCS). Optimization efforts have been demonstrated on complex prediction methods whereas the simplicity of classification can offer some optimal ranges utilizing machine learning classifications in an accelerated process. In this study, a novel procedure using the supervised and semi-supervised learning approaches was conducted to classify and optimize the rate of penetration (ROP) and torque on bit (TOB). Firstly, in the case well, user-defined classes were assigned based on geomechanical units (GMU) and the ranges of high ROP and low TOB, thus classes divided drilling factors as GMUs of the case. Secondly, the feature selection was carried out by neural pattern recognition with three multi-objective optimization methods for classification. The inputs of classifications were WOB, hook load, pump pressure, flow rate, UCS, and internal friction angle. Classification approaches were decision trees, support vector machine (SVM), and ensemble learning. Finally, the bagged trees permutation and Laplacian SVM (LapSVM) algorithm separately revealed the significance of parameters and predicted the optimal ROP and TOB regions. Findings showed (1) in supervised classification of the case well, the cubic SVM and bagged trees had the highest area under the curve (AUC) and accuracy, on average 0.97 and 0.96, respectively. (2) The average accuracy of the supervised classifications in a test well was 91% except for the fine SVM, which makes them reliable for the fields with the least information. (3) The permutation outcomes for significant features, flow rate and UCS, exposed influential parameters for ROP and TOB optimization. (4) The semi-supervised method, LapSVM, not only acquired both ROP and TOB labels with an accuracy of 88% but also presented their optimal ranges in 95% of the assessed zones. (5) LapSVM deals with a limited training section perfectly opposed to the supervised version, which is vital for drilling investigation. (6) Implementing machine learning classification approaches with rock properties is a key factor in achieving effective drilling parameters in less time. More importantly, the recommended drilling factors concerning geomechanical properties can ameliorate both drilling performance and perception of upcoming collapse.

Keywords Drilling optimization · Drilling parameters classification · Geomechanical parameters · Semi-supervised method · Hybrid feature selection · Ensemble learning

Abbreviations

AUC	Area under receiver operating characteristics curve
Bagged.T	Bagged trees method
C.SVM	Cubic support vector machine
DTC	Decision trees classifier
F.Tree	Fine decision trees classifier
GMU	Geomechanical units
GWO	Gray wolf optimizer
NPR	Neural pattern recognition
NSGA-II	Second version of the non-dominated sorting genetic algorithm
PDC	Polycrystalline diamond compact
PDM	Positive displacement motor
PSO	Particle swarm optimization
RBF	Radial basis function

✉ Mohammad Reza Delavar
m.rdelavar@yahoo.com

[1] Department of Mining, Petroleum and Geophysics Engineering, Shahrood University of Technology, Shahrood, Iran

Published online: 19 March 2024

RUSB.T	RUSBoosted trees
SVM	Support vector machine

Latin letters

CALI (inches)	Caliper log
CCS (MPa)	Confined compressive strength
CCS_{DP} (MPa)	Confined compressive strength based on differential pressure
CCS_{sk} (MPa)	Confined compressive strength based on Skempton pore pressure
DP (MPa)	Differential pressure (or confining stress)
DT (μs/ft)	Sonic interval transmit time
DTC_{Best} (–)	Best of decision trees classifier
E and EE (–)	Misclassification error and total error
E_{sta} (GPa)	Static Young's modulus
Flow.R (Gal/min)	Flow rate
GR and CGR (API)	Gamma-ray and corrected gamma-ray
HKL (klbf)	Hook load
IFA (degree)	Internal friction angle
MSE (MPa)	Mechanical specific energy
n_f (–)	Number of selected features
NPHI (V/V)	Neutron porosity
PEF (–)	Formation evaluation photoelectric factor
PHIE (V/V)	Effective porosity
Pump.P (psi)	Pump pressure
q and b (–)	Exponent of cubic and quadratic polynomial kernel functions in SVM and the bias value in SVM function
RHOB (kg/m^3)	Density log
ROP (m/h)	Rate of penetration
RPM (rpm)	Rotary speed or bit revolutions per minute
RT (ohm m)	Resistivity log
$RR_{i,j}$ (–)	Total number of class i values
$r(n)$ and $P(n)$ (–)	Re-substitution estimation error for misclassification and probability of any value equal with criteria
TOB (Lbf.ft)	Torque on bit
UCS (MPa)	Unconfined compressive strength
V_p (m/s)	Compressional wave velocity
V_s (m/s)	Shear wave velocity
V_{shale} (–)	Shale volume
WOB (klbf)	Weight on bit
W_{train} and W_{test}	Weights of the training and test
X_{nor}, X_{max}, and X_{min} (–)	Normal, maximum, and minimum values

Greek letters

α, β, δ, and ω (–)	Alpha, beta, delta, and omega in MOGWO method
ρ (kg/m^3)	Density
$\varnothing(x)$ (–)	Mapping of x from predictor samples space (Rn) in SVM
σ^2 (–)	The variance of RBF or Gaussian kernel (kernel scale)

Introduction

The drilling costs are a significant portion of the expenditures in oil and gas projects. In recent years, attention has been given to developing analytical and data-driven models that could estimate and optimize important factors efficiently. The rate of penetration (ROP) that is obtained by recording the drilled depth per unit of time, is an effective feature for drilling optimization, and a higher rate of it can result in the productive rig, detection of possible kicks, and indicating stick–slip (Elkatatny 2019; Hegde et al. 2017). In the process of drilling optimization, torque on bit (TOB) is another critical factor, which is directly related to the amount of applied pressure of the drilling agent (Motahhari et al. 2009). In addition, TOB is a vital part of the prediction of the specific energy (SE) proposed by Teale (1965). According to studies, some efforts have been made to model ROP and TOB, a sort of the results will be briefly explained.

Rate of penetration models

There are two categories of ROP predictive models, including the physics-based and data-driven models. The mathematical formulas are used based on the physics-based or the training of data-driven models via algorithms to obtain ROP. As regards the physics-based model, an ROP model was proposed by Maurer (1962) (namely perfect cleaning) defined a drilling mechanism consisting of the weight-on-bit (WOB), the rotational speed (RPM), TOB, and rock strength. There are studies focused on a cost-effective ROP by gaining the proper WOB and RPM (Galle and Woods 1963). In terms of data-driven models by artificial neural networks (ANN), Bilgesu et al. (1997) played a pioneering role and introduced a methodology to predict ROP. Concerning the usage of artificial intelligence (AI), numerous useful studies were presented with high accuracy (Bezminabadi et al. 2017; Elkatatny 2019). In this way, Table 1 expresses some proposed both physics-based and data-driven studies in terms of ROP prediction and optimization.

In addition, not all of the ROP-affecting parameters have been utilized in previous models due to the complex relationship between ROP and factors. Besides, geomechanical studies have introduced the influence of specific

Table 1 The details of some studies proposed for the rate of penetration (ROP) prediction

Authors	Model	Involved parameters	Results
Warren (1987)	Physics-based	Drill-ability evaluation by drilling factors and pressure	Establishing ROP model with new mechanical properties
Rampersad et al. (1994)	Physics-based	ROP estimation and optimization via the geological drilling log	Improvement of ROP optimization cause of the innovated log
Shirkavand et al. (2009)	Physics-based	ROP prediction utilizing the confined compressive strength	Introducing ROP in underbalanced drilling condition
Hareland and Hoberock (1993)	Physics-based	ROP model exerting tools' features	Representing for polycrystalline diamond compact (PDC) bits
Motahhari et al. (2009)	Physics-based	Optimization of ROP model by pressure and hole condition	Effective model via positive displacement motor (PDM)
Bourgoyne Jr and Young Jr (1974)	Data-driven	Implementing eight factors for drill-ability examination	ROP optimization and representing a multiple regression procedure
Al-Betairi et al. (1988)	Data-driven	Optimization utilizing factors such as weigh on bit (WOB)	Proposed regression-based model for ROP optimization
Al-AbdulJabbar et al. (2019)	Data-driven	ROP estimation using the mud properties and fluid factors	ROP model via mud features and clustering approach
Alali et al. (2021)	Data-driven	Hybrid model and optimization of ROP by dynamic drilling factors	Minimizing the non-productive time and improving efficiency

properties of rocks such as the internal friction angle (IFA) and confined compressive strength (CCS) for drilling operations. Implementing vast geomechanical and drilling data, Mehrad et al. (2020) proposed an accurate prediction ROP model by a hybrid machine learning approach. Delavar et al. (2021) introduced an ROP model using a wide range of both the drilling factors and geomechanical properties exerting the ANN and Bayesian approaches. As a result, some procedures were presented for finding the optimum ROP in future drilling operations.

Torque on bit models

The introduced models for TOB have relied on the type of recorded data and the development of measurement tools. The cutting-edge downhole TOB recording tools were an effective invention in advance, Warren (1984) presented a TOB model considering SE in drilling along with using some parameters, including WOB, RPM, ROP, and bit size. The torque and drag (T&D) models, in the chain of TOB studies, were presented to describe the mechanism of torque in drilling. To this extent, a T&D real-time model was proposed by Shahri et al. (2018), which reduced costs and improved drilling performance. Moreover, Hegde and Gray (2018), Oyedere and Gray (2020b), and Delavar et al. (2023) used machine learning including the random forest and support vector machine (SVM) with metaheuristic optimization methods for the TOB models.

Classification benefits and flowchart

Data-driven models of ROP and TOB have been used widely to improve physics-based models or introduce machine learning regression models via the wells data. The impact of influential parameters on ROP and TOB is very complex thus it is necessary to consider various features on the basis of the case study. Classification methods can deal with the mentioned complicated circumstance because their practical predictive models approximate the mapping function from input variables to desired ROP and TOB. Furthermore, the importance of representing classifications lies in the simplicity of predicting ROP and TOB in the evaluation of the drilling operation performance.

Hybrid classification approaches are accurate and feasible manners to determine challenges related to drilling and reservoir engineering (Delavar 2021; Sanei et al. 2023). Reservoir characteristics such as fracture density were classified by machine learning procedures with high accuracy (Li et al. 2018). Using classification, Oyedere and Gray (2020a) proposed ROP and TOB classification models to classify two classes so-called high ROP and low TOB using WOB, RPM, flow rate, and unconfined compressive strength (UCS). Here, the classes were allocated by geomechanical and drilling factors, and more practical drilling factors were determined for hybrid classification approaches.

Figure 1 illustrates the steps to achieve the goals of this effort. In the flowchart, first, the geomechanical units (GMU)

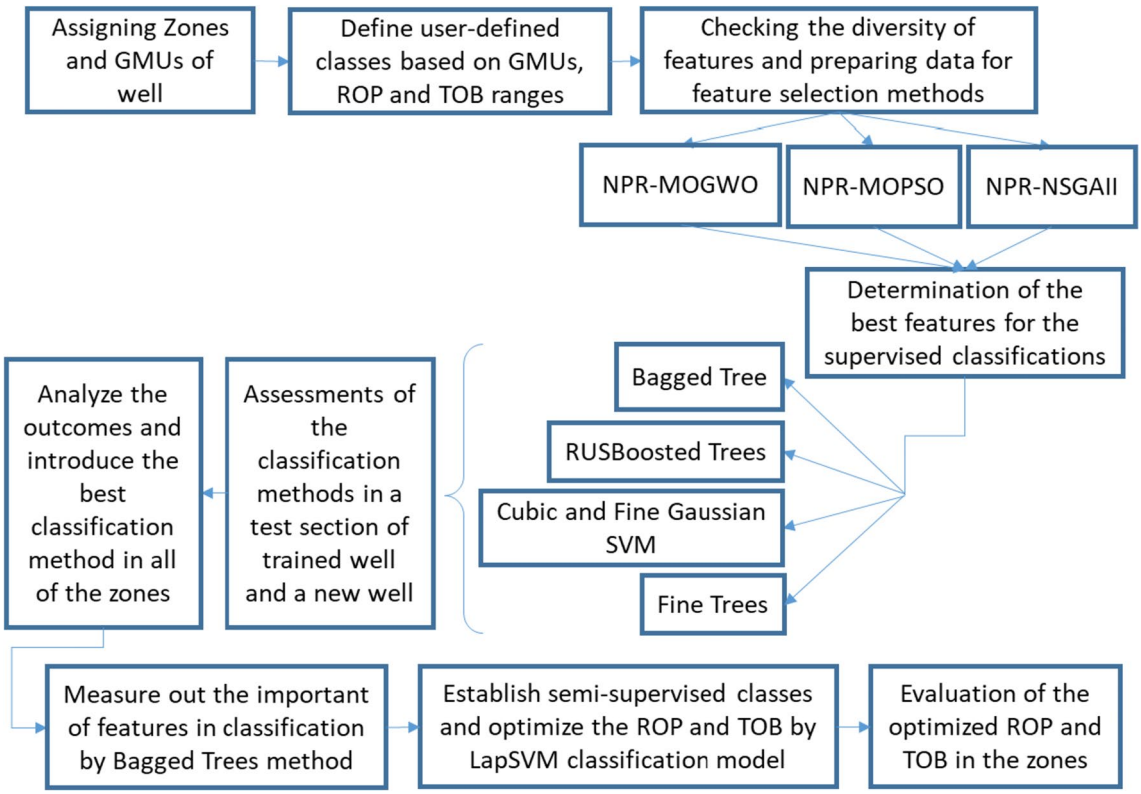

Fig. 1 The flow-chart of all the process of analysis in the present study

were established, and they formed the user-defined classes in addition to ROP and TOB bands. Second, in the feature selection, neural pattern recognition (NPR) was combined with three multi-objective optimization methods, thus the features were selected for the classification methods. Next, five supervised classification methods were applied to the dataset, which included the features and labels. As a result, the best classification method was introduced for the zones. Third, the important features for the classification of ROP and TOB in the zones were presented. Finally, based on a semi-supervised method using the revealed significant factors, the optimized drilling parameters were introduced for ROP and TOB.

Geology and data acquisition

The studied well is located in the southwest of Iran in the Dezful embayment. It was drilled in the Marun oilfield situated in the Middle East (Telmadarreie et al. 2012). In Marun, the Asmari reservoir was considered, which is a carbonate reservoir with five zones and four sub-zones. Here, the interval depth of the case well is 3557 to 3924 m, which includes 4 zones and 3 sub-zones. The lithology of the Asmari generally includes limestone, dolomite-limestone, and thin layers of both sandstone and Shale-Marne stone interbedded (Alavi 2004).

For feature selection methods, the conventional logs of the well were corrected gamma-ray (CGR), DT, NPHI, density log (RHOB), formation evaluation photoelectric factor (PEF), and resistivity log (RT). On the other hand, the mud logs and drilling measurements acquired ROP, TOB, WOB, hook-load (HKL), pump pressure (Pump.P), and flow rate (Flow.R). The dataset of the present study obtained from a wellbore included 2952 points as samples and 738 points of ROP and TOB. The petrophysical and drilling logs of the present study are shown in Fig. 2. As regards the lithology of the well (track 2 in Fig. 2), conventional logs such as CGR and RT, illustrated in tracks 3 and 4, are changed through the diversity of rocks. Concerning the drilling parameters in the dataset, the recorded mud logs are shown on tracks 5 and 6.

In addition, despite the fact that a section of the training well was divided to prepare the test part, the dataset of a second well was deemed to verify the performance of classification methods in the vicinity zones. The formations of the second well are carbonate rocks although the distinguishes between the patterns of geomechanical and drilling factors of the training well with the same parameters

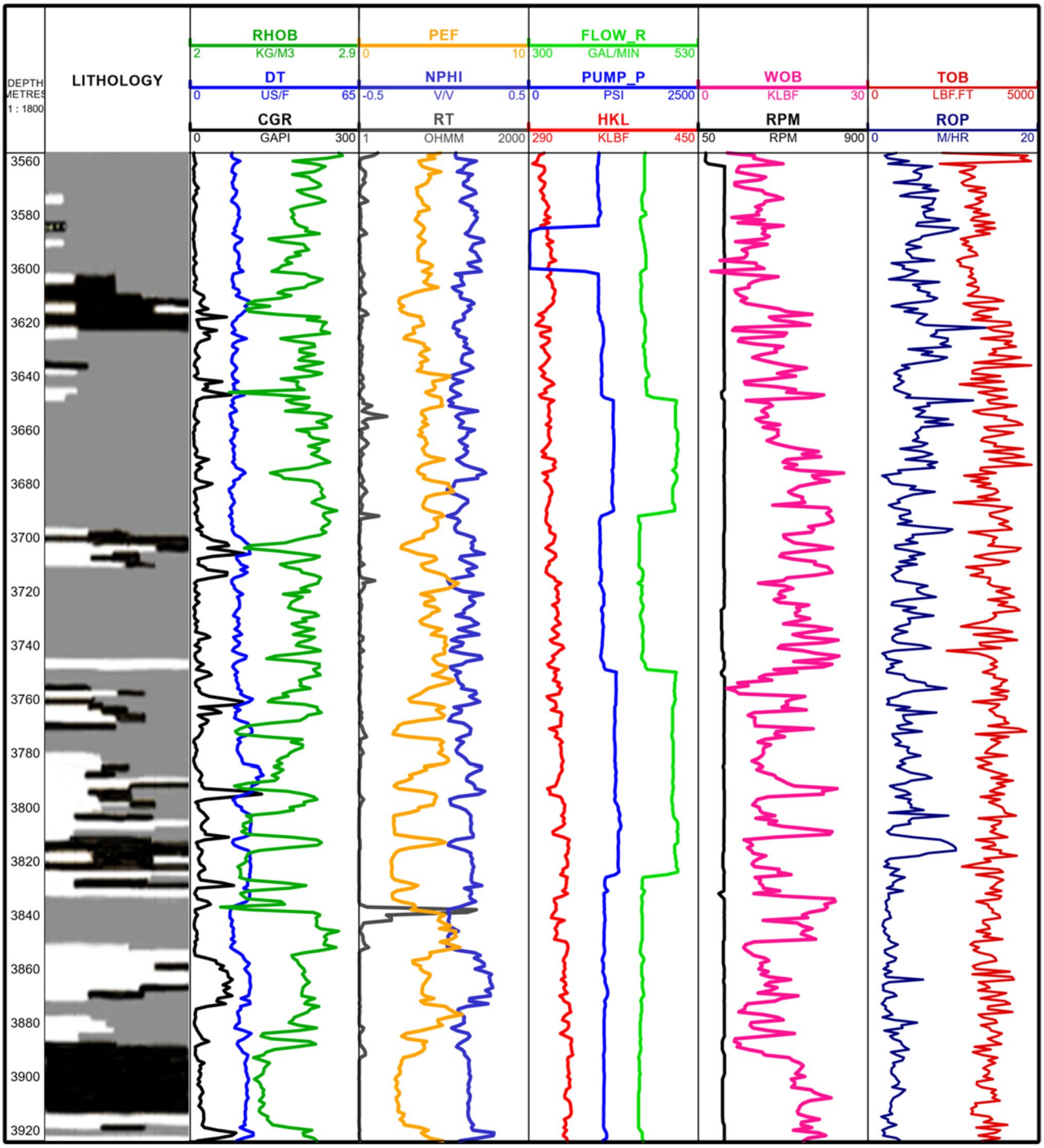

Fig. 2 The conventional and drilling logs of studied well; tracks 1 and 2 are depth and lithology also, petrophysical logs illustrate in tracks 3 and 4; the drilling data show in tracks 5 to 7

in the second well can reveal the capability of models. The interval depth of the second well is 257 m, which provides 2313 points for the evaluation of the proposed method.

The strength parameters of rocks (UCS and CCS) usually have opposite trends toward ROP while pressure and depth of drilling could impact their correlation. On the other hand, IFA is associated with both porosity and shale volume, which could impact TOB and WOB (Anemangely et al. 2018; Calhoun and Ewy 2005). Concerning the influential parameters for ROP and TOB classification and prediction, Table 2 shows

Table 2 The influential parameters used in the previous studies of the rate of penetration and the torque on bit

Applied features	References
WOB	Abbas et al. (2018), Oyedere and Gray (2020a), Oyedere and Gray (2020b)
RPM	Bezminabadi et al. (2017), Oyedere and Gray (2020b)
UCS	Elkatatny (2019), Delavar et al. (2023)
Flow.R	Oyedere and Gray (2020a), Hegde and Gray (2018)
IFA	Delavar et al. (2021), Bezminabadi et al. (2017)
Pump.P	Bourgoyne Jr and Young Jr (1974)
CCS	Delavar et al. (2021), Shirkavand et al. (2009)
HKL	Han et al. (2019), Delavar et al. (2023)

some features executed in the previous studies. To describe the diversity and density of selected petrophysical logs, probability density function (PDF) graphs of the recorded data are illustrated in "Appendix A". The diversity of RHOB, DT, PEF, and NPHI was rather than CGR and RT. In fact, the normal shape of a dataset, spread over in wide range with a peak in the middle, has an influence on results effectively. This checking process proved the credibility of the inputs in terms of statistical features for the classification methods.

In this effort, UCS, CCS, and IFA were obtained by the correlations that have been introduced from the carbonate reservoir case. The dynamic Young's modulus (E_{dyn}) was calculated (Eq. 1) through conventional logs and converted to the static Young's modulus (E_{sta}) (Eq. 2) (Afsari et al. 2010; Zoback 2007). The UCS test on the rock of adjacent wells showed the ranges of UCS at 50–70 MPa. Also, Eq. 3 was presented for the Marun oilfield to predict the UCS of rocks that had been used here (Anemangely et al. 2018). The IFA and CCS were obtained by laboratory-based equation (Eq. 4) and the bottom-hole rock strength and pressure function (Eq. 8) (Calhoun and Ewy 2005; Plumb 1994). The required functions for the calculation of IFA and CCS are (Eqs. 5–7) (Gholami et al. 2014; Kidambi and Kumar 2016).

$$E_{dyn} = \rho V_s^2 \left(\frac{3V_p^2 - 4V_s^2}{V_p^2 - V_s^2} \right) \tag{1}$$

$$E_{sta} = 0.4145 E_{dyn} - 1.0593 \tag{2}$$

$$\mathrm{UCS} = 2.27 E_{sta} + 4.74 \tag{3}$$

$$\mathrm{IFA} = 26.5 - 37.4\left(1 - \mathrm{NPHI} - V_{shale}\right) + 62.1\left(1 - \mathrm{NPHI} - V_{shale}\right)^2 \tag{4}$$

$$V_{shale} = \frac{\mathrm{GR} - \mathrm{GR}_{min}}{\mathrm{GR}_{max} - \mathrm{GR}_{min}} \tag{5}$$

$$\mathrm{CCS}_{DP} = \mathrm{UCS} + \mathrm{DP} + 2\mathrm{DP}\left(\frac{\mathrm{Sin\,IFA}}{1 - \mathrm{Sin\,IFA}} \right) \tag{6}$$

$$\mathrm{CCS}_{SK} = \mathrm{UCS} + \mathrm{DP}_{SK} + 2\mathrm{DP}_{SK}\left(\frac{\mathrm{Sin\,IFA}}{1 - \mathrm{Sin\,IFA}} \right) \tag{7}$$

$$\mathrm{CCS} = \frac{\mathrm{CCS}_{DP}(\mathrm{PHIE} - 0.05)}{0.15} + \frac{\mathrm{CCS}_{SK}(0.2 - \mathrm{PHIE})}{0.15} \tag{8}$$

where E_{dyn}, ρ, V_P, V_S, E_{sta}, NPHI, V_{shale}, GR_{max}, and GR_{min} represent dynamic elastic modulus, density, compressional wave velocity, shear wave velocity, static elastic modulus, neutron porosity, shale volume, maximum gamma ray, and minimum gamma ray. Also, DP, CCS_{DP}, CCS_{SK}, DP_{SK}, and PHIE show differential pressure or confining stress, CCS based on DP, CCS based on Skempton pore pressure, DP based on Skempton pore pressure, and effective porosity or porosity index.

Methodology

The drilling and geomechanical factors of the well were acquired in the first step of the study. The methods used for user-defined labeling, classification, feature selection, and optimization will be briefly described.

User-defined classes

Constructing the classes was the initial step of this study, Table 3 expresses the scheme of the labeling process using geomechanical and drilling factors. Firstly, the thresholds of zones were defined based on previous studies on carbonate formations in the region (Alavi 2004), and considering the well logs. They have evident boundaries if the conventional well logs were in access. Secondly, in assigning the labels of the studied well, the geomechanical units (GMU) were specified. Basically, GMU has close geomechanical features; thus the impact of drilling tools on them is defined in a special pattern.

The zones were divided into 21 GMUs, which GMUs 10 to 21 are shown in Fig. 3 and others are available in "Appendix B". Figure 3 illustrates depth, zones, and GMUs on tracks 1 to 3. Besides, tracks 4 to 7 show the well logs and geomechanical parameters which are a part of the dataset of the well. The GMU's separation criterion in the zones was the average of parameters. For instance, in Z.3 (Fig. 3), the averages of DT (Sonic interval transmit time) and UCS are depicted as yellow dash-lines, which are the trends of common geomechanical features in the GMU. By assigning GMUs, the labeling process proceeded to step 3 of Table 3. In this way, the averages of ROP and TOB in the entire depth were assessed to set their thresholds. Finally, the classes

Table 3 The process of labeling using geomechanical and drilling parameters

Numbers	Labeling steps	Parameters
1	Evaluation of the zones of the carbonate formation	Well logs
2	Dividing the zones to GMU	Geomechanical parameters and well logs
3	Comparing the averages of ROP and TOB of all GMUs in a zone based on thresholds	Drilling parameters
4	Assigning labels as: ROP's thresholds: differences values of > 0.6 m/hr (GMUs with ROP differences ≤ 0.6 m/hr devoted as a class) TOB's thresholds: differences values of > 300 lbf.ft (GMUs with TOB differences ≤ 300 lbf.ft devoted as a class)	Drilling parameters and geomechanical units

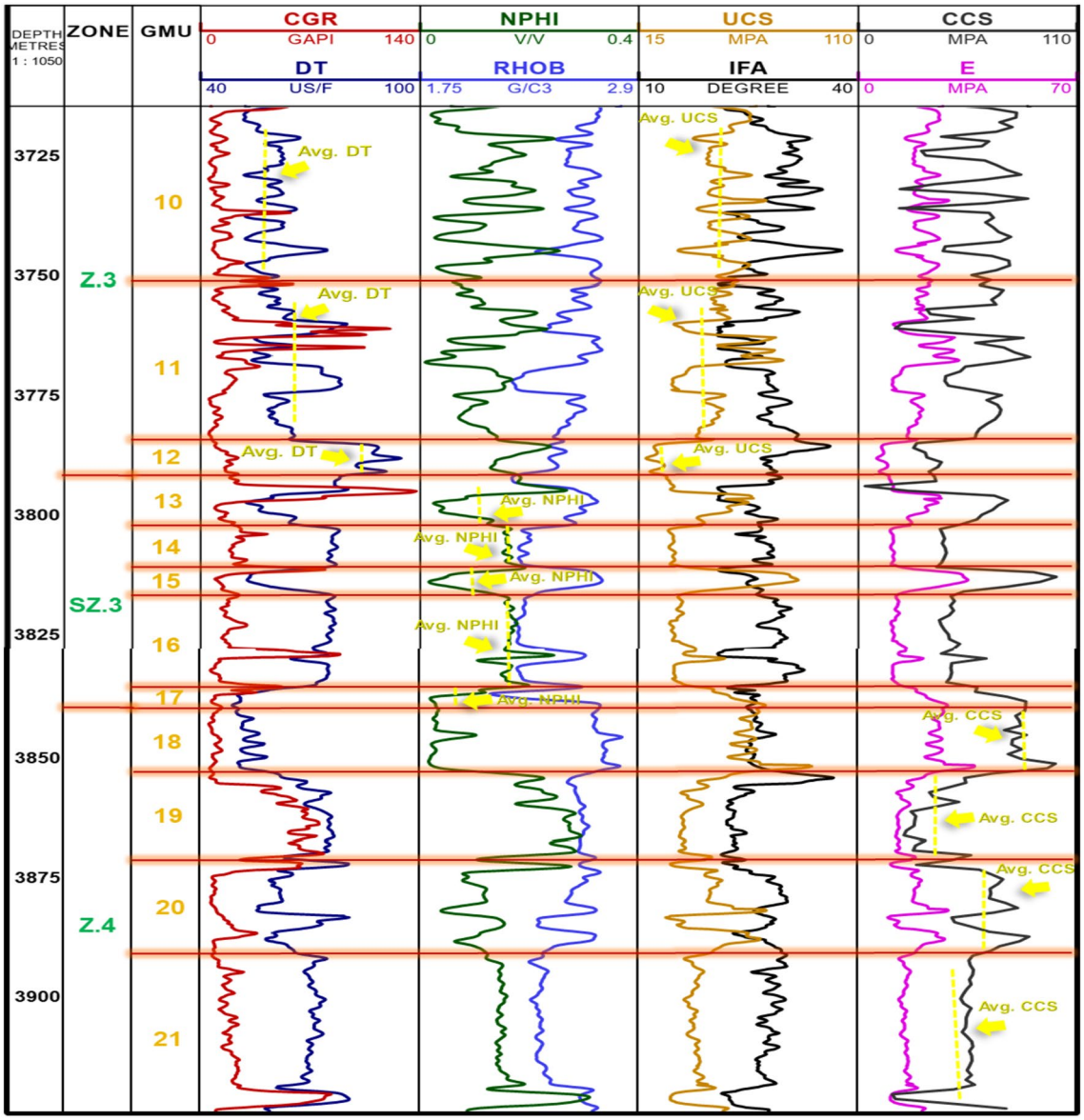

Fig. 3 The conventional well logs and geomechanical parameters of the case in the geomechanical units (GMUs) 10 to 21; tracks 1 to 3 are included depth, zones and GMU, and petrophysical logs are in tracks 4 and 5, and also, tracks 6 and 7 are consisted of the trends of geomechanical features in the well

were achieved by combining the GMUs, dividing the large-scale depth into some particular parts, and evaluating the ROP and TOB bands.

The steps of Table 3 were employed in all of the zones. The changes in the parameters and the average values of ROP and TOB of Z.2 & SZ.2 are respectively illustrated in Figs. 4 and 5, to describe the labeling method of a zone. In the process, first, the zone is divided into specific GMUs (expressed in Fig. 3). Second, in the zone, all GMUs were determined (Fig. 4), and the bands of ROP and TOB were checked via step 4 of Table 3 (Fig. 5). These labels were named for both ROP and TOB classes separately. Therefore, on the ROP side, GMUs 5 and 8 were denoted in class 1 because the differences in ROP in their GMUs were lower than 0.6 m/hr. Where GMU 6 was labeled as class 2 and GMUs 7 and 9 were labeled as class 3 due to staying in the constraint of 0.6 m/hr. For TOB, the classes were named as follows GMUs 5, 8, and 9 in class 1, GMU 6 in class 2, and GMU 7 in class 3 owing to the employed limitations of 300 lbf.ft.

Other zones of the case well were categorized according to the explained procedure and the labels are shown in Table 4. It should be noted that the mentioned labels were used for the supervised classification methods. However, in the semi-supervised method, Laplacian SVM (LapSVM) uses a different scale from the applied manner of the supervised methods (Table 4) for ROP and TOB. Actually, the LapSVM's thresholds were defined as a certain number dividing ROP and TOB to the high and low zones due to

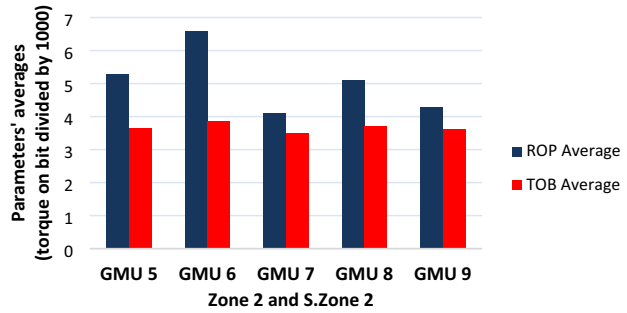

Fig. 5 The average values of the rate of penetration and torque on bit in the five GMUs of the studied well which were used for the user-defined thresholds

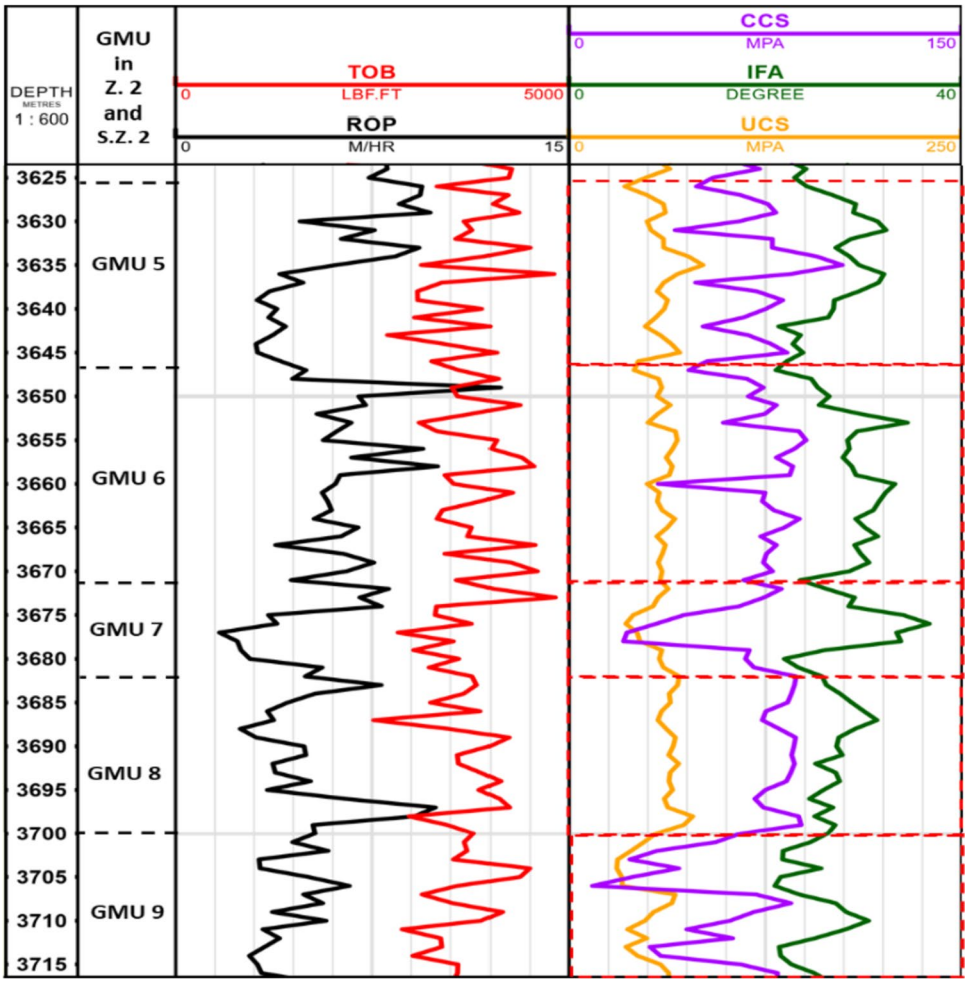

Fig. 4 The drilling and geomechanical logging data in zone 2 and subzone 2 as an example zone of the well and used for user-defined classes; the formulas of the internal friction angle (IFA) and rock strength features can be found in the "Geology and data acquisition" section

Table 4 Details of the zones, GMUs, the user-defined labels of the rate of penetration, and torque on bit of the studied well

Zone	GMU	TOB labels	ROP labels
Z.1	1	1	1
	2	2	2
SZ.1	3	1	1
	4	2	2
Z.2 and S.Z.2	5	1	1
	6	2	2
	7	3	3
	8	1	1
	9	1	3
Z.3	10	1	1
	11	2	2
	12	3	3
SZ.3	13	1	1
	14	2	1
	15	3	2
	16	3	3
	17	4	4
Z.4	18	1	1
	19	2	1
	20	3	2
	21	1	2

achieving optimal parameters. The different limits of ROP and TOB in LapSVM were assigned because the semi-supervised methods can be trained by a few features and also it was aimed at seeking optimal factors.

The significant influences of geomechanical properties in the supervised and semi-supervised labeling procedures were in the process of categorizing zones to GMUs and acquiring the ROP and TOB contributed to geomechanical features, respectively. The second utilization of geomechanics (semi-supervised) leads the method to indicate the optimal ROP and TOB when the important drilling factors (gained via a method) apply as a pattern of drilling factors from GMU with close geomechanical features. This desirable pattern of drilling parameters was obtained in the labeling process of LapSVM. The details of LapSVM's classes are described in the following in its specific section. It should be deemed that presenting a certain amount of drilling factors that offer optimum ROP and TOB is almost impossible due to tools limitations; thus the values were introduced as ranges.

Feature selection

Feature selection can be used in the process of prediction, classification, and pattern recognition. Feature selection methods are often based on the ANN combined with optimization approaches. Here, the multi-objective methods were applied for the optimization aimed at minimizing both the number of features and the errors of the classification (Hamdani et al. 2007). The influence of two features for simultaneous optimization was the main reason to utilize the multi-objective methods.

In this study, the NPR approach was the major learning method for the feature selection. The scaled conjugate gradient (SCG) algorithm was applied to the training process including 70% of the dataset. The SCG algorithm is a fast learning method with good convergence (Castillo et al. 2006; Møller 1993). The remaining 30% of the dataset was assigned to the testing and validation process and 20 neurons in the two hidden layers were used for the ANN's structure in the Matlab software. In the hybrid feature selection methods and classification methods of the present study, training, validation, and testing sections were randomly split because it ensures that the sections are representative of the original dataset. The data normalization step was achieved in Eq. 9, and the cost function of NPR (Z) is shown in Eq. 9. The weights (W_{train} and W_{test}) were assigned for two types of error consisting of 0.8 and 0.2 for the train and test sections, respectively. The main function is shown in Eq. 10. Knowing that NPR hybridized by the optimization method at each run, different error values (EE), and hence the average value (E_{Total}) were considered (Eq. 11).

$$X_{\text{nor}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}; \quad \left\{ \begin{array}{l} \min Z_1 = E \\ \min Z_2 = n_f \end{array} \right\} \Rightarrow \quad Z = \begin{bmatrix} E & n_f \end{bmatrix} \tag{9}$$

$$EE = W_{\text{train}} \times E_{\text{train}} + W_{\text{test}} \times E_{\text{test}} \tag{10}$$

$$E_{\text{Total}} = \text{mean}\ (EE); \tag{11}$$

where X_{nor}, $X_{\max}$, $X_{\min}$, W_{train}, E_{train}, W_{test}, E_{test}, E, and n_f represent normal value, maximum value, minimum value, weight of train, error of train, weight of test, error of test, difference of observed and predicted values by NPR, and number of features.

Multi-objective genetic optimization

The non-dominated sorting genetic algorithm II (NSGA-II) is a multi-objective optimization method. The NSGA-II finds the optimal parameters with minimum errors and presents the results as a Pareto-optimal diagram. In the case of complex datasets, this natural-inspired algorithm is useful due to its lower computational costs (Liu et al. 2015). Moreover, the NSGA-II presents the distribution parameters by a novel crowding distance method (Deb et al. 2002; Delavar and Ramezanzadeh 2023).

The flowchart of the NSGA-II algorithm is shown in Fig. 6. First, the problem definition, cost function, and variables limitation were set. Next, the NSGA-II parameters,

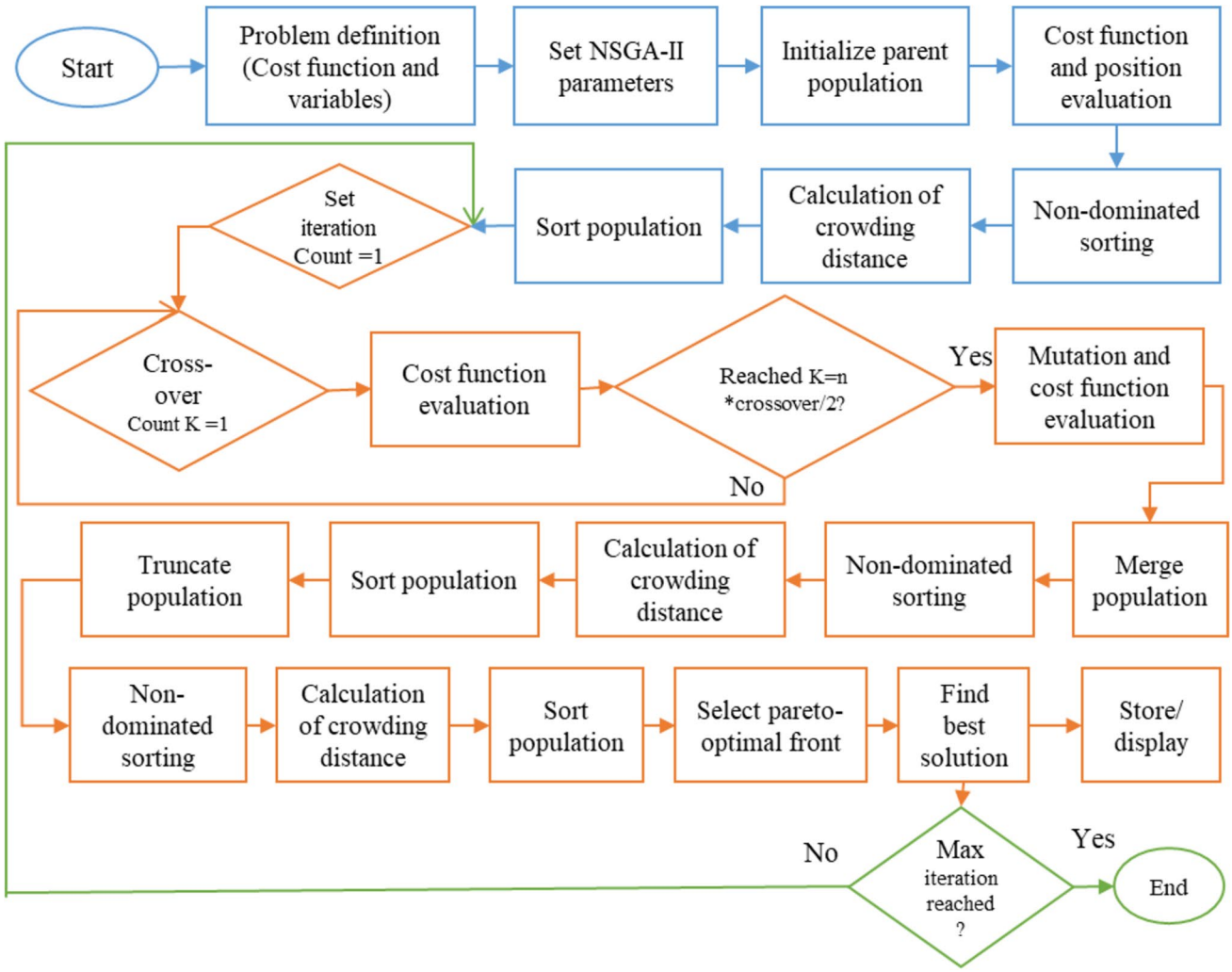

Fig. 6 The flowchart of the non-dominated sorting genetic algorithm II method, adapted from (Deb et al. 2002)

including mutation percentage, crossover percentage, and mutation rate were set as 0.4, 0.7, and 0.1, respectively. There were some steps for the optimization process of the swarm such as non-dominated sorting. Finally, the main loop of NSGA-II was started by determining the calculations in every mutation to reach the best solution.

Multi-objective particle swarm optimization

The multi-objective particle swarm optimization (MOPSO) algorithm is based on heuristic particle swarm optimization (PSO). In MOPSO, there is a population of particles, a repository as external memory, and the definition of the new generation with space of objective. In fact, in MOPSO, the solution of storing each particle is employed instead of the elitism procedure that had been used in NSGA-II (Coello and Lechuga 2002). Figure 7 shows the process of MOPSO where the number of grids per dimension and the inflation rate were set at 7 and 0.1, respectively. Besides, both the leader selection pressure and deletion selection pressure were equal to 2. The mutation rate was also adjusted to 0.1. In the steps of the MOPSO, computations reach the main loop, and the solutions are presented (Coello 2011).

Multi-objective Gray wolf optimization

The multi-objective Gray wolf optimizer (MOGWO) presents the hunting behavior of Gray wolves. The main procedure of MOGWO was adapted from the proposed Gray wolf optimizer (GWO) by adding some features. In the algorithm, the alpha (α), beta (β), and delta (δ) are search agents and are updated by the leader (Mirjalili et al. 2016). Figure 8 shows the performance of MOGWO in different steps. First, the cost function was defined; then, three specific parameters of the algorithm, including grid inflation parameters, archive size, and the number of grids were set as 0.1, 100, and 10,

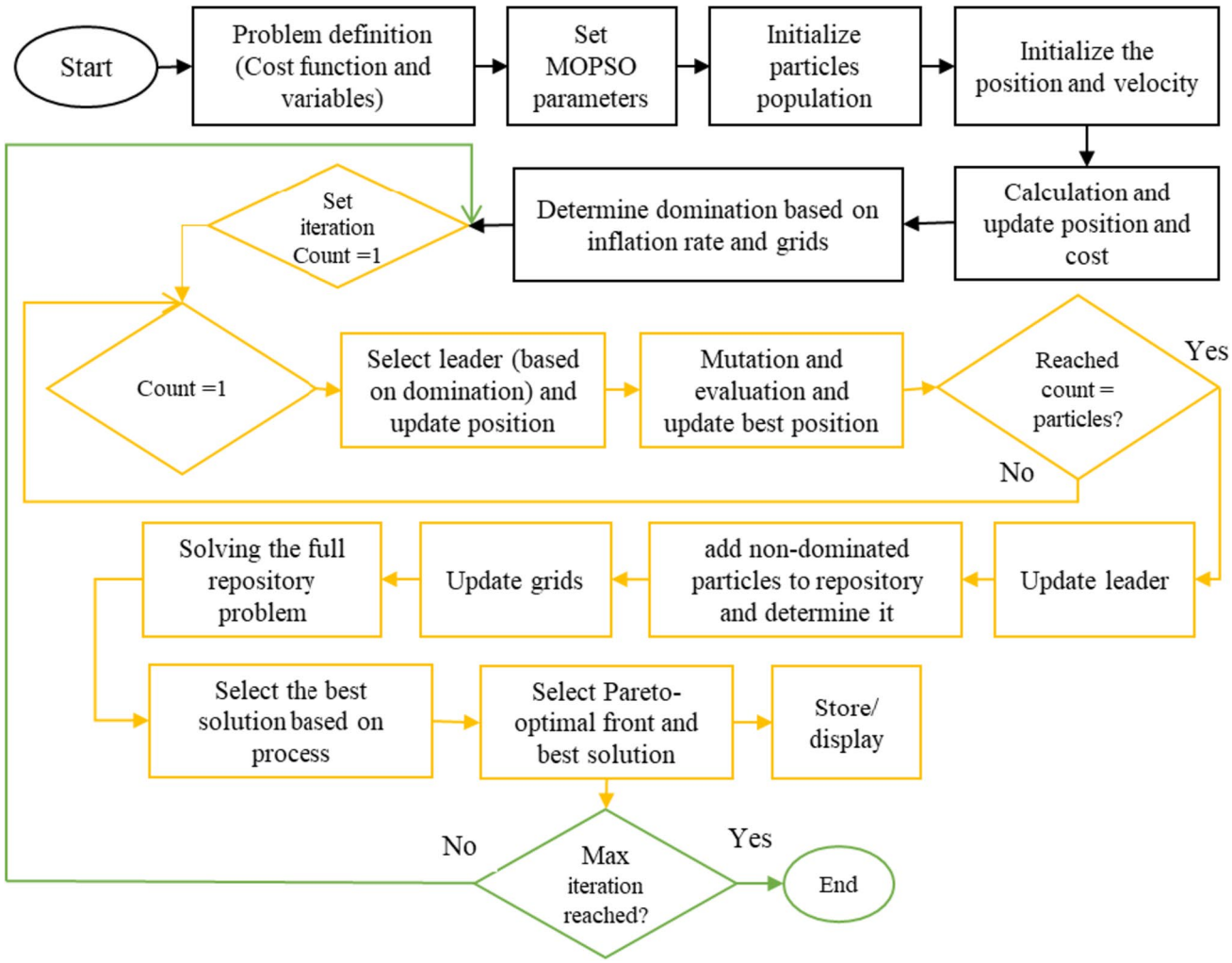

Fig. 7 The flowchart of the multi-objective particle swarm optimization method, adapted from (Coello and Lechuga 2002)

respectively. Then, random values were assigned as the initial guess of a, A, and C (Mirjalili et al. 2016). In the main process of the MOGWO algorithm, the solutions that had been found by search agents were examined at every step.

Classification methods

The class labeling and feature selection steps to classify ROP and TOB were explained. Here, the applied supervised classification methods will be briefly described.

Fine decision tree classifier

Decision trees classifier (DTC) learns from the general pattern of training data and predicts the class of test via the acquired pattern. In DTC's elements, the features are nodes that are scattered in the solution spaces (Singh and Singh 2017). The criteria and features of the classifier are expressed in Table 5. Here, the fine DTC was used as follows:

$$\mathrm{DTC}_{\mathrm{Best}} \cong \min\left(E_{\mathrm{DTC}}\left(T_w\right)\right); \quad w = 1, 2, \ldots, W \tag{12}$$

$$E\left(T_w\right) = \sum_{n \in T_w} r(n)P(n) \tag{13}$$

where $\mathrm{DTC}_{\mathrm{Best}}$ shows the best DTC. E_{DTC} represents the error of classification of the tree T_w and w denotes the tree index number, $r(n)$ and $P(n)$ are re-substitution prediction errors for the wrong classified nodes and the probability in comparison values with criteria, respectively (Singh and Singh 2017).

Support vector machine classification

The support vector machine (SVM) methods are powerful in nonlinear analysis. There are two types of SVM

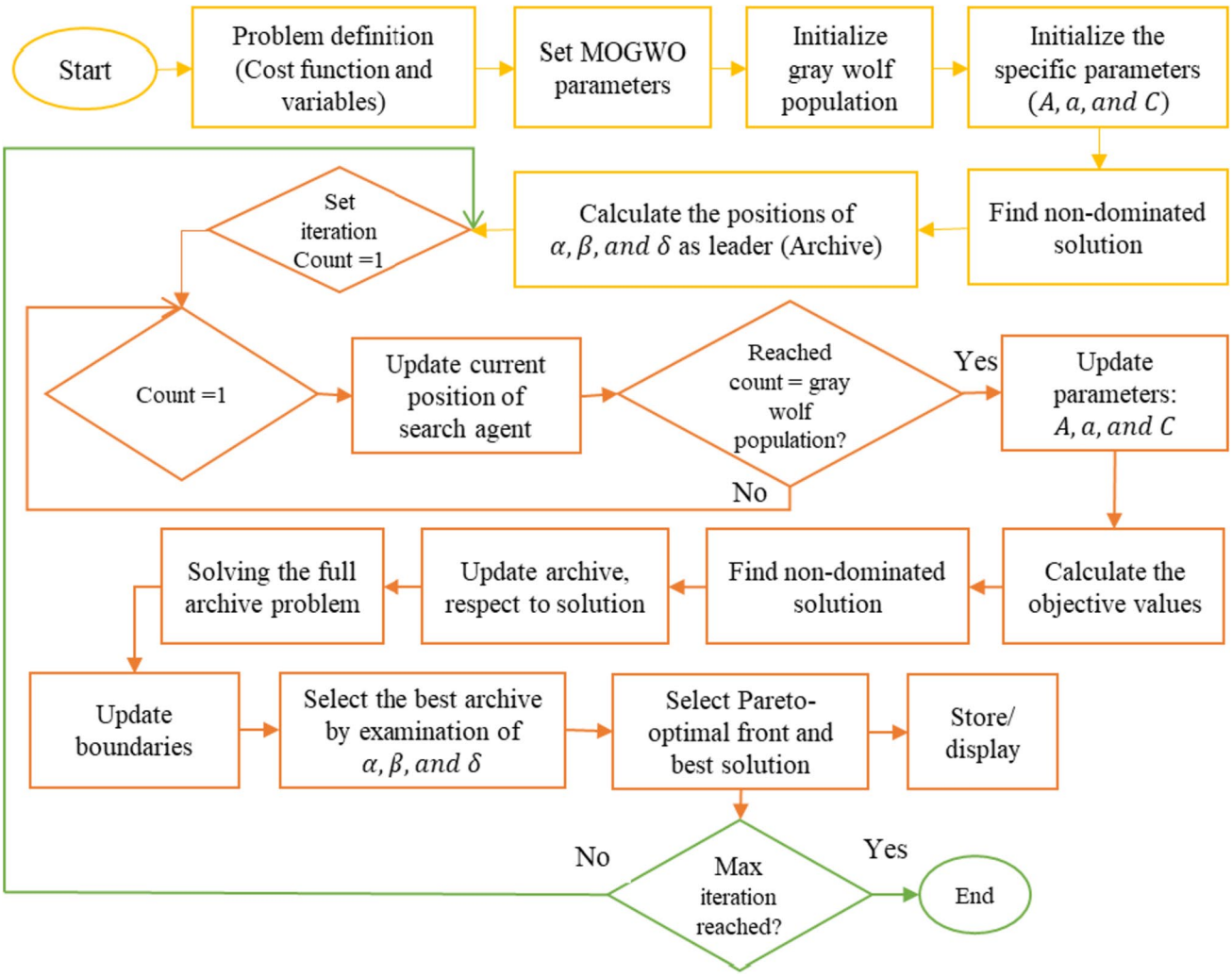

Fig. 8 The flowchart of the multi-objective Gray wolf optimizer method, adapted from (Mirjalili et al. 2016)

Table 5 The details of user-defined features which applied on the three classification approaches

Classification approach	User-defined parameter	Value/type
Decision trees classifier (DTC)	Split criterion	Gini's diversity index
	Maximum number of distinctions	100
	DTC version	Fine tree
Cubic SVM	Kernel function	Cubic polynomial
	Box constraint level	1
	Kernel function power (q)	3
Fine SVM	Kernel function	Gaussian kernel
	Box constraint level	1
	Variance of Gaussian kernel (σ^2)	0.61

classifications, linear and nonlinear (Cortes and Vapnik 1995; Tatar et al. 2014). The main function of SVM classification is expressed in Eq. 14. The SVM structure has been modified to utilize kernel function in studies due to its powerful capability (Ding et al. 2017). In this effort, the cubic polynomial and fine Gaussian kernel functions were applied to the SVM, their kernel functions are implied via Eqs. 15 and 16, respectively. The details of the setting for both SVM classification methods are depicted in Table 5.

$$f(x) = sgn\left(w.\emptyset(x) + b\right) \tag{14}$$

$$K\left(x_j, x_k\right) = \left(1 + x_j^T x_k\right)^q \tag{15}$$

$$K\left(x_j, x_k\right) = \exp\left(-\frac{x_j - x_k}{\sigma^2}\right); \quad \sigma^2 = \sqrt{P}/4 \tag{16}$$

where $\emptyset(x)$ represents the mapping of x from estimator samples space (R^n) to feature space, p is the number of predictors, σ^2 represents the variance of the Gaussian kernel (Kernel scale), w is weight vector and b denotes the bias value. Also, x_j and x_k are input variables as features and labels, respectively.

Ensemble classification

In machine learning classification collection, ensemble learning is an improvement in the capability and flexibility of the classifications (Mousavi and Eftekhari 2015). Here, two types of ensemble methods, the RUSBoosted trees (RUSB.T) and bagged trees (Bagged.T) classifications were used. The special parameters of the ensemble models are the number of splits and learners. In fact, by increasing the number of splits, the capability and overfitting of results are improved (Saeed et al. 2019). Table 6 shows the details of ensemble classifications used in the present study. The maximum number of splits in the bagged trees method varies due to using the classification method in different zones of the well.

Semi-supervised learning method

Laplacian SVM (LapSVM) is a semi-supervised learning method that predicts unlabeled data through a few labeled samples (Melacci and Belkin 2011). The LapSVM relies on SVM methods, and the main problem is described in Eq. 17 with three parts. The first part of Eq. 17 is related to the labeled samples, and the two other parts involve unlabeled and labeled samples (Dong et al. 2020). Where $V\left(x_i, y_i, f\right)$ is determined by the SVM function. Also, f is the decision function of f_A^2 which formed the kernel of the problem; for other functions and details refer to (Ding et al. 2017; Dong et al. 2020; Melacci and Belkin 2011).

$$\min \frac{1}{l} \sum_{i=1}^{l} V\left(x_i, y_i, f\right) + \gamma_A f_A^2 + \gamma_I f_I^2 \tag{17}$$

Based on Eq. 17 that is the main function of LapSVM, it can be applied to ROP and TOB analysis. Generally, three labels are typically used for LapSVM: labels '1' and '− 1' for learning the algorithm and label '0' for prediction. First, the zones with proper ROP and TOB ranges, whose averages are higher than 4.6 (m/hr) for ROP and below 3593.5 (lbf.ft) for TOB, were labeled '− 1'. Improper regions of ROP and TOB were assigned to label '1'. The areas whose LapSVM should be predicted were labeled as '0'. Condition 1 has its prediction by LapSVM. In condition 2, the coefficients of drilling features with the same geomechanical characteristics from the proper ROP and TOB (labeled '− 1') were multiplied by the features in the improper area (labeled '1'). The results can indicate the optimal ranges of the drilling factors in the labeled zones. Figure 9 illustrates the process of prediction by LapSVM in two mentioned conditions.

Classification analysis methods

The confusion matrix, a classification assessment, indicates the performance matrix, including the true and predicted classes, and it can be explained by Eq. 18 (Chamkalani et al. 2017). Where, $RR_{i,j}$ is the total number of class i values, and the principal diagonal elements are $\left(RR_{1,1}, \ldots, RR_{N,N}\right)$. The performance of classification could be evaluated considering the principal diagonal on the total values as accuracy (Eq. 19). The area under the curve (AUC) is an assessment and is defined as Eq. 20 (Fawcett 2006).

$$\text{Confusion matrix} = \begin{bmatrix} RR_{1,1} & RR_{1,2} & \ldots & RR_{1,N} \\ RR_{2,1} & RR_{2,2} & \ldots & RR_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ RR_{N,1} & RR_{N,2} & \ldots & RR_{N,N} \end{bmatrix} \tag{18}$$

$$\text{Accuracy} = \frac{\text{Principal diagonal elements } \left(RR_{1,1}, RR_{2,2}, \ldots, RR_{N,N}\right)}{\text{Total values } \left(RR_{i,j}\right)} \tag{19}$$

Table 6 The feature of applied ensemble classifiers

Methods	Learner type	Maximum number of splits	Number of learners	Learning rate
Bagged trees	Decision trees	60 to 120 (changeable for zones)	30	–
RUSBoosted trees	Decision trees	20	30	0.1

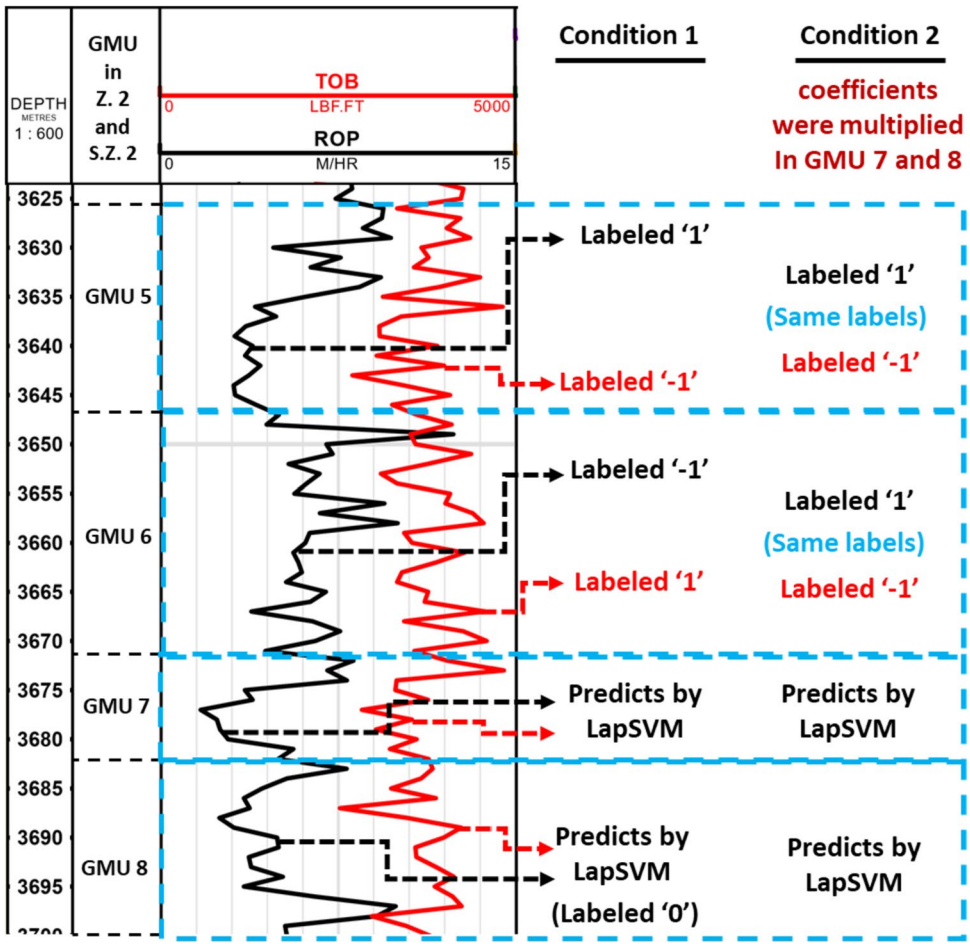

Fig. 9 Description of how the Laplacian support vector machine method used in two conditions in zone 2 and sub-zone 2 of the well

Table 7 The output of a confusion matrix as binary problem

Confusion matrix	Target	
	True	False
Model of classification		
True	True positive	False positive
False	False negative	True negative

$$\mathrm{AUC} = \frac{S_a - n_p(n_n + 1)/2}{n_p n_n} \tag{20}$$

where S_a, n_n and n_p are the sum of all positive, the number of negative, and the number of positive values, respectively. Table 7 shows the binary outputs of a classification model (Oyedere and Gray 2020a).

Results and discussion

The user-define classes of the ROP and TOB and also the methodologies of classification approaches were described in "Methodology" section. In the forthcoming sections, the results will be explained.

Feature selection results

The three feature selection methods, including the NPR-NSGA-II, NPR-MOPSO, and NPR-MOGWO, were implemented on the dataset. In the NPR method, the structure was selected by trial and error. The features were WOB, RPM, HKL, Pump.P, Flow.R, UCS, IFA, and CCS where the targets were the ROP and TOB. The number of classes was 7, equal to the number of the well zones. The results are presented in Fig. 10 using a confusion matrix. Figure 10 (sum up of train and test) shows 94.6% and 5.4% for the true classification and misclassification, respectively. Besides, the results of the training and testing parts are shown in "Appendix C".

Here, the NPR model was implemented with the multi-objective optimization methods, and their results illustrated

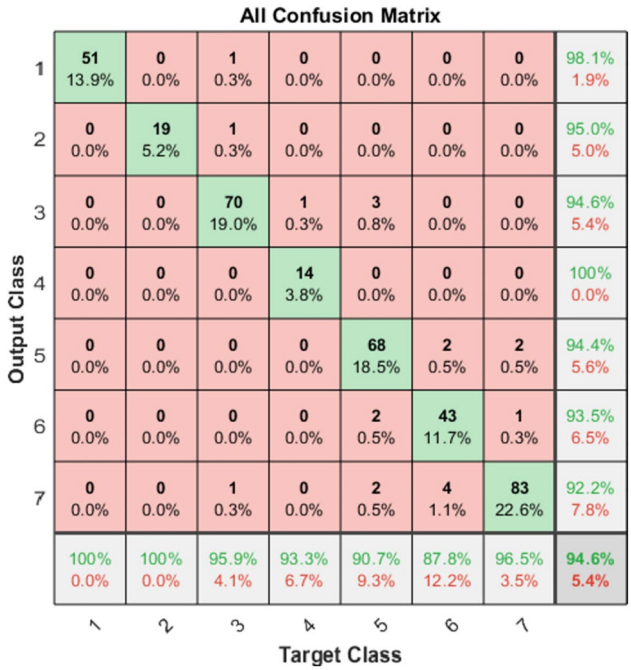

Fig. 10 Target classes with all zones based on confusion matrix of the neural pattern recognition method in all (sum up of train and test) version

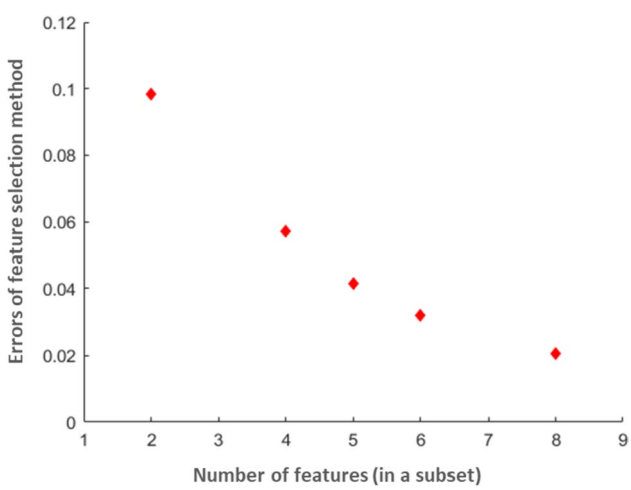

Fig. 11 The results of the neural pattern recognition combined with the multi-objective Gray wolf optimizer in 30 populations, where the model with 8 features depicts the lowest error

the options for the features in every scenario. Figure 11 shows the results of NPR-MOGWO with a number of 30 for the population and in 100 iterations using the Pareto-optimal points. The vertical and horizontal labels represent the errors (E) or cost function and number of features (n_f), respectively. The best feature selection with a minimal error occurred by 8 features for NPR-MOGWO (Fig. 11). Concerning the effect of population on optimization, five different populations were tested and the least errors (likewise the point in the right corner of Fig. 11) were gathered. So, the best n_f and minimum errors are listed in Table 8.

Table 8 expresses the results of the three methods. All of the iterations' numbers of optimization were 100. For every hybrid method, the bolded values are the best option with the minimum error. In the final step, common aspects of the results were considered. Therefore, the features that had been selected two or more times, among the three bolded options, were chosen for ROP and TOB classification. The features were WOB, HKL, Pump.P, Flow.R, UCS, and IFA, which are expressed in Table 9 for the case well by statistical indexes. The statistical aspects of the second well are shown in "Appendix D".

Classification results

Five classification procedures were applied to the dataset (Table 9). Table 10 shows the results in AUC and the accuracy of supervised methods for ROP. Moreover, the AUC of ROP is depicted in Fig. 12. In Table 10, there are a few zones that their superior classification procedure was not evident. Therefore, in addition to AUC, the accuracy and confusion matrix can indicate the superior method. In Z.1 and Z.3, the AUC values were close thus, Fig. 13 presents the corresponding confusion matrices. As a result, the Bagged.T classified labels better than other methods; its accuracy was also high (Table 10).

Table 11 shows the results of TOB classification in the zones with AUC and accuracy. Moreover, the AUC comparison in Fig. 14 clarifies the assessment. In the zones of Fig. 14, the outperformed classification was not obvious. Figure 15 presents the confusion matrices of the two zones (Z.2 & SZ.2 and Z4), which reveal the superiority of C.SVM. Other accurate and reliable methods for TOB and ROP are presented in Table 12. According to the results, the Bagged.T and C.SVM had the least errors among approaches. Besides, Table 13 shows the outcomes of the ROP and TOB classification methods in the testing section of zones of the well using AUC and accuracy.

In addition, the classification methods were assessed in new zones when they were trained by the case well and tested using the dataset of the second well. Figure 16 shows the outcomes of the given process as the average of ROP and TOB classification where Bagged.T and C.SVM showed the best performance in terms of both AUC and accuracy. Based on the results, the F.SVM method cannot fulfill a reliable classification manner for ROP and TOB classification. Despite the fact that the classification accuracies were dropped around an average of 2–8% for the second well as opposed to the models of the case well, they depicted precise results. Therefore, the superior approaches of the findings, Bagged.T and C.SVM, can be implemented in other

Table 8 The results of feature selection methods with the lowest error and the best features for each population

Population(s)	Methods	Features	Error (*E*)
5	NPR-MOGWO	WOB-RPM-HKL-Flow.R-UCS-CCS	0.0413
	NPR-NSGA-II	WOB-RPM-HKL-Flow.R-UCS-IFA-CCS	0.0204
	NPR-MOPSO	**WOB-HKL-Pump.P-Flow.R-UCS-IFA**	**0.0203**
10	NPR-MOGWO	WOB-Pump.P-Flow.R-UCS-IFA	0.0469
	NPR-NSGA-II	WOB-HKL-Pump.P-Flow.R-IFA-CCS	0.0211
	NPR-MOPSO	WOB-RPM-HKL-Flow.R-UCS-IFA	0.0247
15	NPR-MOGWO	WOB-HKL-Pump.P-Flow.R-IFA-CCS	0.0324
	NPR-NSGA-II	**WOB-HKL-Pump.P-Flow.R-UCS**	**0.0166**
	NPR-MOPSO	WOB-RPM-HKL-Pump.P-Flow.R-IFA-CCS	0.0240
20	NPR-MOGWO	WOB-RPM-HKL-Pump.P-Flow.R-UCS-IFA-CCS	0.0269
	NPR-NSGA-II	WOB-HKL-Pump.P-UCS-IFA	0.0215
	NPR-MOPSO	WOB-RPM-HKL-Pump.P-Flow.R-UCS	0.0259
30	**NPR-MOGWO**	**WOB-RPM-HKL-Pump.P-Flow.R-UCS-IFA-CCS**	**0.0206**
	NPR-NSGA-II	WOB-RPM-HKL-Pump.P-Flow.R-UCS-IFA	0.0170
	NPR-MOPSO	WOB-HKL-Pump.P-Flow.R-UCS-IFA-CCS	0.0232

The bolded rows show the least error for each hybrid method in the five population sets

Table 9 Three statistical features of the last selected parameters for classification methods

Statistical Index	Features					
	WOB	HKL	Pump.P	Flow.R	UCS	IFA
Minimum	2.22	293.7	12.5	448.06	19.73	20.9
Average	14.2	315.4	1098.6	467.9	50	27.8
Maximum	25.6	331.4	1358.6	502.4	85.27	37.2

Table 10 The results of classification procedures as training section in all of the zones for the rate of penetration

Zone	Methods									
	Bagged.T		RUSB.T		C.SVM		F.SVM		F.Tree	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Z.1	0.98	92.2	0.99	80.4	0.98	88.2	0.94	76.5	0.95	92.2
SZ.1	0.92	84.2	1	73.7	0.99	89.5	0.81	73.7	0.75	68.4
Z.2 & SZ.2	0.96	80.7	0.87	77.3	0.97	87.5	0.87	63.6	0.8	70.5
Z.3	0.99	97.3	0.99	96	0.99	96	0.97	74.7	0.99	96
SZ.3	0.94	81.6	0.88	71.4	0.9	79.6	0.75	55.1	0.83	73.5
Z.4	0.98	91.9	0.91	83.7	0.99	94.2	0.93	82.6	0.86	81.4

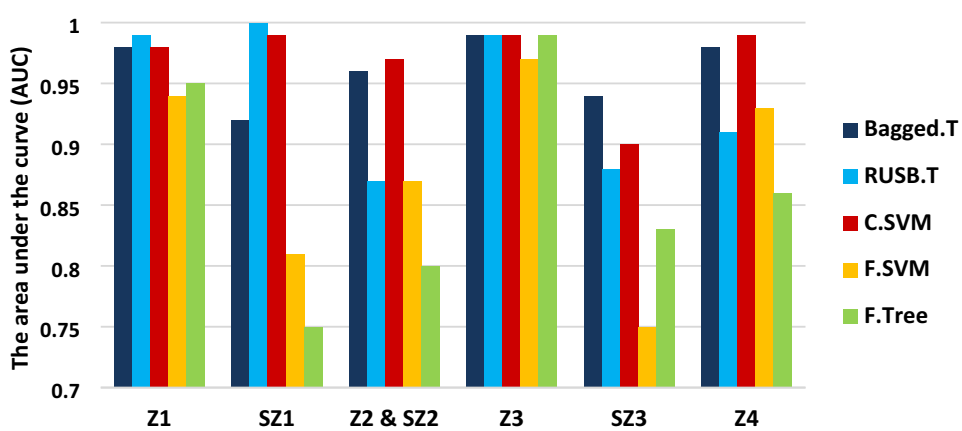

Fig. 12 The bar-chart of classification methods as the area under the curve (AUC) in zones for the rate of penetration

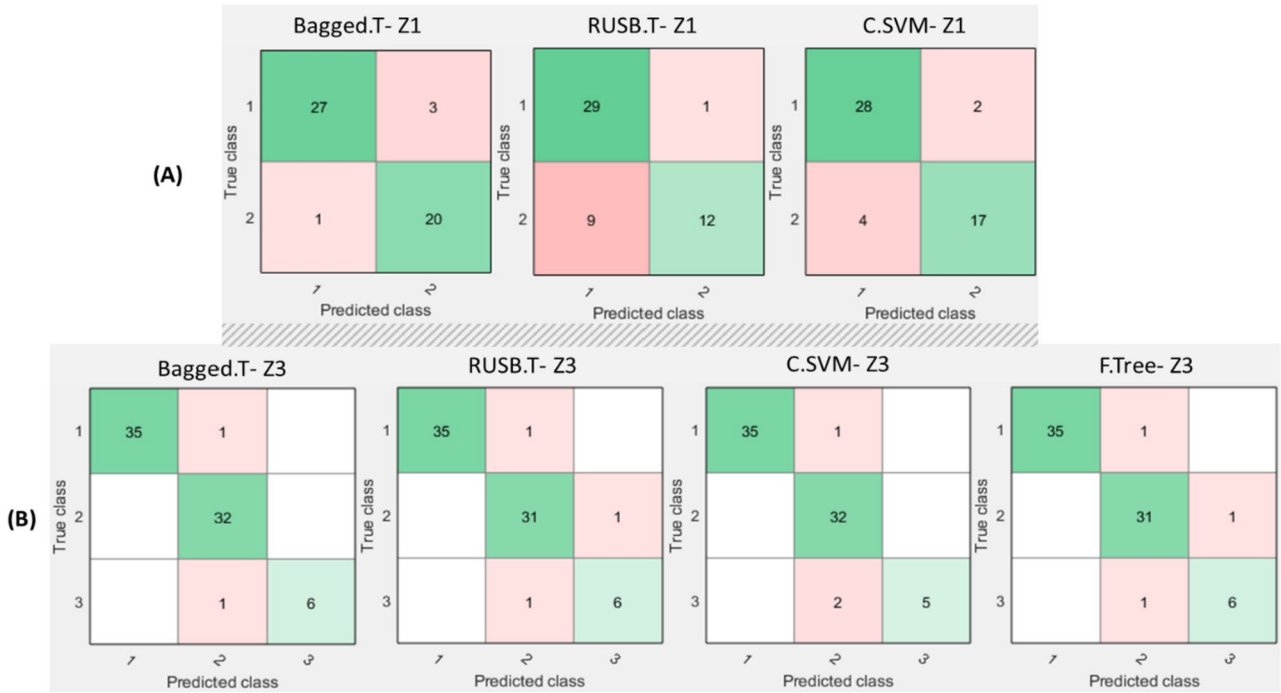

Fig. 13 The confusion matrices of the rate of penetration classification in two zones with close results as the area under the curve

Table 11 The results of classification methods as training section for the torque on bit in all of the zones

Zone	Methods									
	Bagged.T		RUSB.T		C.SVM		F.SVM		F.Tree	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Z.1	0.98	92.2	0.99	80.4	0.98	88.2	0.94	76.5	0.95	92.2
SZ.1	0.92	84.2	1	73.7	0.99	89.5	0.81	73.7	0.75	68.4
Z.2 & SZ.2	0.98	87.5	0.97	87.5	0.99	90.9	0.94	76.1	0.91	84.1
Z.3	0.99	97.3	0.99	96	0.99	96	0.97	74.7	0.99	96
SZ.3	0.99	89.8	0.97	81.6	0.98	77.6	0.89	59.2	0.97	79.6
Z.4	0.99	86	0.94	83.7	0.97	91.9	0.97	66.3	0.92	84.9

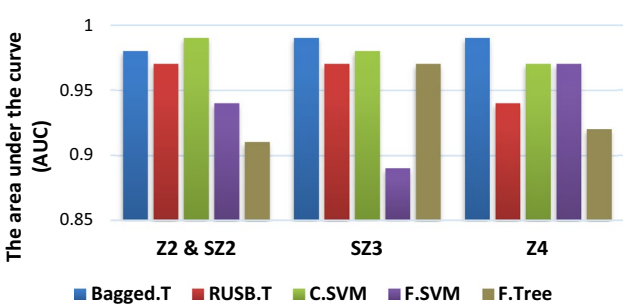

Fig. 14 The area under the curve (AUC) of classifications for the torque on bit in three zones

carbonate zones to classify ROP and TOB based on controllable drilling factors and geomechanical properties.

The importance of features

The importance of features was examined utilizing the classification-based sensitivity analysis for the best methods (C.SVM and Bagged.T). Based on the results (Table 12), either C.SVM or Bagged.T was the most precise method in every three zones of the well. In fact, the SVM structure maps the dataset into multi-dimensional space, thus it

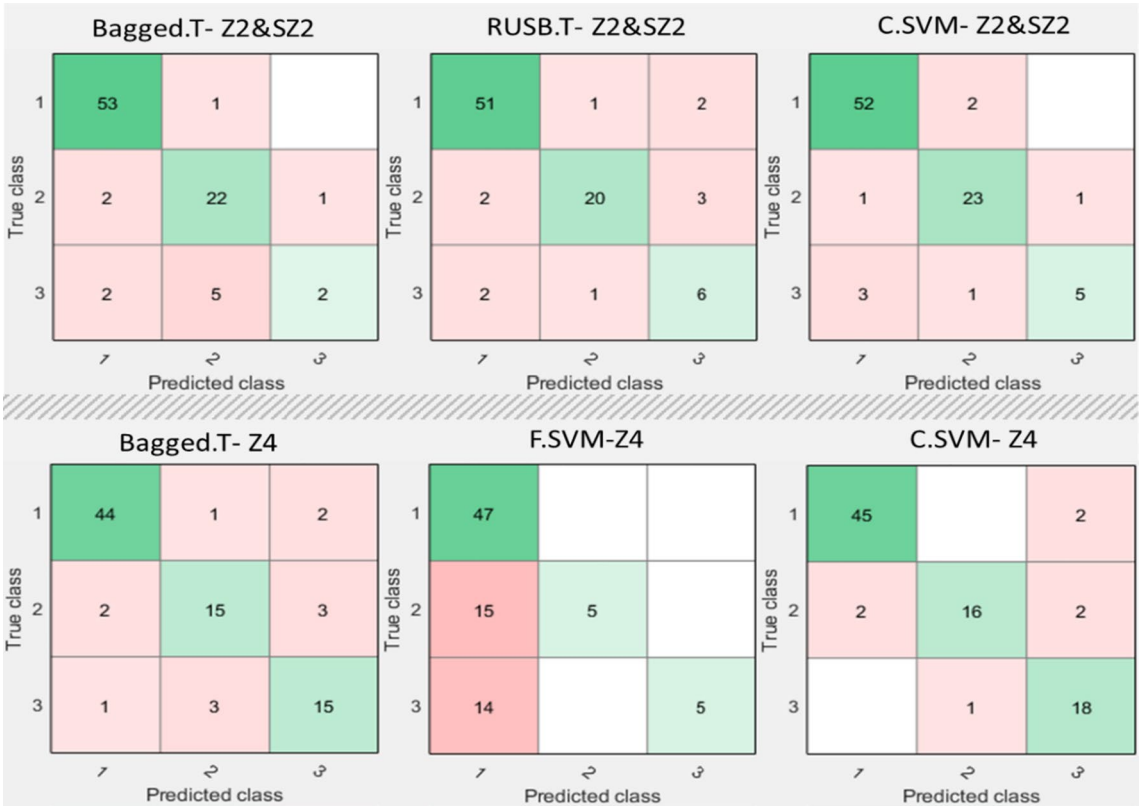

Fig. 15 The confusion matrices of the torque on bit classification in two zones with close results as the area under the curve

Table 12 The best procedure of classification the rate of penetration and torque on bit in all the zones of well

Zones	The best method	
	ROP method	TOB method
Z.1	Bagged.T	Bagged.T
SZ.1	C.SVM	C.SVM
Z.2 & SZ.2	C.SVM	C.SVM
Z.3	Bagged.T	Bagged.T
SZ.3	Bagged.T	Bagged.T
Z.4	C.SVM	C.SVM

makes the process of feature ranking via C.SVM impossible (Oyedere and Gray 2020a). Consequently, the significance of parameters was determined based on the feature ranking in Matlab software as the permuted predictor importance using classification Bagged.T in 50 cycles.

Figure 17 shows the relative importance of features in Z. 4 for TOB and ROP with bar charts. Flow.R and WOB were the most significant features for ROP and TOB, respectively. The total importance of parameters was gathered in box plots. Figure 18 shows the significance of parameters using classification for ROP and TOB in the zones. As a result of TOB (Fig. 18), the Pump.P had the highest importance although the maximum of Flow.R and WOB were higher than it. Besides, both UCS and IFA are illustrated as minus numbers because of declining the true classification. Concerning ROP (Fig. 18), the order of importance among drilling parameters was Flow.R, Pump.P, HKL, and WOB, the highest to least, respectively. Moreover, the most crucial geomechanical feature was UCS. Comparing the ROP and TOB boxplots (Fig. 18), Flow.R and Pump.P were more important than others in the zones. The higher importance of UCS on ROP classification can be justified by the influence of rock strength on the drilling rate.

Optimization results

In this section, the results of semi-supervised classification, LapSVM, and the optimum ranges of drilling parameters are presented. The details of labels, two conditions ("Semi-supervised learning method" section), and significant factors ("The importance of features" section) were constructed in the algorithm settings. Table 14 expresses the labels of GMU for ROP and TOB. The LapSVM algorithm predicted the GMUs' classes (for label '0') utilizing the training section (labels '1' and '− 1') in the condition 1. The labels of the condition 1 were selected from a limited number

Table 13 The results of the rate of penetration and torque on bit classification in testing section in all of the zones as the area under the curve and Accuracy

Target	Zone	Methods									
		Bagged.T		RUSB.T		C.SVM		F.SVM		F.Tree	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
ROP	Z.1	0.96	83.3	0.8	58.3	0.99	91.7	0.94	75	0.83	83.3
	SZ.1	1	99	1	99	1	99	1	80	0.83	80
	Z.2 & SZ.2	0.97	81.8	0.93	72.7	1	86.4	0.9	59.1	0.95	77
	Z.3	1	99	1	90.9	1	95.5	0.96	59.1	1	95.5
	SZ.3	1	71.4	1	71.4	0.98	85.7	0.75	50	0.92	64.3
	Z.4	1	99	1	95.2	1	95.2	1	85.7	0.93	95.2
TOB	Z.1	0.96	83.3	0.8	58.3	1	91.7	1	75	0.83	83.3
	SZ.1	1	99	1	99	1	99	1	80	0.83	80
	Z.2 & SZ.2	1	90.9	0.99	90.9	0.98	95.5	0.94	72.7	0.95	86.4
	Z.3	1	99	1	90.9	1	95.5	0.96	59.1	1	95.5
	SZ.3	0.97	83.3	0.97	75	1	75	0.69	50	0.69	66.7
	Z.4	1	85.7	1	85.7	1	85.7	0.95	61.9	0.95	81

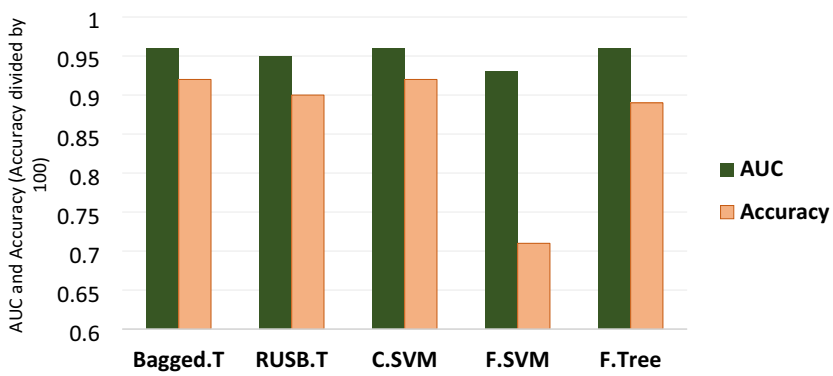

Fig. 16 The results of the classification methods in the second well

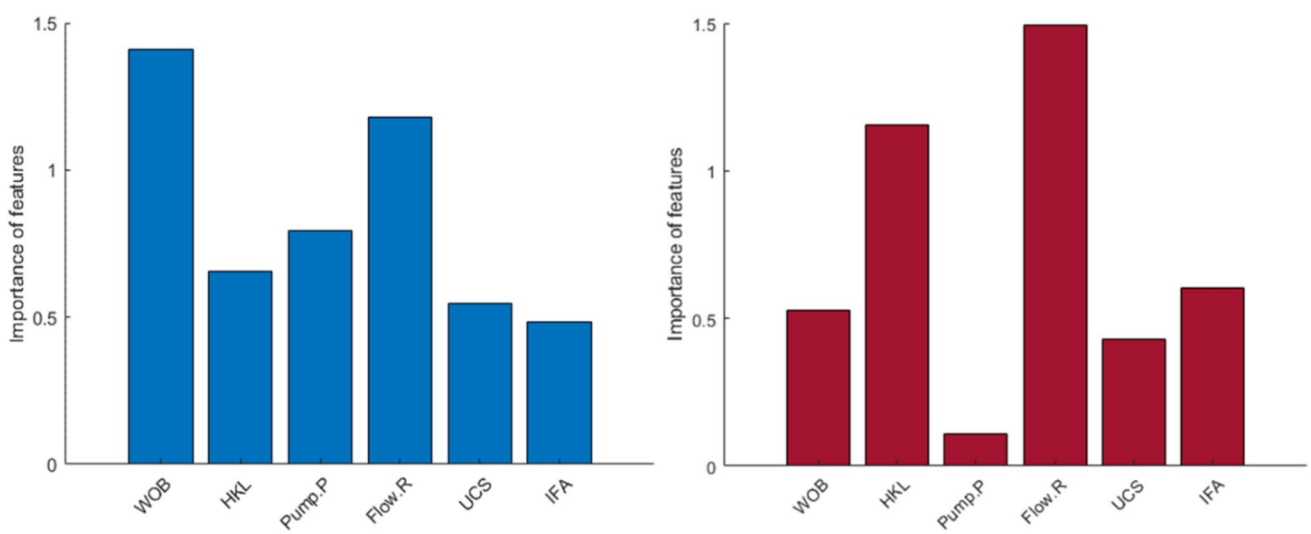

Fig. 17 The importance of features using classifications in the left and right bar-charts for the torque on bit and rate of penetration in zone 4, respectively

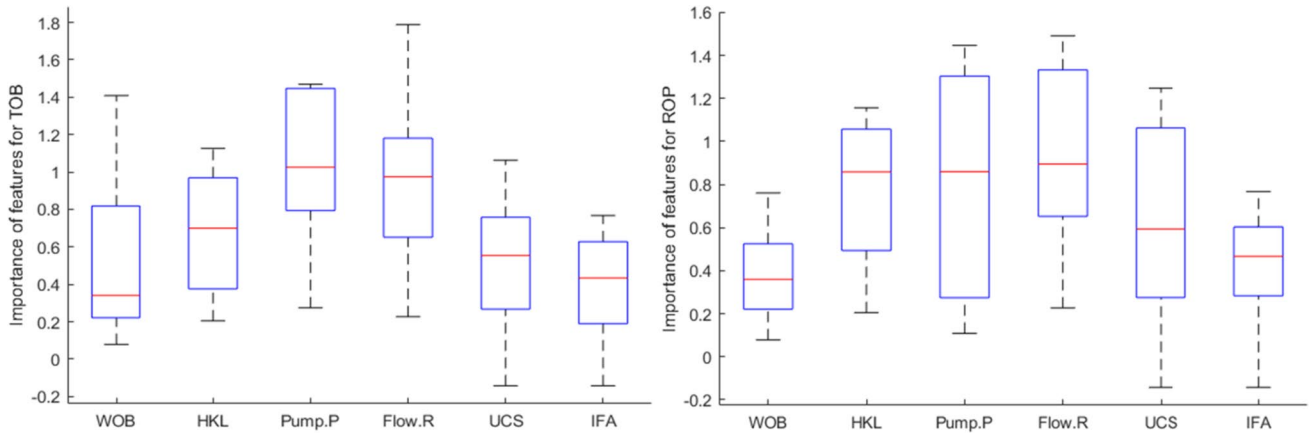

Fig. 18 The results of bagged trees for the importance of features in all the zones as boxplots; the left and right boxplots are related to the torque on bit and rate of penetration, respectively

Table 14 Details of two conditions applied in the Laplacian support vector machine algorithm included labels and geomechanical units' names

Parameters	Conditions	Labels	GMU name
ROP	1	− 1	2, 5, 6, 8, 12, 15
		1	7, 9, 11, 16, 18, 19
		0	1, 3, 4, 10, 13, 14, 17, 20, 21
	2	− 1	1, 2, 3, 4, 5, 6, 8, 12, 15
		1	7, 9, 11, 16, 18, 19
		0	10, 13, 14, 17, 20, 21
		Similar GMUs as average of UCS and IFA (used in condition 2): 10►15; 13►3; 14►12; 17►3; 20►3; 21►3	
TOB	1	− 1	1, 2, 7, 17, 18, 19
		1	4, 6, 8, 12,16, 20
		0	3, 5, 9, 10, 11, 13, 14, 15, 21
	2	− 1	1, 2, 3, 7, 17, 18, 19, 21
		1	4, 6, 8, 12, 16, 20
		0	5, 9, 10, 11, 13, 14, 15
		Similar GMUs as average of UCS and IFA (used in condition 2): 9►3; 5►1; 10►1; 13►3; 14►19; 11►3; 15►1	

intentionally to depict the capability of the semi-supervised classification as opposed to the supervised methods.

Concerning the condition 2, it expresses a broader range of labels compared the condition 1. Furthermore, in the condition 2, the significant factors that were revealed in previous section, including HKL, Pump.P, and Flow.R, had an important role. Regarding this condition, the LapSVM was trained with the given classes while the coefficients of significant parameters from a similar GMU (shown in Table 14) were multiplied in the improper GMUs (GMUs with low ROP and high TOB). The averages of IFA and UCS established the similar GMU of the zones. Then, a new class prediction of the unseen testing part by LapSVM indicated the intervals where the ROP and TOB were turned to the optimal sections (high ROP and low TOB). The outcomes of both LapSVM's conditions were interpreted and illustrated.

Figure 19 illustrates the results of LapSVM in 7 tracks. The first and second tracks are depth and GMU, respectively. The third track represents the logs of ROP and TOB in the well. Besides, the 4th and 5th tracks are related to ROP, whereas the 6th and 7th tracks describe two conditions of TOB. In the first condition of ROP (track 4), there are three lines. The left and right dash lines (black color) are labels '− 1' and '1', respectively. The middle lines represent labels '0' which should be predicted by the algorithm. Additionally, the red middle lines show true labels of low ROP, whereas the green ones represent high ROP. Colors can help in the determination of LapSVM-predicted

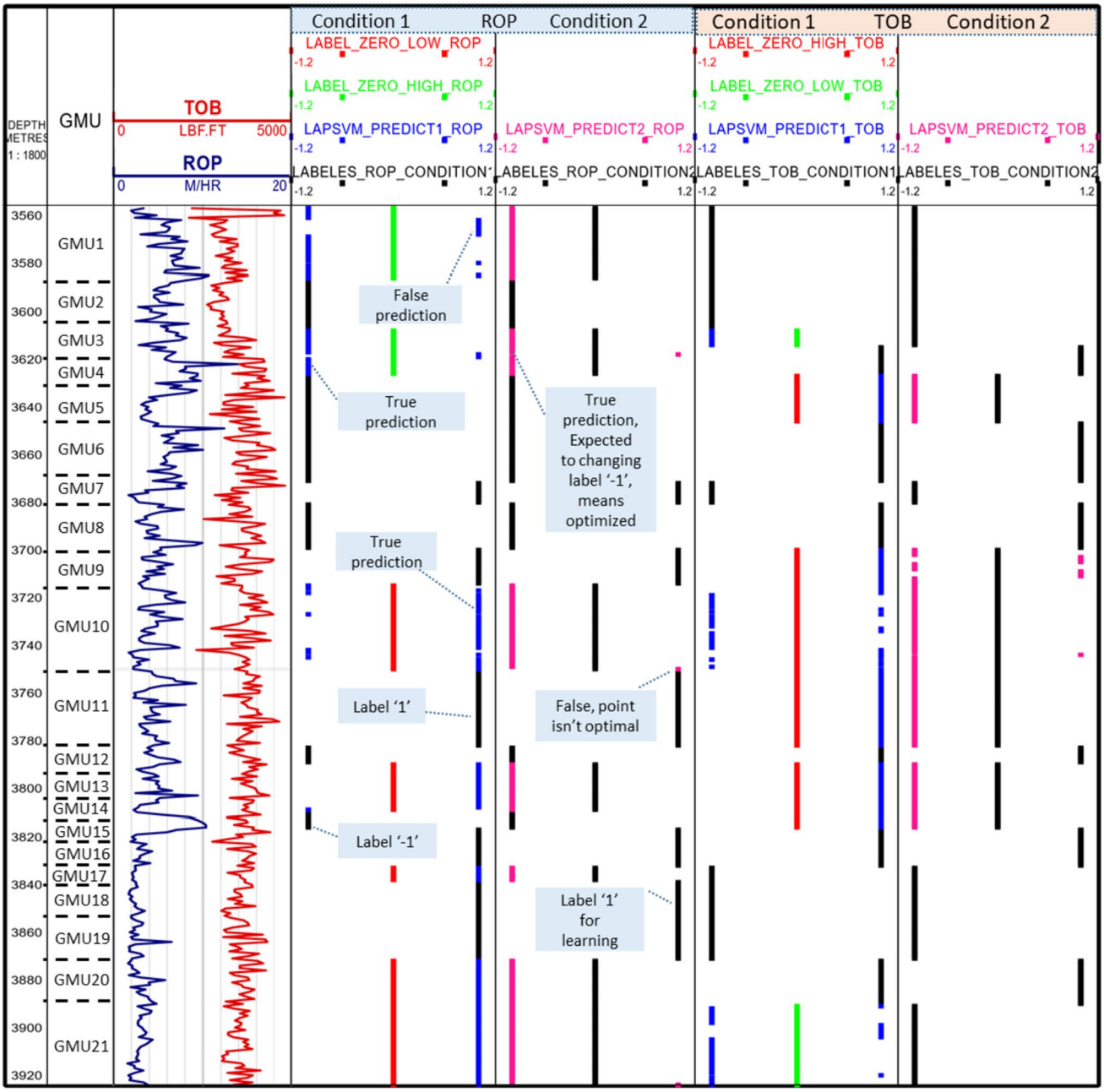

Fig. 19 The Laplacian support vector machine predictions for the rate of penetration (ROP) and torque on bit (TOB) in two conditions, left to right, tracks 1 to 3 include depth, geomechanical units, and ROP and TOB, tracks 4 and 5 are predictions of the algorithm in conditions 1 and 2 for ROP, tracks 6 and 7 are the same predictions for TOB, the description of track 4 is the same as track 6 and also, tracks 5 is like track 7

labels. Predicted classes are denoted by blue lines. So, if the green lines (high ROP) are predicted as label '1', they are false (misclassification, blue lines), and the true labels are '− 1'. On the other hand, when the red lines (low ROP) are predicted as label '1', they are true (blue lines) while the false labels are '− 1'. As a result, most parts (88% of the entire depth) of the studied zones were classified as true labels for ROP and TOB despite the limited training labels in condition 1.

The results of the second condition of ROP are shown on track 5. In this section, same as the condition 1, the testing part was labeled '0' (low ROP, black lines in the middle). As regulation of the condition 2, the important drilling parameters of the intervals (shown in black color in the middle) were multiplied by the same proper GMU (high ROP and low TOB). The training dataset was labeled '1' and '− 1' (black lines on the right and left parts of track 5), including the low and high classes of ROP, respectively. In

the results of condition 2, if they are classified as label '1', they did not turn to the optimal class (label '− 1', higher ROP). In this way, the outcomes of the second condition are shown via purple lines on track 5. According to the classification, roughly 95% of the testing section as the label '0' (black middle lines, low ROP) transformed to the label of high ROP (pink lines on the left of the track).

The LapSVM classification of TOB is illustrated to resemble ROP in Fig. 19 in tracks 6 and 7, whereas the low TOB (proper) was labeled as '− 1' (opposite of ROP). The LapSVM for TOB in the target GMUs was as reliable as ROP. In track 6 of Fig. 19, green lines should be predicted as label '− 1', and red lines were expected to be label '1' which occurred in most depths. Moreover, in track 7, approximately entire intervals were optimized (label '− 1', low TOB).

Consequently, three GMUs were employed for acquiring the optimal drilling parameters, which led to the optimization of ROP and TOB, whose ranges are shown in Table 15. Other GMUs were implemented for optimization either ROP or TOB, and also some GMUs were used for training the algorithm. Importantly, target parameters of the optimization method (ROP and TOB) interact in the process of drilling. Therefore, ROP and TOB should be presented in a reasonable range due to the mentioned constraints. The optimal values of the three GMUs were obtained through the optimization process of ROP and TOB (condition 2). These outcomes, Table 15, are practical to reach a desired drilling operation with such geomechanical characteristics. Furthermore, evaluating zones for their classification by using the optimal drilling parameters in the future project not only offers a higher ROP and lower TOB but also protects reservoirs and tools from unforeseen hazards.

Conclusions

Analysis of the rate of penetration (ROP) and torque on bit (TOB) is the primary subject of drilling assessments. In this paper, the given parameters were determined by the nonlinear classification in a carbonate reservoir. The novelty of the study lies in implementing the supervised and semi-supervised machine learning classification approaches and utilizing both geomechanical and drilling factors to establish classes. First, the user-defined labels were assigned to zones. Next, as the feature selection (FS) procedure, a neural pattern recognition (NPR) hybridized with three multi-objective optimization algorithms, the best features were revealed for classification. The supervised classifications were the decision trees classifier (DTC), support vector machine (SVM), and ensemble classifiers. Then, the bagged trees method indicated the importance of features in classification. The findings of this section of the study were as follows:

(1) The combined FS methods depicted the most effective subset of factors for classification with errors of less than 0.02;
(2) The results of classifications showed that bagged trees and cubic SVM had the least errors compared to others;
(3) The results confirmed the credibility of the two superior approaches, their average area under the curve (AUC) of ROP classification were respectively 0.96 and 0.97;
(4) In the importance rate of bagged trees, the flow rate and pump pressure were introduced as the most vital features of drilling factors.

Finally, the semi-supervised method, Laplacian SVM, was trained via a limited number of labels (high/low ROP and TOB) and accuracy was assessed. Next, in similar GMUs, the crucial drilling factors, introduced by bagged trees, were multiplied in non-optimal GMUs. The final step indicated the following results:

(1) The LapSVM showed the classes with an accuracy of 88% despite utilizing half of the inputs in the supervised classification;
(2) The LapSVM exposed the optimal classes in 95% of the studied areas, which recommend optimized values for the zones;
(3) The outcomes proved the classifications can expedite the process of seeking effective drilling parameters concerning geomechanical properties, which is lacking in the previous efforts;
(4) The optimal drilling parameters can be applied to optimize the ROP and TOB although they can be extended with more reservoir characteristics in future efforts.

Table 15 The optimal values of drilling parameters and the geomechanical features of three geomechanical units, all the values are in ranges

GMU	Optimal ranges of drilling parameters						Ranges of geomechanical features			
	HKL		Pump.P		Flow.R		UCS		IFA	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
10	355.2	302.7	1655.5	823.1	554.2	403.2	64	37	35	22
13	338.7	305.2	1701.4	1563.9	543.6	540	60	22	29	21
14	329.7	306.3	1701.3	1376.6	559.5	496.4	30	28	28	25

Appendix A

See Fig. 20.

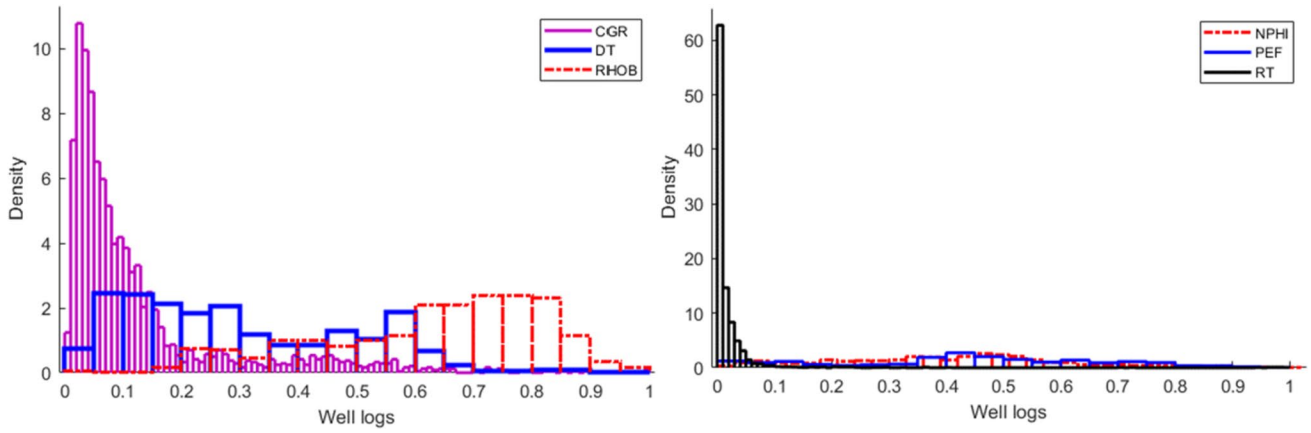

Fig. 20 The probability density function (PDF) graphs of the features that show the range of the input factors

Appendix B

See Fig. 21.

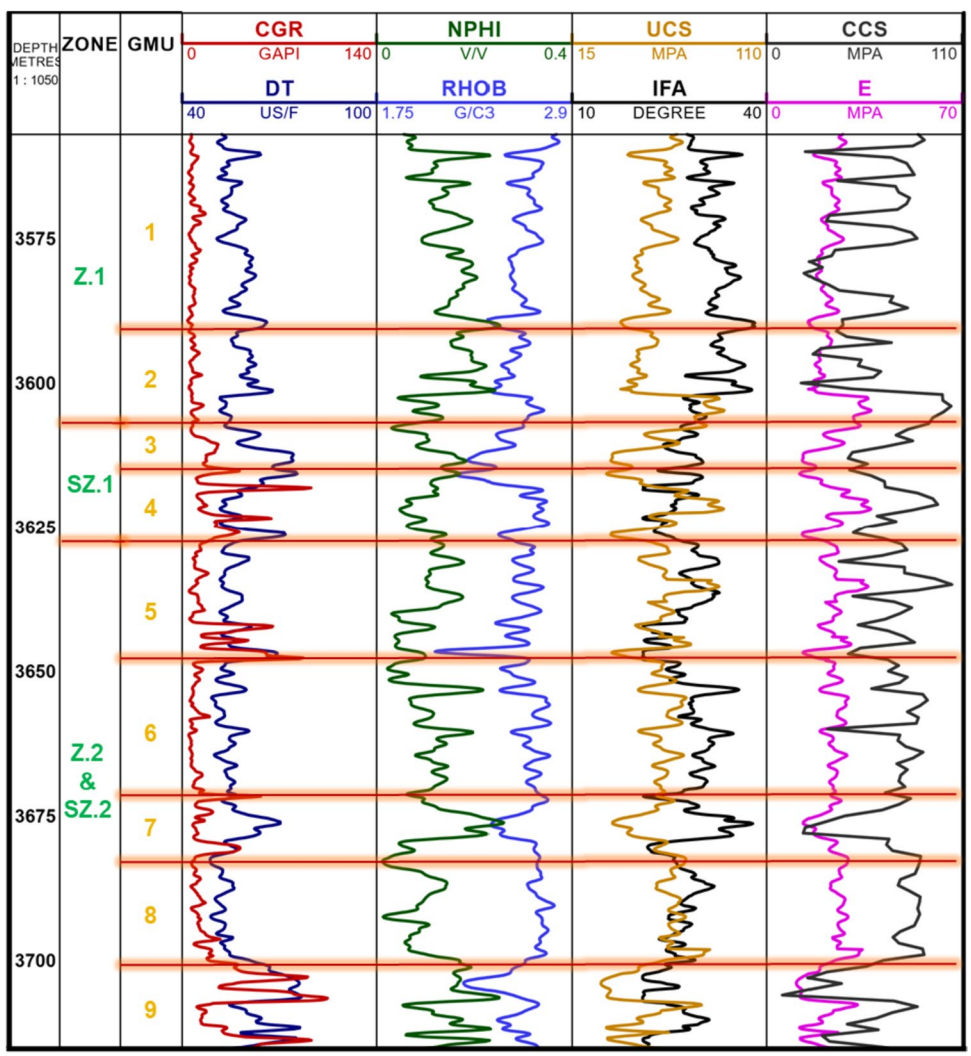

Fig. 21 The conventional well logs and geomechanical parameters of the case in the geomechanical units (GMUs) 1 to 9; tracks 1 to 3 are included depth, zones and GMUs, and petrophysical logs are in tracks 4 and 5, and also, tracks 6 and 7 are consisted of the trends of geomechanical features in the well

Appendix C

See Fig. 22.

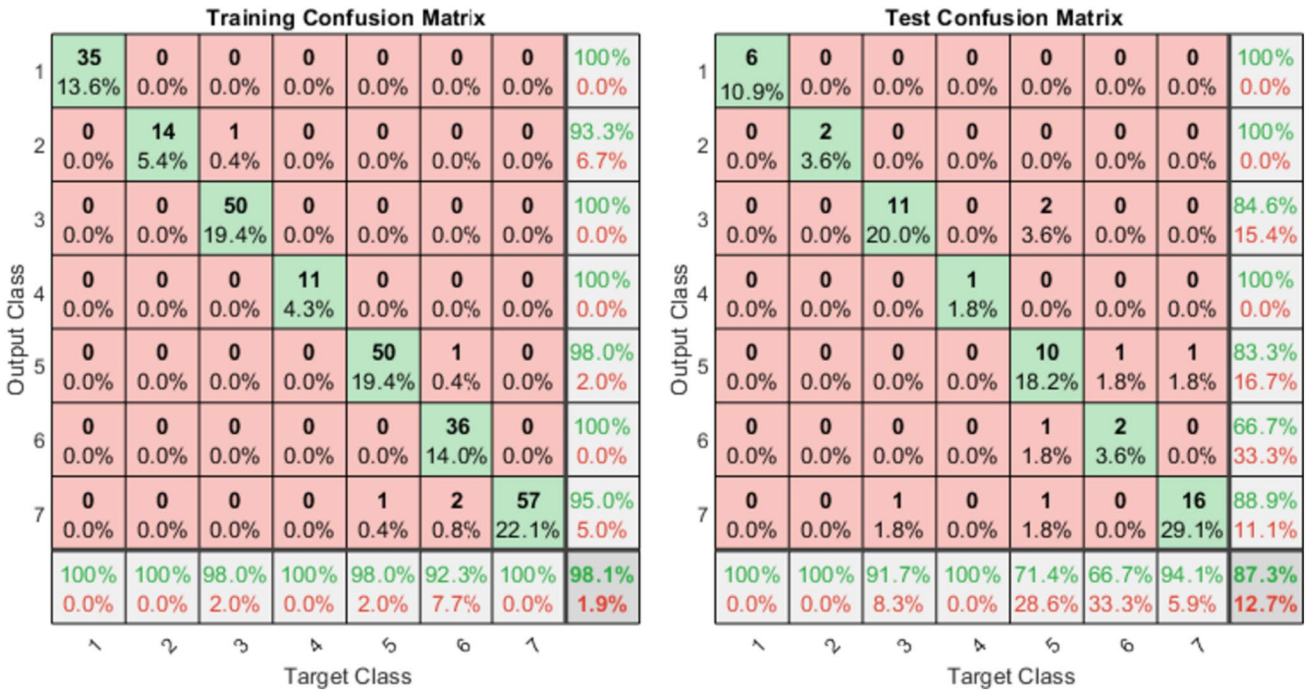

Fig. 22 The confusion matrices of the well zones included outputs and target classes for training and testing section

Appendix D

See Table 16.

Table 16 The statistical indexes of the features of the second studied well

Statistical Index	Features							
	ROP	TOB	WOB	HKL	Pump.P	Flow.R	UCS	IFA
Minimum	0.41	859.9	4.28	213.9	1842.09	291.1	37.1	22.3
Average	2.44	1009.4	8.93	225.5	2018.03	307.1	67.05	37.6
Maximum	6.35	1357.2	14.74	256.5	2279.83	418.4	96.5	50.7

Author contributions MRD: conceptualization, software, writing—review and editing, validation, data curation, formal analysis, visualization. AR: review and editing, validation.

Funding All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

Abbas AK, Alameedy U, Alsaba M, Rushdi S (2018) Wellbore trajectory optimization using rate of penetration and wellbore stability analysis. In: SPE international heavy oil conference and exhibition, Kuwait City, Kuwait, December 2018. https://doi.org/10.2118/193755-MS

Afsari M, Amani M, Razmgir SAM, Karimi H, Yousefi S (2010) Using drilling and logging data for developing 1d mechanical earth model for a mature oil field to predict and mitigate wellbore stability challenges. In: International oil and gas conference and exhibition in China, Beijing, China, June 2010. https://doi.org/10.2118/132187-MS

Al-AbdulJabbar A, Elkatatny S, Mahmoud M, Abdelgawad K, Al-Majed A (2019) A robust rate of penetration model for carbonate formation. J Energy Resour Technol 141(4):042903

Alali AM, Abughaban MF, Aman BM, Ravela S (2021) Hybrid data driven drilling and rate of penetration optimization. J Petrol Sci Eng 200:108075

Alavi M (2004) Regional stratigraphy of the Zagros fold-thrust belt of Iran and its proforeland evolution. Am J Sci 304(1):1–20

Al-Betairi EA, Moussa MM, Al-Otaibi S (1988) Multiple regression approach to optimize drilling operations in the Arabian Gulf area. SPE Drill Eng 3(01):83–88

Anemangely M, Ramezanzadeh A, Tokhmechi B, Molaghab A, Mohammadian A (2018) Development of a new rock drillability index for oil and gas reservoir rocks using punch penetration test. J Pet Sci Eng 166:131–145

Bezminabadi SN, Ramezanzadeh A, Jalali S-ME, Tokhmechi B, Roustaei A (2017) Effect of rock properties on ROP modeling using statistical and intelligent methods: a case study of an oil well in southwest of Iran. Arch Min Sci 62(1):131–144

Bilgesu H, Tetrick L, Altmis U, Mohaghegh S, Ameri S (1997) A new approach for the prediction of rate of penetration (ROP) values. In: SPE eastern regional meeting, Lexington, Kentucky, October 1997. https://doi.org/10.2118/39231-MS

Bourgoyne AT Jr, Young F Jr (1974) A multiple regression approach to optimal drilling and abnormal pressure detection. Soc Pet Eng J 14(04):371–384

Calhoun W, Ewy R (2005) New confined compressive strength calculation improves bit selection and bit performance. In: Proceeding of the AADE national technical conference and exhibition. Houston, Texas, April 2005

Castillo E, Guijarro-Berdinas B, Fontenla-Romero O, Alonso-Betanzos A, Bengio Y (2006) A very fast learning method for neural networks based on sensitivity analysis. J Mach Learn Res 7(7):1159–1182

Chamkalani A, Zendehboudi S, Amani M, Chamkalani R, James L, Dusseault M (2017) Pattern recognition insight into drilling optimization of shaly formations. J Pet Sci Eng 156:322–339

Coello CC, Lechuga MS (2002) MOPSO: a proposal for multiple objective particle swarm optimization. In: Proceedings of the 2002 congress on evolutionary computation. CEC'02 (Cat. No. 02TH8600), vol 2, pp 1051–1056

Coello CAC (2011) An introduction to multi-objective particle swarm optimizers. In: Soft computing in industrial applications. Springer, pp 3–12

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197

Delavar MR (2021) Hybrid machine learning approaches for classification and detection of fractures in carbonate reservoir. J Pet Sci Eng 208:109327

Delavar MR, Ramezanzadeh A (2023) Pore pressure prediction by empirical and machine learning methods using conventional and drilling logs in carbonate rocks. Rock Mech Rock Eng 56(1):535–564

Delavar MR, Ramezanzadeh A, Tokhmechi B (2021) An investigation into the effect of geomechanical properties of reservoir rock on drilling parameters—a case study. Arab J Geosci 14(17):1763. https://doi.org/10.1007/s12517-021-08168-y

Delavar MR, Ramezanzadeh A, Gholami R, Sanei M (2023) Optimization of drilling parameters using combined multi-objective method and presenting a practical factor. Comput Geosci 175:105359

Ding S, Zhu Z, Zhang X (2017) An overview on semi-supervised support vector machine. Neural Comput Appl 28(5):969–978

Dong S, Zeng L, Lyu W, Xu C, Liu J, Mao Z, Tian H, Sun F (2020) Fracture identification by semi-supervised learning using conventional logs in tight sandstones of Ordos Basin, China. J Nat Gas Sci Eng 76:103131

Elkatatny S (2019) Development of a new rate of penetration model using self-adaptive differential evolution-artificial neural network. Arab J Geosci 12(2):19

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874

Galle E, Woods H (1963) Best constant weight and rotary speed for rotary rock bits. In: Drilling and production practice, January 1963

Gholami R, Moradzadeh A, Rasouli V, Hanachi J (2014) Practical application of failure criteria in determining safe mud weight windows in drilling operations. J Rock Mech Geotech Eng 6(1):13–25

Hamdani TM, Won J-M, Alimi AM, Karray F (2007) Multi-objective feature selection with NSGA II. In: International conference on adaptive and natural computing algorithms. Springer, Berlin, pp 240–247

Han J, Sun Y, Zhang S (2019) A data driven approach of ROP prediction and drilling performance estimation. In: International petroleum technology conference

Hareland G, Hoberock L (1993) Use of drilling parameters to predict in-situ stress bounds. In: SPE/IADC drilling conference, Feb 1993
Hegde C, Gray K (2018) Evaluation of coupled machine learning models for drilling optimization. J Nat Gas Sci Eng 56:397–407
Hegde C, Daigle H, Millwater H, Gray K (2017) Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. J Pet Sci Eng 159:295–306
Kidambi T, Kumar GS (2016) Mechanical earth modeling for a vertical well drilled in a naturally fractured tight carbonate gas reservoir in the Persian Gulf. J Pet Sci Eng 141:38–51
Li T, Wang R, Wang Z, Zhao M, Li L (2018) Prediction of fracture density using genetic algorithm support vector machine based on acoustic logging data. Geophysics 83(2):D49–D60
Liu T, Gao X, Wang L (2015) Multi-objective optimization method using an improved NSGA-II algorithm for oil–gas production process. J Taiwan Inst Chem Eng 57:42–53
Maurer W (1962) The "perfect-cleaning" theory of rotary drilling. J Pet Technol 14(11):1270–1271,1274
Mehrad M, Bajolvand M, Ramezanzadeh A, Neycharan JG (2020) Developing a new rigorous drilling rate prediction model using a machine learning technique. J Pet Sci Eng 192:107338
Melacci S, Belkin M (2011) Laplacian support vector machines trained in the primal. J Mach Learn Res 12(3):1149–1184
Mirjalili S, Saremi S, Mirjalili SM, Coelho LdS (2016) Multi-objective Gray wolf optimizer: a novel algorithm for multi-criterion optimization. Expert Syst Appl 47:106–119
Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6(4):525–533
Motahhari HR, Hareland G, Nygaard R, Bond B (2009) Method of optimizing motor and bit performance for maximum ROP. J Can Pet Technol 48(06):44–49
Mousavi R, Eftekhari M (2015) A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches. Appl Soft Comput 37:652–666
Oyedere M, Gray K (2020a) ROP and TOB optimization using machine learning classification algorithms. J Nat Gas Sci Eng 77:103230
Oyedere M, Gray K (2020b) Torque-on-bit (TOB) prediction and optimization using machine learning algorithms. J Nat Gas Sci Eng 84:103623
Plumb R (1994) Influence of composition and texture on the failure properties of clastic rocks. In: Rock mechanics in petroleum engineering, Delft, Netherlands, August 1994. https://doi.org/10.2118/28022-MS
Rampersad P, Hareland G, Boonyapaluk P (1994) Drilling optimization using drilling data and available technology. In: SPE Latin America/Caribbean petroleum engineering conference, Buenos Aires, Argentina, April 1994. https://doi.org/10.2118/27034-MS
Saeed MS, Mustafa MW, Sheikh UU, Jumani TA, Mirjat NH (2019) Ensemble bagged tree based classification for reducing non-technical losses in Multan electric power company of Pakistan. Electronics 8(8):860
Sanei M, Ramezanzadeh A, Delavar MR (2023) Applied machine learning-based models for predicting the geomechanical parameters using logging data. J Pet Explor Prod Technol 13:1–23
Shahri M, Wilson T, Thetford T, Nelson B, Behounek M, Ambrus A, D'Angelo J, Ashok P (2018) Implementation of a fully automated real-time torque and drag model for improving drilling performance: case study. In: SPE annual technical conference and exhibition, Dallas, Texas, USA, September 2018. https://doi.org/10.2118/191426-MS
Shirkavand F, Hareland G, Aadnoy B (2009) Rock mechanical modeling for a underbalanced drilling rate of penetration prediction. In: 43rd US rock mechanics symposium & 4th US–Canada rock mechanics symposium, Asheville, North Carolina, June 2009
Singh U, Singh SN (2017) Optimal feature selection via NSGA-II for power quality disturbances classification. IEEE Trans Ind Inf 14(7):2994–3002
Tatar A, Yassin MR, Rezaee M, Aghajafari AH, Shokrollahi A (2014) Applying a robust solution based on expert systems and GA evolutionary algorithm for prognosticating residual gas saturation in water drive gas reservoirs. J Nat Gas Sci Eng 21:79–94
Teale R (1965) The concept of specific energy in rock drilling. Int J Rock Mech Min Sci Geomech Abstr 2(1):57–73
Telmadarreie A, Shadizadeh S, Alizadeh B (2012) Research note: investigation of hydrogen sulfide oil pollution source: asmari oil reservoir of marun oil field in the Southwest of Iran. Iran J Chem Eng 9(3):63–74
Warren T (1984) Factors affecting torque for a roller cone bit. J Pet Technol 36(09):1500–1501,1508
Warren T (1987) Penetration rate performance of roller cone bits. SPE Drill Eng. https://doi.org/10.2118/13259-PA,2(01),9-18
Zoback MD (2007) Reservoir geomechanics: Earth stress and rock mechanics applied to exploration. In: Production and wellbore stability, p 449