Article

Application of Machine Learning Methods in Modeling the Loss of Circulation Rate while Drilling Operation

Ahmed Alsaihati, Mahmoud Abughaban, Salaheldin Elkatatny,* and Dhafer Al Shehri

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Fluid losses into formations are a common operational issue that is frequently encountered when drilling across naturally or induced fractured formations. This could pose significant operational risks, such as well control, stuck pipe, and wellbore instability, which, in turn, lead to an increase in well time and cost. This research aims to use and evaluate different machine learning techniques, namely, support vector machine (SVM), random forest (RF), and K nearest neighbor (K-NN) in predicting the loss of circulation rate (LCR) while drilling using solely mechanical surface parameters and interpretation of the active pit volume readings. Actual field data of seven wells, which had suffered partial or severe loss of circulation, were used to build predictive models with an 80:20 training-to-test data ratio, while Well No. 8 was used to compare the performance of the developed models. Different performance metrics were used to evaluate the performance of the developed models. The root-mean-square error (RMSE) and correlation coefficient (R) were used to evaluate the performance of the models in predicting the LCR while drilling. The results showed that K-NN outperformed other models in predicting the LCR in Well No. 8 with an R of 0.90 and an RMSE of 0.17.
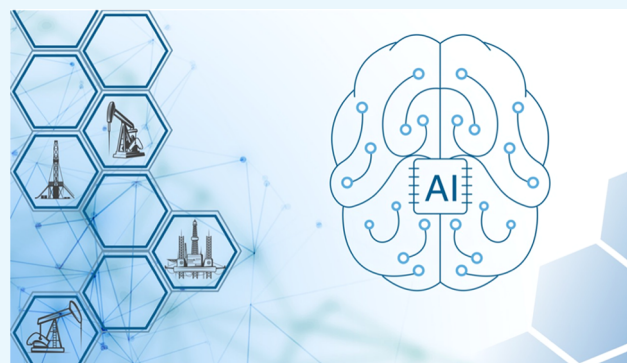
1. INTRODUCTION

Loss of circulation can be defined as the partial or complete loss of the drilling fluid into the formation at any depth during drilling, circulation, or while running tubular. Loss of circulation events is a common occurrence during drilling activities and has subsequent consequences, such as an increase in the overall well cost and nonproductive time, sticking of the drill string, and well control incidents.[1] Loss of circulation is typically caused by a pressure imbalance, wherein the bottom-hole pressure is higher than the formation pore pressure, which results in losing the drilling fluid into the induced fractures.[2]

1.1. Traditional Methods of Monitoring Lost of Circulation while Drilling. A paddle flowmeter, which is installed in the drilling fluid return line, is the most accustomed outflow sensor on drilling rigs.[3] It provides a qualitative measurement that needs additional indicator support before the drilling crew positively confirms loss of circulation occurrences. It enables the drilling crew to know the position of the paddle in a relation to the full opening. The paddle flowmeter has some disadvantages as follows:[3]

- It does not consider the velocity of the drilling fluid that passes through, which makes it impossible to measure the flow rate accurately.

- The constant oscillations of the drilling fluid inside the return line can induce errors in the reading of the paddle

flowmeter. Hence, its accuracy is very poor compared to more advanced techniques.

Monitoring the drilling fluid level in the mud pits by a pit volume totalizer system is another technique used to identify loss of circulation occurrences while drilling. It is a device that collects the level of drilling fluid data inside mud pits to alert the drilling crew if an abnormality has occurred. The fluid level inside the mud pits is usually measured using different types of sensors. It can either be measured using a floating sensor or an acoustic reflector that is located inside the active mud pits.[4] Most drilling contractors utilize a float that travels along a pole that activates a sequence of reed switches, where the drilling fluid level is related to the electrical signal produced by the status of these magnetic switches. On the other hand, some drilling contractors measure the differences in resistance with a potentiometer that is connected to the buoyant float through a chain mounted on a wheel.[5] The ultrasound sensor is another type of sensor that is used to measure the drilling fluid level in

Table 1. Applications of AI Related to the Loss of Circulation Problems

| refs	model	input variables	output variable	data	statistical metric
Jeirani and Mohebbi[18]	ANN	pressure drop across filter cake and water/NACL weight percent	permeability of filter cake; filtrate volume	not available	$R^2 = 0.94$; $R^2 = 0.98$
Moazzeni et al.[19]	ANN	current depth of the well related to the ground elevation; current depth of the well related to the sea level, drilled depth, time of drilling, length of open hole, formation's top depth, the north and east coordinates of the well, hole size, pumping rate, pump pressure, mud density, solid content percentage, viscosity readings, fluid loss, and the mud losses before the considered day	volume of losses; severity of losses (seepage, severe, complete)	32 wells from an oil field	$R = 0.944$; $R = 0.982$
Jahanbakhshi et al.[20]	ANN	Young's modulus, fracture orientation, tensile strength, unconfined compressive strength, minimum horizontal stress, API fluid loss, mud filtrate viscosity, solids percent, mud gel strength, plastic viscosity, yield point, temperature across the loss of circulation zone, pump pressure, drilling speed, equivalent circulation density, porosity, formation permeability, the differential pressure between formation and wellbore hydrostatic pressure, and hole depth	volume of losses	not available	$R^2 = 0.94$
Jahanbakhshi and Keshavarzi[21]	SVM	daily drilling reports, geological information, and geomechanical properties	volume of losses	260 data sets	$R = 0.985$
Toreifi et al.[22]	multilayer perceptron	east and north coordinates of the well, current depth, formation's tip angle, drilling speed, formation type, annulus capacity, pump pressure, hydrostatic pressure, pumping rate, filter cake viscosity, plastic viscosity, and yield point	volume of losses	1756 data points collected from 38 wells	$R = 0.94$
	particle swarm optimization algorithm		severity of losses		$R = 0.98$
Manshad et al.[23]	SVM RBF	north and east coordinates, loss volume one day prior to the day of interest, and loss volume during the two days prior to the day of interest	volume of losses	30 wells	not available
Far and Hosseini[24]	ANN	pumping rate, pump pressure, and mud weight	volume of losses; genetic algorithm was used to minimize the severity of losses	not available	$R^2 = 0.99$
Solomon et al.[25]	ANN	Poisson's ratio, horizontal stress, Young's modulus, well depth, and well pressure	width of the induced fracture	30 data points	$R^2 = 0.79$
Li et al.[26]	RF	depth, drilling speed, circulating pressure, pumping rate, mud weight, plastic viscosity, yield point, gel strength, API filtration, lithology, pore pressure, stresses	loss of circulation occurrence (losses, no losses)	6976 data points collected from an oil field	accuracy of predicting points with losses correctly= 56%
Abbas et al.[27]	ANN	lithology, mud weight, pumping rate, drilling speed, circulating pressure, inclination, solids content, fluid loss, pipe rotation, weight exerted on the drilling bit, yield point, plastic viscosity, marsh funnel viscosity, gel strength, azimuth, measured depth, and hole size	corrective treatment for curing losses in vertical and deviated wells	385 wells	$R^2 = 0.95$
Abbas et al.[28]	SVM	lithology, mud weight, pumping rate, drilling speed, circulating pressure, inclination, solids content, fluid loss, pipe rotation, weight exerted on the drilling bit, yield point, plastic viscosity, marsh funnel viscosity, gel strength, azimuth, measured depth, and hole size	likelihood of lost circulation occurrences	385 wells	accuracy = 0.91%
Shi et al.[29]	SVM	inlet flow, outlet flow, annuals pressure, annulus temperature, hook load, well depth, bit depth, pit volume, pumping pressure, pipe rotation, fluid outlet density, and fluid outlet temperature	type of event (influx, losses, normal)	4 wells	accuracy = 93.72%
Ahmed et al.[30]	ANN	pumping rate, pump pressure, weight exerted on the drilling bit, rotary torque, and drilling speed	loss of circulation zones	3 wells	$R = 0.99$
Hou et al.[31]	ANN	mud weight, yield point, solid content%, plastic viscosity, pumping rate, drilling speed, weight exerted on the drilling bit, pumping pressure, nozzles flow area, measured depth, lithology, fracture and pore pressure, and unconfined compressive strength	type of losses	50 wells	F_1 score = 0.9
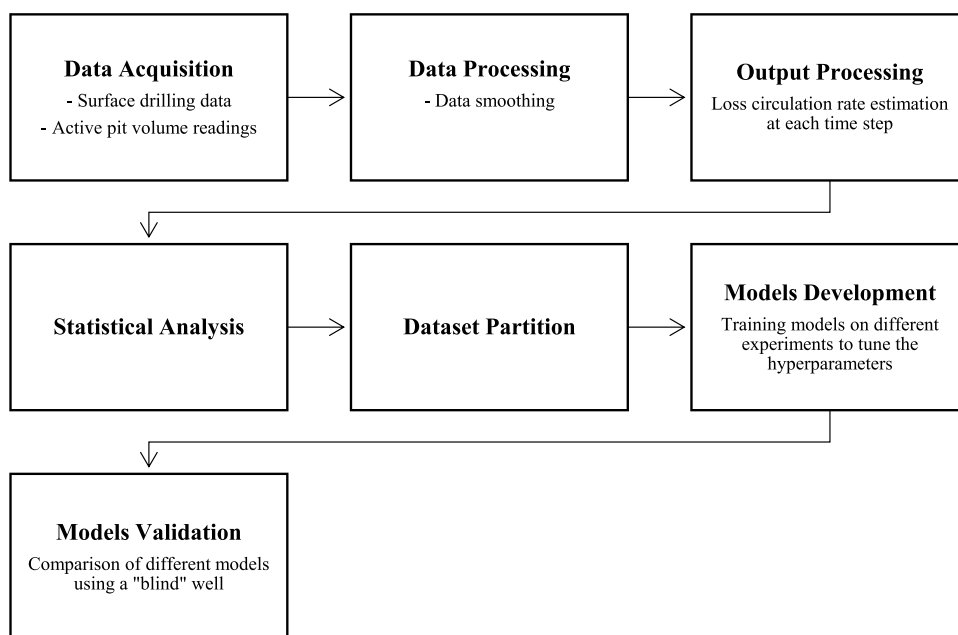
Figure 1. Methodology flowchart.

the mud pits. It emits a series of ultrasonic pulses that are reflected off the fluid surface and received by the same sensor. The transit time is proportional to the distance, which enables the system to calculate the drilling fluid levels in the active or reserve pit tanks. Once the reduction of the drilling fluid level is identified by the pit volume totalizer system during the loss of circulation events, the drilling crew can manually calculate the loss rate by dividing the reduction in volume over the time of observation. The pit volume totalizer system has some disadvantages as follows:

- It has a borderline detectable fluid volume below which it would be difficult for the drilling crew to identify losses. The alarm is typically set by the drilling crew at 5 bbl in most rigs.
- It requires manual calculation for estimating the loss rate if the rig is not equipped with advanced mud-logging instruments. Thus, it does not provide a continuous reading of loss rates while drilling activities.
- It is not installed in old drilling rigs, where a measuring stick is still used to allocate the fluid level inside mud pits to calculate the loss rate.

The aforementioned limitations of the current methods used to monitor or calculate the loss of circulation hinder the drilling crew to have a continuous profile of the loss of circulation that helps in finding the optimum combined drilling parameters to reduce the intensity of losses. Thus, we introduced a new machine learning approach based on surface drilling parameters to estimate the loss rate while drilling and allow the drilling crew to alter these parameters—if possible—to control the loss rate.

1.2. Application of Artificial Intelligence and Machine Learning Related to Drilling Applications. Most artificial intelligence (AI) and machine learning (ML) techniques can potentially solve practical problems by learning from large historical data sets, something that conventional analytical models cannot do.[6,7] The applications of AI in drilling operation activities have evolved over recent years due to their flexibility in classification, optimization, prediction, and selection.[8] These applications include, but are not limited to, identification of

formation lithology,[9] estimation of pore and fracture pressures during the drilling operation,[10,11] real-time prediction of drilling fluid properties,[12] formation identification while drilling using mechanical surface parameters,[13] early warning signs detection while drilling horizontal wells,[14] use of an Internet-of-things (IoT) environment integrated with cameras and high-computation edge server to implement a deep learning model for proper drill string space out when a well control incident occurs during drilling,[15] employing of raw drilling data to estimate the drilling bit- wear in real time using a bidirectional long short-term memory-based variational autoencoder,[16] and determination of downhole vibrations while drilling surface hole sections to mitigate premature drill string failures.[17]

Table 1 also sheds light on some works on AI applications related to the loss of circulation problems with their accuracy in terms of the correlation coefficient (R) and coefficient of determination (R^2).

The aforementioned studies showed that AI and ML were used to either detect losses events (losses or no losses) or estimate the volume of the losses. Apart from Shi et al.,[29] most of the previous studies used *static* historical data, i.e., the volume of losses or loss rates that were documented in the daily drilling reports, to build a predictive model. Shi et al.[29] used the real-time readings of the inlet flow rate and the outlet flow rate to label each data point as losses or gain. It should be mentioned that the flowmeter is not commonly available at the rig site due to the high cost in our region. The previous studies also used some input parameters that are difficult to obtain and require laboratory or logging measurements such as Young's modulus, fracture orientation, tensile strength, unconfined compressive strength, minimum horizontal stress, pore pressure, formation's top angle, Poisson's ratio, and unconfined compressive strength. Instead, our approach used data that is always available at the rig site, i.e., the sensor data readings of the APV, to estimate the loss of circulation rate (LCR) at each time span, which was linked later with the drilling surface parameters to develop a predictive model.

The new approach shows potential for continuous monitoring of the LCR during the drilling operation, which assists the

Table 2. Dummy Example of How to Identify Loss Occurrence at Each Time Step

time (min.)	depth (ft.)	Q (gal/min)	SPP (psi)	WOB (kIbf)	ROP (ft/h)	RS (RPM)	T (kIbf-ft)	APV (bbl.)	APV_a (bbl.)	APV_{TH} (bbl.)	ΔAPV (bbl.)	LCR (bbl/min)
0	0	###	###	###	###	###	###	600				
1	10	###	###	###	###	###	###	597	3	3[a]	0	0
2	20	###	###	###	###	###	###	592	5	2[a]	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
100	1000	###	###	###	###	###	###	420	7	7	0	0

[a]These values are assumed, not calculated, in this particular example.

Table 3. Dummy Modified Data Set with the New Continuous Variable

Q (gal/min)	SPP (psi)	WOB (kIbf)	ROP (ft/h)	RS (RPM)	T (kIbf-ft)	LCR (bbl/min)
###	###	###	###	###	###	
###	###	###	###	###	###	0
###	###	###	###	###	###	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
###	###	###	###	###	###	0

drilling crews to alter the drilling surface parameters to minimize losses, detect small amounts of losses, and take corrective actions before losses are escalated. Furthermore, it will aid drilling engineers to design a predrill model with the optimum surface drilling parameters, which are learned from offset wells, before drilling a specific section to lessen drilling fluid losses into formations.

2. METHODS

Figure 1 shows a flowchart that displays the technical approach to reaching our objective.

2.1. Data Collection. This study used field data from an intermediate open-hole section of seven vertical wells. The intermediate section consisted of dolomite/limestone formations that caused partial or severe loss of circulation due to pore pressure depletion and natural fractures. The initial data set consisted of operational surface drilling parameters, which included flow rate (Q), standpipe pressure (SPP), weight on bit (WOB), rate of penetration (ROP), rotary speed (RS), and surface torque (T), and raw data of the active pit volume (APV) during the drilling operation, which correspond to the drilling parameters. The operational surface parameters were collected from surface real-time transmitter sensors, while the raw data of the APV were obtained from the pit volume totalizer system.

2.2. Data Processing. Noise in a data set can significantly impact the predictability accuracy of any machine learning model.[32] Therefore, it is essential to use an established technique of the industry, known as smoothing, to reveal the underlying trend in the original data set. An exponential moving average technique was applied to the original data set with a damping factor of 0.9, which is used most often in time-series data.[33−35] The exponential moving average technique was applied using eq 1.

$$Y_t = \beta X_{t-1} + (1-\beta)Y_{t-1} \tag{1}$$

where Y_t is the smoothed observation of any variable at time t, β is a damping factor that varies between 0 and 1, X_{t-1} is the unsmoothed observation of the variable at the previous time $t-1$, and Y_{t-1} is the smoothed observation of the variable at the previous time $t-1$.

2.3. Output Processing. Three important quantities were calculated to be able to estimate the loss of circulation rate (LCR) while drilling at each time step in the original data set: the actual reduction of APV, the theoretical reduction of APV, which is defined as the drilling fluid volume needed to drop from the active mud pit to fill up the new drilled footage, and the difference between the actual and theoretical reduction. These quantities can be calculated using eqs 2, 3, and 4

$$APV_a = APV_{t-1} - APV_t \tag{2}$$

where APV_a is the actual reduction of active pit volume (bbl.), APV_{t-1} is the active pit volume reading at the previous time $t-1$ (bbl.), and APV_t is the active pit volume reading at the current time t (bbl.).

$$APV_{TH} = \left[\frac{OH^2 - (D_o^2 - D_i^2)}{1029.4} \times (\Delta D) \right] \tag{3}$$

where APV_{TH} is the theoretical reduction of active pit volume (bbl.), OH is the open-hole size (in.), D_o is the outer diameter of the bottom-hole assembly (in.), D_i is the inner diameter of the bottom-hole assembly, and ΔD is the difference between the depth at time t and the depth at time $t-1$ (ft.).

$$\Delta APV = APV_a - APV_{TH} \tag{4}$$

where ΔAPV is the difference between the actual and theoretical reduction (bbl.).

At each time step, the LCR was calculated if the difference between the actual reduction and theoretical reduction is greater than zero, and if the difference is equal to zero, the LCR was set to be equal to zero. On the other hand, if the difference is less than zero, the data point was removed from the data set, which indicates the need for either a treatment for the drilling fluid or the addition of further mud volume by drilling crews. In this study, a threshold of 0.5 bbl. was used to account for the mud loss over the surface solids control equipment at the rig site and for filtrate volume due to being overbalanced. A different threshold can be selected by the user depending on the efficiency of the solids control equipment and the condition of the drilling fluid used to drill with. The LCR was calculated using eq 5.

Table 4. Descriptive Statistics of the Training Set (11,022 Data Points)

statistical parameters	Q (gal/min)	ROP (ft/h)	RS (RPM)	SPP (psi)	T (kIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	466.9	4.9	35.7	566.6	3.3	3.5	0.0
max	1105.7	149.9	186.7	3964.2	25.3	65.2	12.9
range	638.8	145.0	151.0	3397.6	22.1	61.6	12.9
mean	832.5	44.1	118.7	2184.3	15.1	39.2	1.0

Table 5. Descriptive Statistics of the Testing Set (2872 Data Points)

statistical parameters	Q (gal/min)	ROP (ft/h)	RS (RPM)	SPP (psi)	T (kIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	466.6	5.0	39.4	577.9	3.4	4.6	0.0
max	1104.6	149.6	186.3	3963.5	25.0	65.1	12.7
range	638.0	144.6	146.9	3385.6	22.0	60.5	12.7
mean	832.4	43.4	118.6	2180.3	15.0	39.0	0.9

$$LCR = \frac{\Delta APV}{\Delta t} \tag{5}$$

where LCR is the loss of circulation rate (bbl/min) and Δt is the incremental time between two consecutive data points (min.)

Table 2 shows a dummy example of how to determine the LCR for each observation in a data set. The new continuous variable (i.e., LCR) was defined to be able to accomplish our objective, which is building a machine learning model for the prediction of the LCR based on drilling surface parameters. Table 3 shows a dummy modified data set with the input and output variables that were used for building different machine learning models.

2.4. Statistical Analysis. The prepared data set consisted of seven wells with 13,894 data points. Statistical analysis was performed on the prepared data set for each well in the data set. The descriptive statistics for the seven wells used to build a predictive model are presented in Appendix A. The LCR in each well was determined using the approach discussed in Section 2.3.

2.5. Data Set Partition. In this study, the data set was divided with a ratio of 80:20. Eighty percent of the data was selected for training to ensure that the models captured most of the loss of circulation behaviors while drilling. The training set had 11,022 data points, while the testing set had 2872 data points. Tables 4 and 5 present the descriptive statistics for the training and testing sets.

2.6. Relative Importance of Variables. Understanding the influence of input variables on output variables is imperative when developing AI and ML models because including more inputs than required can decrease the learning speed and efficiency. Mutual information (MI), which is a statistical quantity that measures the amount of information between two variables, was used to study the relationship between the input and output variables. The MI determines how much two variables are dependent. MI is a dimensionless quantity and it is always greater than or equal to zero. A high MI value indicates a large reduction in uncertainty, while a low MI value indicates a small reduction in uncertainty; zero MI indicates that the variables are independent.[36] The MI of two jointly discrete variables X and Y can be calculated using eq 6.[37]

$$MI(X, Y) = \int_y \int_x P[X = x, \ Y = y]\cdot$$
$$\log\left(\frac{P[X = x, \ Y = y]}{P[X = x].\ P[Y = y]}\right) dxdy \tag{6}$$

where MI is the mutual information, X is a random variable, Y is a random variable, $P(x)$ is the marginal probability density function of variable X, and $P(y)$ is the marginal probability density function of variable Y.

How all of this can be interpreted? For example, in studying SPP and WOB, the MI values are 0.82 and 0.31, respectively. This means that if SPP is removed as an input variable from our data set, the model will not be able to capture most of the trend of LCR, compared to if WOB is removed. Similarly, other input variables can be interpreted. In this study, we used domain knowledge as well to select the variables needed for modeling the LCR. The surface drilling parameters are very important in determining the intensity of the LCR, and therefore, none of them was removed from the data set. Figure 2 compares the
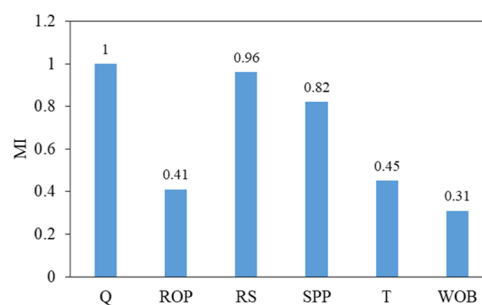
Figure 2. Comparison of the dependency between the input variables and the output.

dependency between the input parameters and the output variable. Figure 2 shows that Q, RS, and SPP have MI values of 1.0, 0.96, and 0.82, respectively, while ROP and T have MI values of 0.41 and 0.45, respectively. On the other hand, WOB has a value of MI of 0.31.

3. MODEL DEVELOPMENT

Three ML models were used to build a model to predict the LCR using mechanical surface parameters, which included Q, ROP, RS, SPP, T, and WOB.

3.1. Support Vector Machine (SVM) Model. SVM is supervised machine learning that arises from the concept of having the maximum margin of separation. Given a data set $\boldsymbol{DT}$ = $[(\boldsymbol{x_1,y_1}), (\boldsymbol{x_2,y_2}),...,(\boldsymbol{x_n,y_n})]$ that is linearly separable (i.e., it is separable with no error by a decision function in the form of $g_{w,b}(\boldsymbol{x}) = sign(f_{w,b}(\boldsymbol{x}))$ *with* $f_{w,b}(\boldsymbol{x}) = \langle \boldsymbol{w,x} \rangle + \boldsymbol{b})$, then the optimal separating hyperplane is defined as the hyperplane that separates the data such that the distance, known as margin, between the hyperplane and the closet point is maximum. The

optimum hyperplane can be obtained as a solution to the following constraints

$$\begin{cases} \min_{w \in \mathbf{R}^d, b \in \mathbf{R}} \frac{1}{2} \|w\|^2, \\ y_i(\langle w, \ x_i \rangle + b) \geq 1, \quad i \in \{1, 2, ..., n\} \end{cases} \quad (7)$$

The above equation is quadratic with linear constraints and its solution is a linear combination of a subset of samples denominated as support vectors, s.v., which are the closest points to the hyperplane.

$$w = \sum_{s.v.} \beta_i x_i \Rightarrow f_{w,b}(x) = \sum_{s.v.} \beta_i \langle x_i, \ x \rangle + b \quad (8)$$

The inner product in a large dimension space can be represented by a positive definite function k (kernel) such that if the transformation is chosen in $\phi: \Omega \subset R^d \to R^r$, then the inner product is defined as[38]

$$\langle \phi(x), \ \phi(x') \rangle = \sum_i \phi_i(x)\phi_i(x') = k(x, \ x') \quad (9)$$

The above expression can now be substituted in eq 7 to resolve the problem as follows

$$w = \sum_{s.v.} \beta_i \phi(x_i) \Rightarrow f_{w,b}(x) = \sum_{s.v.} \beta_i \langle \phi(x_i), \ \phi(x) \rangle + b$$
$$= \sum_{s.v.} \beta_i k(x_i, \ x) + b \quad (10)$$

There are different kernel functions such as linear, radial basis function, and polynomial functions[39] that give rise to different feature spaces verifying the condition in eq 9. In practice, it is not advisable to have a perfect separation, therefore, it is necessary to have misclassified observations (i.e., soft margin). This can be achieved by adding a slack variable $\xi_i \geq 0$, $i = 1,2,3,...,n$

$$y_i(\langle w, \ \phi(x_i) \rangle + b) + \xi_i \geq 1 \quad (11)$$

The interest here is to minimize the misclassified observations and have a robust model. This can be expressed mathematically as[40]

$$\min_{w,b,\xi} \frac{1}{2} \left\{ \|w\|^2 + C \sum_{i=1}^n \xi_i \right\}$$
$$\begin{cases} y_i(\langle w, \ \phi(x_i) \rangle, \ b) \geq 1 - \xi_i \\ \xi_i \geq \end{cases} \bigg| \ i = 1, 2, 3, ..., n \quad (12)$$

Python's Scikit-Learn was used to build an SVM model for the prediction of the LCR. Different kernel functions (i.e., linear, RBF, polynomial) were applied to map the original data set onto a higher-dimensional space to become separable. The regularization parameter C is an important parameter in an SVM model used to control the variance−bias tradeoff, i.e., to have almost similar accuracy in the training and testing sets, and to generalize well on unseen data. The performance of the SVM classifier with different C values using different kernel functions was evaluated. In the case of a polynomial kernel function, different degrees from 3 to 10 were assessed.

The γ parameter, which signifies the influence of points either near or far away from the hyperplane, was also tuned using the GridSearchCV function. GridSearchCV is a built-in function in Python that helps to iterate through predefined hyperparameters

and fit a model on a training set. Two options for γ in the SVM classifier (i.e., scale or auto) were evaluated.

3.2. Random Forest (RF) Model. RF is an ensemble learning model that can be used to solve classification and regression problems.[41] RF operates by constructing hundreds or thousands of decision trees and training each one on a different set of observations. Formally, for a random vector $x = (x_1, x_2,..., x_p)^T$ representing the independent variables and a random variable y representing the independent variable, we assume unknown joint distribution, i.e., $P_{xy}(x,y)$. The function that predicts y is determined by a loss function, $L(y,f(x))$, and defined to minimize the expected value of the loss function.

$$E_{xy}(L(y, f(x))) \quad (13)$$

$L(y,f(x))$ measures how close $f(x)$ is to y, and it penalizes values of $f(x)$ that are a long way from y. The loss function for regression is the squared error loss and zero-one loss for classification.

$$L(y, f(x)) =$$
$$\left| \begin{array}{ll} (y - f(x))^2, & \text{regression} \\ I(y \neq f(x)) = \begin{cases} 0 \text{ if } y = f(x), \\ 1 \text{ if } y \neq f(x) \end{cases} & \text{classification} \end{array} \right| \quad (14)$$

Minimizing $E_{xy}(L(y,f(x)))$ for a regression task gives the conditional expectation

$$f(x) = E(y|X = x) \quad (15)$$

If we set the possible value of y denoted by y', minimizing $E_{xy}(L(y,f(x)))$ for a classification task gives the conditional expectation

$$f(x) = \underset{y \in y}{\mathrm{argmax}} \ P(Y = y|X = x) \quad (16)$$

The ensembles construct f in terms of a collection of base learners $h_1(x), h_2(x),..., h_j(x)$ that are combined to give predictor $f(x)$. In a regression task, the base learners are averaged

$$f(x) = \frac{1}{J} \sum_{j=1}^J h_j(x) \quad (17)$$

For a classification task, the final output of the RF is the most frequently predicted class

$$f(x) = \underset{y \in y}{\mathrm{argmax}} \sum_{j=1}^J I(y = h_j(x)) \quad (18)$$

The jth base learner for an RF is a decision tree. RF outperforms a single decision tree because of its ability to limit overfitting without substantially increasing the margin of error.[41,42]

Python's Scikit-Learn was used to build an RF model. The hyperparameters in an RF are either used to enhance the predictive power or to make the model run quicker. The number of estimators is a hyperparameter; this is the number of trees that an RF builds before computing the average of predictions. The number of estimators was tuned from 1 to 150 using the GridSearchCV function to find the optimal value. A high number of estimators enhances the performance of the model

and makes the prediction more stable, but it slows down the process of computation.

The maximum features is another hyperparameter; this is the number of features to be considered to split a node in each decision tree. If the maximum features is "sqrt", then the number of features to be considered is the square root of the number of input variables in a data set. If the maximum features is "Log$_2$", then the number of features to be considered is the base-2 logarithm of the number of input variables in a data set. Both maximum feature types were tried during the construction of the RF to find the optimum performance. The maximum depth is another important hyperparameter; this represents the depth of each tree in a forest. In practice, deep trees can capture more information about the data set, but they can cause model overfitting. Therefore, the maximum depth, in this case, was tuned from 1 to 19 using the GridSearchCV function to find the optimum value.

3.3. K Nearest Neighbor Model. K-NN is instance-based learning that, instead of performing an explicit generalization, classifies new instances based on a direct comparison and similarity to known training instances, which have been stored in a memory. Instance-based methods assume a function for determining the similarity or the distance between any two instances. Giving a training set $DT = [(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)]$, K-NN starts by calculating the distance between a test sample and the training samples. Different distance metrics can be used to calculate the distance between samples.[43] For continuous feature vectors, the Euclidean or Manhattan distance is the generic choice that can be calculated using eqs 19 and 20.

$$d_e(x_i, x_j) = \sqrt{\sum_{p=1}^{n} (a_p(x_i) - a_p(x_j))^2}$$

(19)

$$d_m(x_i, x_j) = \sum_{p=1}^{n} |x_i - x_j|$$

(20)

The output of the test sample in a classification problem is a label that is the most frequent among the K nearest training samples or the average of the K nearest training samples.[44] The number of neighbors, i.e., K, is a user-defined hyperparameter that can be selected using heuristic techniques.

When the class distribution of a data set is highly skewed (right/left) where training samples of one class dominate the prediction of a new sample because of their large number, a downside of the basic majority of voting occurs.[45] To resolve this problem, the distance from the test point to each of its nearest neighbors is considered, where the class label or the value in regression problems of each of the K nearest points is multiplied by a weight proportional to the inverse of the distance from that point to the test point.

The K-NN algorithm is considered to be one of the simplest machine learning algorithms, and it is effective when the training data set is large. However, while there are advantages, K-NN has some disadvantages such as the need to determine the value of K, which can be complex, and the high computational cost because of calculating the distances between new test samples to all of the training samples.

Python's Scikit-Learn was used to build a K-NN model. Different combinations of numbers of K and distance metrics (i.e., Euclidean and Manhattan) were used to evaluate the performance of the model.

4. VALIDATION OF THE DEVELOPED MODELS USING WELL NO. 8

Well No. 8, with more than 1123 unseen data points, was used to compare the capability of the developed models in predicting the LCR. The descriptive statistics for the data set of Well No. 8 is presented in Appendix A.

The evaluation of models is an important component in ML projects that aims to estimate the ability of a developed model to generalize on future unseen data sets. Model evaluation metrics are required to quantitively measure the performance of a developed model. The selection of evaluation metrics depends on the type of machine learning problems such as classification, regression, clustering, etc. The models' performance is evaluated using the root-mean-square error (RMSE) and R. The RMSE is one way of quantifying the difference between the values implied by the predicted and actual values. The RMSE values may indicate how well the model output values fit the desired output values, but it is often useful to investigate the model performance by calculating R. The RMSE and R can be calculated using eqs 21 and 21, respectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}$$

(21)

$$R = \frac{n(\sum_1^n Y_i \hat{Y}_i) - (\sum_1^n Y_i)(\sum_1^n \hat{Y}_i)}{\sqrt{\left[n \sum_1^n Y_i^2 - (\sum_1^n Y_i)^2\right]\left[n \sum_1^n \hat{Y}_i^2 - (\sum_1^n \hat{Y}_i)^2\right]}}$$

(22)

where n is the number of samples in the data set, Y_i is the actual output, and $\hat{Y}_i$ is the predicted output.

5. RESULTS AND DISCUSSION

The SVM model performance with different combinations of C, γ, and kernel function was evaluated. Table 6 shows the results of the SVM when RBF was used as a kernel function with different values of C. Table 6 shows that a value of C equal to 1 and a γ option of "auto" are the best-combined model parameters that had an RMSE and an R of 0.91 and 0.79, respectively, in the training set. On the other hand, the RMSE and R were 1.22 and 0.53, respectively, in the case of the testing set. The model with a

Table 6. Performance of the SVM Model with the RBF Kernel Function

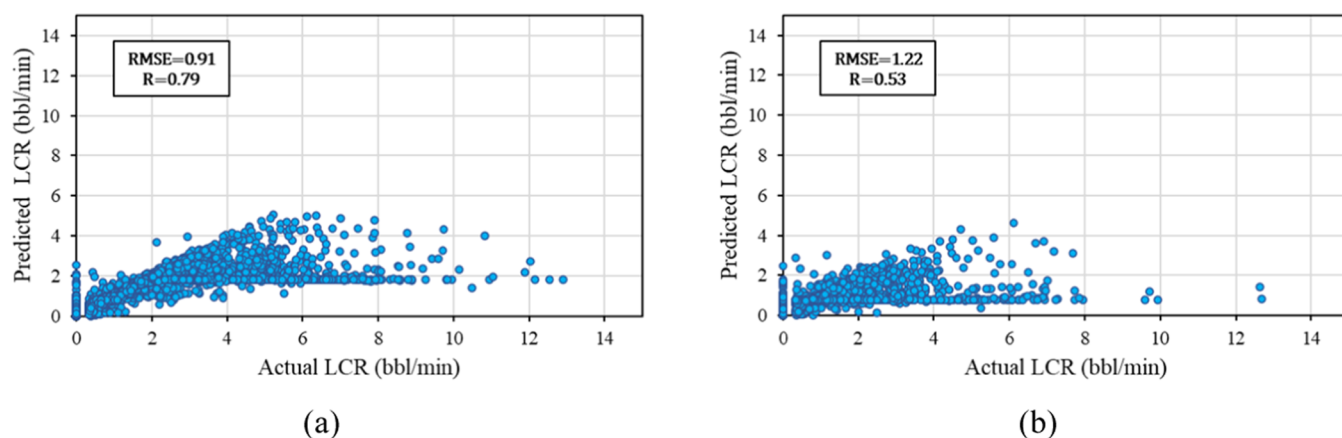| model parameters		training set		testing set	
C	γ	RMSE	R	RMSE	R
0.001	auto	1.62	0.43	1.56	0.42
0.01		1.61	0.42	1.56	0.41
0.1		1.5	0.12	1.49	0.27
[a]1		0.91	0.79	1.22	0.53
10		0.14	0.97	1.1	0.65
100		0.1	0.98	1.11	0.64
1000		0.1	0.98	1.11	0.63
0.001	scale	1.55	0.29	1.49	0.27
0.01		1.46	0.18	1.42	0.18
0.1		1.45	0.22	1.41	0.21
1		1.44	0.25	1.4	0.23
10		1.42	0.29	1.39	0.26
100		1.41	0.31	1.39	0.27
1000		1.42	0.31	1.39	0.25

[a]Best results.

Figure 3. Cross-plots of the actual LCR versus the predicted LCR of the optimum SVM model. (a) Training set. (b) Testing set.

value of C of more than 1 produced a high R in the training set but a low value of R in the testing set, which indicates model overfit. The results showed that the SVM model with linear and polynomial kernel functions did not perform well and therefore was not presented here. Figure 3a,b shows the cross-plots of the actual LCR versus the predicted LCR of the optimum SVM model.

Table 7 shows the performance of the RF model with different maximum depth values in predicting the LCR. The maximum

Table 7. Performance of the RF Model with Different Maximum Depth Values

model parameter	training set		testing set	
maximum depth	RMSE	R	RMSE	R
1	1.4	0.39	1.37	0.36
2	1.36	0.44	1.33	0.41
3	1.31	0.49	1.3	0.45
4	1.26	0.55	1.26	0.49
5	1.21	0.6	1.22	0.54
6	1.15	0.65	1.18	0.59
7	1.09	0.7	1.13	0.63
8	1.03	0.74	1.09	0.67
9	0.96	0.78	1.04	0.71
10	0.88	0.83	0.98	0.75
11	0.81	0.86	0.93	0.78
12	0.75	0.88	0.87	0.81
13[a]	0.67	0.91	0.82	0.84
14	0.58	0.94	0.78	0.86
15	0.54	0.95	0.74	0.87
16	0.46	0.96	0.71	0.88
17	0.42	0.97	0.68	0.89
18	0.37	0.98	0.66	0.9
19	0.34	0.98	0.64	0.9

[a]Best results.

depth of 13 was selected as an optimum value since there was an increase in the difference of the error between the training and testing sets when the maximum depth went beyond it. It is worth mentioning that different options of maximum features ($\log_2$, sqrt) led to similar model performance. The optimum RF model predicted the LCR with an RMSE of 0.67 and an R of 0.91 in the training set, while the RMSE and R were 0.82 and 0.84, respectively, in the testing set. Figure 4a,b shows the cross-plots

of the actual LCR versus the predicted LCR of the optimum RF model.

Table 8 shows the performance of the K-NN model with different K values when Manhattan was used as a distance metric. Table 8 shows that the K-NN model with a K value of 2 achieved its best performance with an RMSE of 0.35 and an R of 0.97 in the training set, while the RMSE and R were 0.69 and 0.94, respectively. Similarly, Table 9 shows the performance of the K-NN model with different values of K when Euclidian was used as a distance metric. Table 9 shows that the K-NN model with a K value of 2 achieved its best performance with an RMSE of 0.37 and an R of 0.97 in the training set, while the RMSE and R were 0.73 and 0.88, respectively. The results show that the K-NN model with the Manhattan distance slightly outperformed the K-NN model with the Euclidian distance. Figure 5a,b shows the cross-plots of the actual LCR versus the predicted LCR of the optimum K-NN model. Table 10 summarizes the optimum parameters for each model.

5.1. Comparison of the Developed Models Using Well No. 8. The K-NN model outperformed the other models with a small RMSE of 0.17 and a high R of 0.90. SVM had a relatively small RMSE of 0.41 but an extremely low value of R of 0.14. RF, on the other hand, had a slightly better value of R of 0.16 compared to the SVM models, but it is still considered to be a low value of R. Figure 6 shows a stack plot of the input variables and the output variable, i.e., LCR, as a function of adjusted well depth in Well No. 8. The adjusted depth was used to hide the actual depth of the well. These results show that the K-NN's performance was very satisfactory, whereas the RF and SVM models were not able to accurately predict the LCR.

5.2. Leverage Approach Implementation. The applicability scope of the best-developed model was studied by applying the Leverage method[46] to assess the validity of the model in the estimation of the LCR. The standard residuals that calculate the deviation of the model predictions from the actual values can be computed as follows[47]

$$Re_j = \frac{e_j}{(\text{MSE}(1 - H_{jj}))^{0.5}} \tag{23}$$

where H_{jj} is the Hat index of the jth data, e_j is the difference between the prediction and actual value of the jth data, and MSE is the mean square error of the model.

The Hat indices can be determined as below[48]
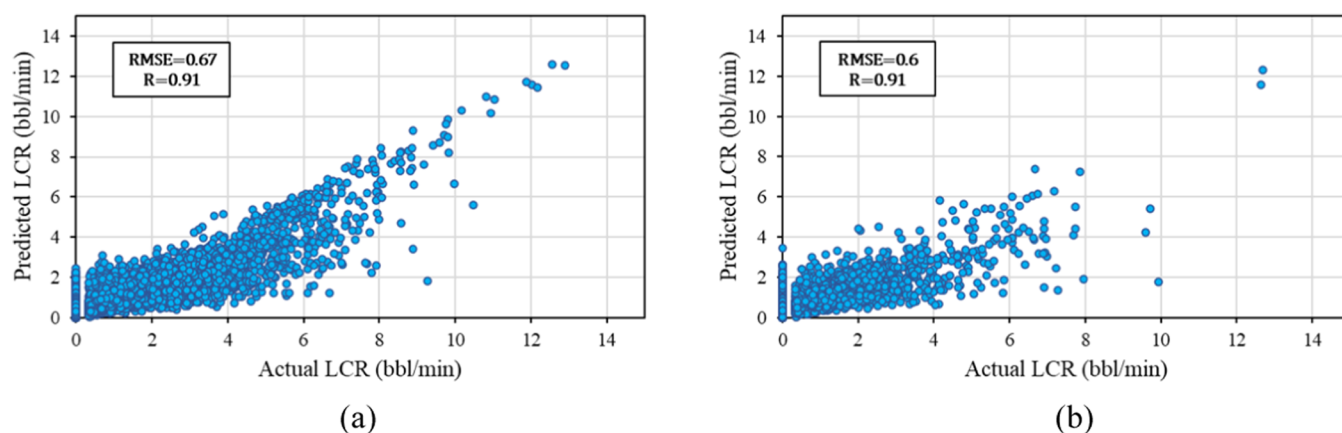
$$H = X(X^T X)^{-1} X^T \tag{24}$$

Figure 4. Cross-plots of the actual LCR versus the predicted LCR of the optimum RF model. (a) Training set. (b) Testing set.

Table 8. Performance of the K-NN Model with the Manhattan Distance

model parameters		training set		testing set	
K	distance	RMSE	R	RMSE	R
2^a		0.35	0.97	0.69	0.94
3		0.44	0.95	0.73	0.88
4		0.52	0.94	0.75	0.87
5		0.57	0.92	0.78	0.85
6		0.62	0.91	0.83	0.83
7		0.66	0.9	0.86	0.82
8	Manhattan	0.7	0.88	0.89	0.8
9		0.74	0.87	0.92	0.79
10		0.77	0.86	0.95	0.77
11		0.8	0.84	0.97	0.76
12		0.82	0.83	1	0.75

^{a}Best results.

Table 9. Performance of the K-NN Model with the Euclidian Distance

model parameters		training set		testing set	
K	distance	RMSE	R	RMSE	R
2		0.37	0.97	0.73	0.88
3		0.47	0.95	0.75	0.87
4		0.54	0.93	0.78	0.86
5		0.61	0.91	0.81	0.84
6		0.65	0.9	0.85	0.82
7		0.69	0.89	0.89	0.8
8		0.73	0.87	0.91	0.79
9		0.77	0.86	0.94	0.78
10	Euclidian	0.8	0.84	0.97	0.76
11		0.83	0.83	0.99	0.75
12		0.85	0.82	1.01	0.74

where X is a matrix of size nxq, X^T is the transpose of matrix X, n is the count of data points, and q is the number of input parameters. The warning Leverage is H^* is calculated as follows

$$H^* = \frac{3(q + 1)}{n} \tag{25}$$

The suspected data for training and validation data sets and the application area of the best-developed model, i.e., K-NN, were identified by plotting William's plot, Figure 7. The data points that have $-3 \leq Re \leq 3$ and $H_{jj} \leq H^*$ are within the application scope of the developed model. Figure 7 shows that the majority of the data are enclosed inside the area of $-3 \leq Re \leq 3$ and $H_{jj} \leq H^*$ which is considered crucial for the validity of the model. There are seven data points that are out of leverage limit (i.e., 0.002) from the training data set. There are only 182 data points from 11,022 data points of the training data set that are labeled as suspected, which were identified outside the acceptable application area of the developed model. These outcomes of the Leverage method confirm that the developed model provides the reliability estimation of the LCR.

5.3. General Discussion. The best-developed model can be only applied in a similar application with the same hole section and range of input parameters, as discussed in Section 2.1. The data set of any future well has to be recorded at a similar data frequency to be able to apply the best-developed model. Otherwise, interpolation techniques need to be applied at the required time to estimate the missing channels, i.e., Q, ROP, RS, SPP, T, and WOB. It is important to mention that if a well is to be drilled with a different well profile, the models would need to be retrained.

One use of our developed model is within the drilling optimization process. Once the drilling data is captured by drilling one or more stands (1 stand = 93 ft), an ROP model for drilling optimization is trained based on the acquired data. This developed model is then used to optimize the ROP ahead of the drilling bit by adjusting the controllable parameters, mainly RS, WOB, and Q. These optimal parameters along with the expected ROP are fed to the developed LCR model to ascertain that these losses are below the desired threshold set by the end-user. The accepted parameters now are used to drill the next one or two stands. Once drilling the previous one or two stands is completed, the ROP model now is updated, i.e., trained, based on the newly acquired drilling data. The cycle is continued until the drilling activities are completed. Figure 8 shows a flowchart that describes how the drilling optimization process can be linked to our machine learning-based algorithm described in this paper.

Another use of the best-developed model is within the planning phase of drilling a similar section that was used in this study. Using the best-developed model, the drilling crew can have a predrill model with the optimum operational parameters to drill a new offset well, that is nearby the wells used in this study, with minimal losses. This can be achieved by dividing each well used for training to a small interval of depth (e.g., 5 ft) and obtaining the drilling surface parameters corresponding to the minimum LCR. Then, the arithmetic average for each
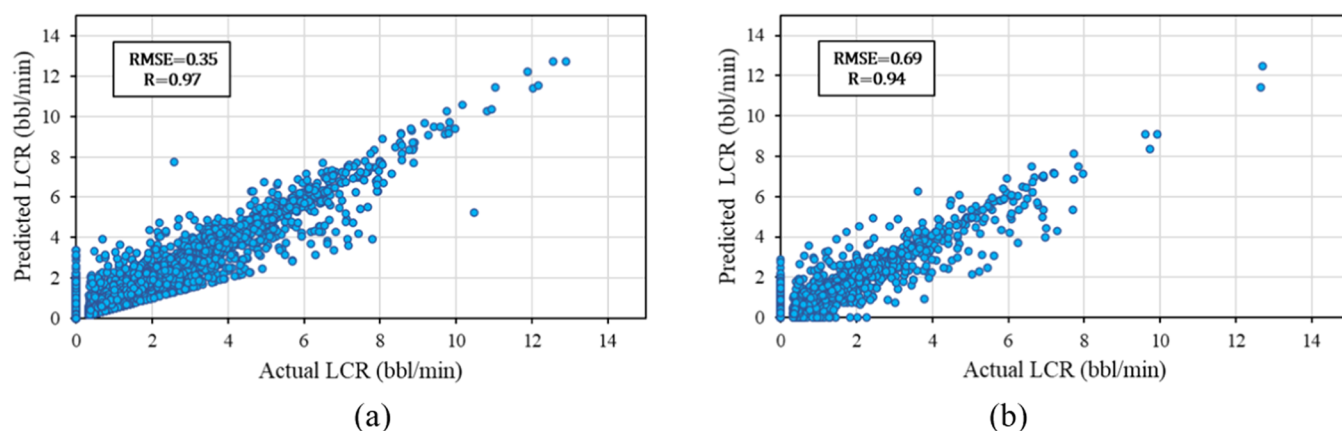
Figure 5. Cross-plots of the actual LCR versus the predicted LCR of the optimum K-NN model. (a) Training set. (b) Testing set.

Table 10. Optimum Design Parameters of the Developed Models

	SVM		RF		K-NN
Kernel function	RBF	number of estimators	100	K	2
C	1	maximum depth	13	distance	Manhattan
γ	auto	maximum features	sqrt or Log_2		

drilling parameter can be computed to generate a road map for every 5 ft while drilling the open-hole section in the subject well. The weighted average can be used if the user prefers to give more importance to the nearest wells.

6. CONCLUSIONS

The conventional methods used to monitor the loss of circulation have some limitations as discussed above; hence, a data-driven model to predict the LCR was established. This paper introduced an ML application to continuously predict the LCR based on surface drilling parameters including Q, SPP, WOB, ROP, RS, and T. The main findings of the study are listed below:

- Using the best-developed model of those compared, the drilling crew can have a predrill model with optimization to drill a new offset well, that is nearby the wells used in this study, with minimal losses.

- The best-developed model of those compared can be used to predict the LCR based on the current drilling parameters applied by the drilling crew and hence advise the drilling crew to alter the operational drilling parameters if the LCR was predicted to be high.

- The good accuracy of the best-developed model, i.e., K-NN, indicates that it would be possible to use a data-driven model to predict the LCR while drilling activities. This will be most beneficial in the old drilling rigs where a measuring stick is still used to know the fluid level inside the mud pits to calculate the LCR.
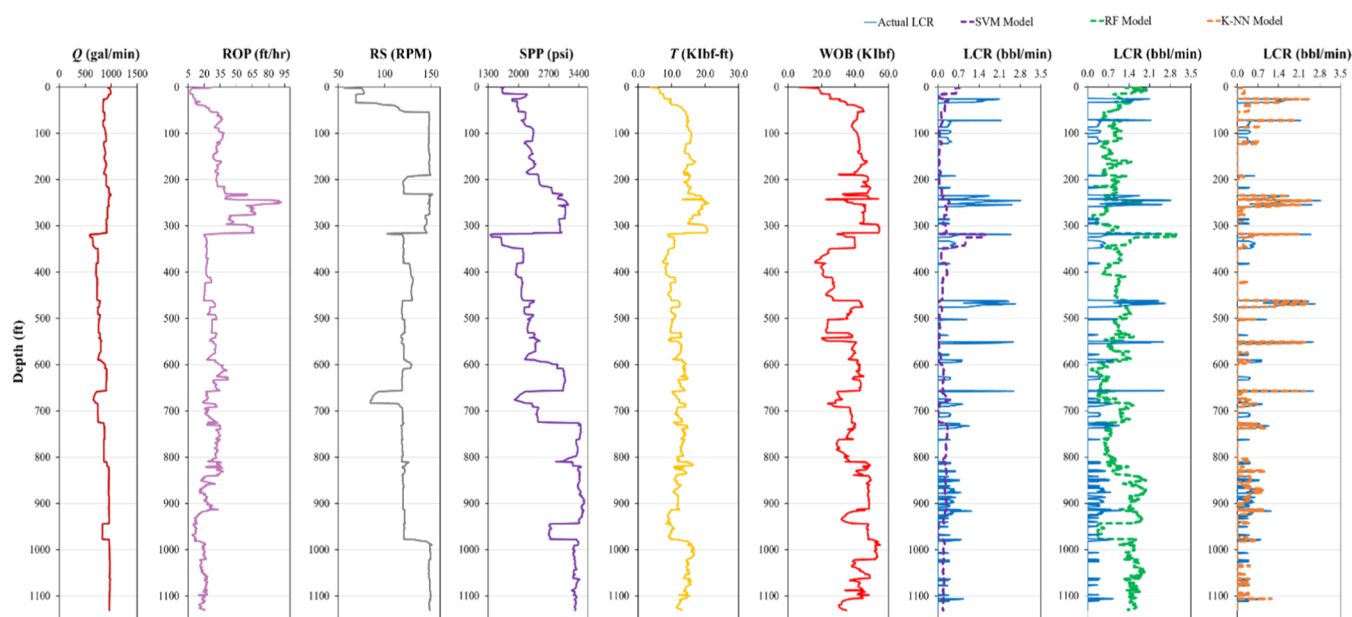

Figure 6. Stack plot of the dependent variables and the output variable (i.e., LCR) versus the adjusted well depth. The SVM-based model (dashed purple curve), the RF-based model (dashed green curve), and the K-NN-based model (dashed orange curve) are superimposed onto the actual LCR (blue curve) for Well No. 8 (1123 unseen data points).
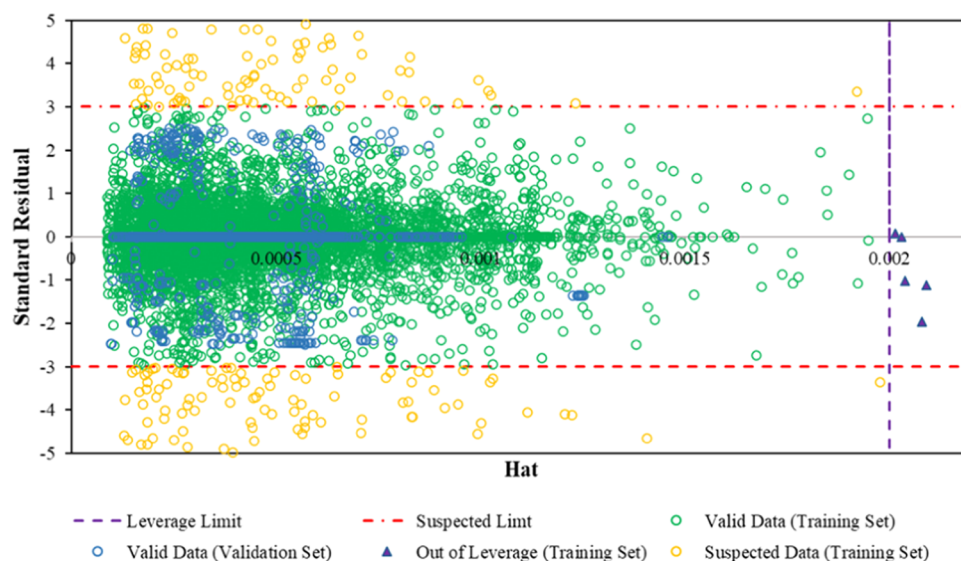
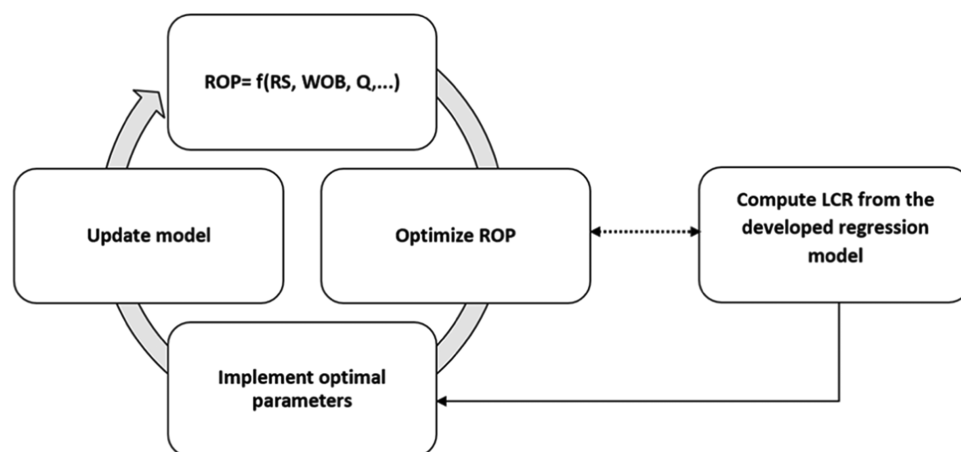**Figure 7.** William's plot for identifying the application area of the K-NN model and doubtful data.

Figure 8. Flowchart that describes how the developed LCR model can be used for the drilling optimization process.

Table 11. Descriptive Statistics of Well No. 1 (2379 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	609.1	4.9	60.3	682.8	7.3	7.1	0.0
max	1105.7	135.2	186.7	3045.7	23.0	63.1	8.1
range	496.5	130.3	126.4	2362.9	15.8	56.1	8.1
mean	814.5	39.9	144.0	2260.4	16.1	42.2	0.5

6.1. Limitations and Future Plans. There are some limitations associated with the data-driven models used in this study that can be summarized as follows:

- The data set was collected from a specific hole section and a range of input parameters, as discussed in Section 2.1. Therefore, the models would need to be retrained if a well is to be drilled with different well profiles and mechanical drilling parameters.

- Data-driven models are extremely dependent on data quality. Most sensors are exposed to harsh environmental conditions at the rig site, such as mechanical shocks, temperature changes, humidity, etc. Therefore, it is essential to establish a recalibration cycle as inaccurate data can mislead the drilling crew.

- It is imperative to mention that the data sets used in this study were collected from different rigs with different acquisition systems that transmit data at a different frequency, which is defined as the number of data points recorded per second. The data set of any future well has to be recorded at a similar frequency to be able to apply the best-developed model. Otherwise, interpolation techniques need to be applied at the required time span to estimate the missing channels, i.e., Q, ROP, RS, SPP, T, and WOB.

■ **APPENDIX A—DESCRIPTIVES STATISTICS OF ALL WELLS**

Table 12. Descriptive Statistics of Well No. 2 (2117 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	635.9	8.9	50.5	566.6	8.8	12.3	0.0
max	1086.8	126.7	157.5	3161.9	25.3	65.2	8.1
range	450.9	117.7	106.9	2595.3	16.6	52.9	8.1
mean	942.8	43.3	112.7	2436.3	17.6	58.0	0.9

Table 13. Descriptive Statistics of Well No. 3 (2143 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	653.9	12.2	60.5	791.5	10.9	5.4	0.0
max	1104.4	91.9	134.7	2837.7	23.2	59.3	8.0
range	450.5	79.8	74.2	2046.2	12.3	53.8	8.0
mean	816.7	33.7	112.1	1493.4	18.7	36.0	1.7

Table 14. Descriptive Statistics of Well No. 4 (1800 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	466.6	24.5	35.7	845.9	3.3	8.0	0.0
max	1044.1	215.1	159.8	3446.0	16.4	55.4	10.0
range	577.5	190.6	124.2	2600.1	13.1	47.4	10.0
mean	705.3	63.7	117.1	1927.6	10.7	34.7	1.6

Table 15. Descriptive Statistics of Well No. 5 (2375 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	633.7	9.0	59.2	797.4	9.6	8.5	0.0
max	967.7	90.7	160.0	2980.0	19.2	50.5	12.9
range	333.9	81.7	100.8	2182.6	9.6	42.0	12.9
mean	758.0	35.0	102.8	2068.9	14.5	30.1	0.7

Table 16. Descriptive Statistics of Well No. 6 (996 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	700.2	17.4	62.0	857.6	3.3	19.2	0.0
max	1010.3	140.0	157.5	3964.2	18.8	53.5	9.9
range	310.0	122.5	95.5	3106.7	15.5	34.3	9.9
mean	900.7	75.1	139.1	3125.9	13.1	43.5	0.9

Table 17. Descriptive Statistics of Well No. 7 (2084 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	662.3	13.8	55.6	998.9	3.9	3.5	0.0
max	1003.1	72.0	130.9	2953.3	17.2	56.3	12.2
range	340.8	58.2	75.4	1954.4	13.3	52.8	12.2
mean	929.7	45.2	114.1	2433.3	13.2	32.0	0.8

Table 18. Descriptive Statistics of Well No. 8 (1123 Data Points)

statistical parameters	Q (gal/min)	ROP (fph)	RS (RPM)	SPP (psi)	T (KIbf-ft)	WOB (klbf)	LCR (bbl/min)
min	584.9	9.7	68.7	1345.4	3.4	20.4	0.00
max	988.4	93.0	151.2	3523.4	21.1	55.0	2.79
range	403.5	83.3	82.5	2177.9	17.7	34.6	2.79
mean	889.3	27.4	128.4	2953.5	12.9	41.1	0.15

■ AUTHOR INFORMATION

Corresponding Author

Salaheldin Elkatatny − *Department of Petroleum Engineering, King Fahd University of Petroleum &Minerals, Dhahran 31261, Saudi Arabia;* ● orcid.org/0000-0002-7209-3715; Email: elkatatny@ kfupm.edu.sa

Authors

Ahmed Alsaihati − *Department of Petroleum Engineering, King Fahd University of Petroleum &Minerals, Dhahran 31261, Saudi Arabia*

Mahmoud Abughaban − *Drilling Technology Division, EXPEC Advanced Research Center, Dhahran 34466, Saudi Arabia*

Dhafer Al Shehri − *Department of Petroleum Engineering, King Fahd University of Petroleum &Minerals, Dhahran 31261, Saudi Arabia;* ● orcid.org/0000-0002-7032-5199

Author Contributions

Conceptualization: A.A., M.A., S.E., and D.A.S.; methodology: A.A. and M.A.; software: A.A.; formal analysis: A.A.; investigation: A.A., S.E., and D.A.S.; data curation: S.E.; writing—original draft preparation: A.A.; writing, review, and editing: S.E. and M.A.; supervision: S.E., D.A.S., and M.A. All authors have read and agreed to the published version of the manuscript.

Notes

The authors declare no competing financial interest.

The data that support the finding of this study are available upon reasonable request from the corresponding author. The data are not publicly available as they contain sensitive information.

■ NOMENCLATURE

AI   artificial intelligence
ANN   artificial neural network
APV   active pit volume, bbl.
APV_a actual reduction of APV, bbl.
APV_{TH} theoretical reduction of active pit volume, bbl.
β damping factor, unitless
C regularization parameter, unitless
D_i inner diameter of the bottom-hole assembly, in.
ΔD the difference between two consecutive depths, ft.
ΔAPV difference between the actual and theoretical reduction of APV, bbl.
D_o outer diameter of the bottom-hole assembly, in.
K number of neighbors, unitless
K-NN   K nearest neighbors
LCR   loss of circulation rate, bbl/min
ML   machine learning
OH hole size, in.
Q flow rate, gal/min
R correlation coefficient, unitless
R^2 coefficient of determination, unitless
RBF   radial basis function
RF   random forests
RMSE   root-mean-square error
ROP   rate of penetration, ft/h
RPM   revolution per minute
RS   rotary speed, RPM
SPP   standpipe pressure, psi
SVM   support vector machines
T surface drilling torque, kIbf-ft
WOB   weight on bit, kIbf

■ REFERENCES

(1) Howard, G. C.; Scott, P. An Analysis and the Control of Lost Circulation. *J. Pet. Technol.* **1951**, *3*, 171−182.

(2) Ghalambor, A.; Salehi, S.; Shahri, M. P.; Karimi, M. *Integrated Workflow for Lost Circulation Prediction*, Presented at the SPE International Symposium and Exhibition on Formation Damage Control, Society of Petroleum Engineers, Lafayette, Louisiana, USA, Feb 26−28, 2014.

(3) Schafer, D. M.; Loeppke, G. E.; Glowka, D. A.; Scott, D. D.; Wright, E. K. In *An Evaluation of Flowmeters for the Detection of Kicks and Lost Circulation During Drilling*, Presented at the IADC/SPE Drilling Conference, New Orleans, Louisiana, Feb 18−21, 1992.

(4) Lavrov, A. *Lost Circulation Mechanisms and Solutions*, 1st ed.; Gulf Professional Publishing: Amsterdam: The Netherlands, 2016.

(5) Beda, G.; Carugo, C. In *Use of Mud Micro loss Analysis While Drilling to Improve the Formation Evaluation in Fractured Reservoir*, Presented at the SPE Annual Technical Conference and Exhibition held in New Orleans, Louisiana, Sept 30−Oct 3, 2001.

(6) Shahab, M. Virtual-Intelligence Applications in Petroleum Engineering: Part 1—Artificial Neural Networks. *J. Pet. Technol.* **2000**, *52*, 64−73.

(7) Balaji, K.; Rabiei, K.; Suicmez, V.; Canbaz, C.; Agharzeyva, Z.; Tek, S.; Bulut, U.; Temizel, C. In *Status of Data-Driven Methods and Their Applications in Oil and Gas Industry*, Presented at 80th EAGE Conference and Exhibition held in Copenhagen, Denemark, June 11−14, 2018.

(8) Bello, O.; Holzmann, J.; Yaqoob, T.; Teodoriu, C. Application of Artificial Intelligence Methods in Drilling System Design and Operations: A Review of the State of The Art. *J. Artif. Intell. Soft Comput. Res.* **2015**, *5*, 121−139.

(9) Ren, X.; Hou, J.; Song, S.; Liu, Y.; Chen, D.; Wang, X.; Dou, L. Lithology identification using well logs: A method by integrating artificial neural networks and sedimentary patterns. *J. Pet. Sci. Eng.* **2019**, *182*, No. 106336.

(10) Ahmed, A. S.; Mahmoud, A. A.; Elkatatny, S. In *Fracture Pressure Prediction Using Radial Basis Function*, Proceedings of the AADE National Technical Conference and Exhibition, Denver, CO, USA, April 9−10, 2019a.

(11) Ahmed, A. S.; Mahmoud, A. A.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A. In *Prediction of Pore and Fracture Pressures Using Support Vector Machine*, Proceedings of the 2019 International Petroleum Technology Conference, Beijing, China, March 26−28, 2019b.

(12) Elkatatny, S.; Mahmoud, M. Development of new correlations for the oil formation volume factor in oil reservoirs using artificial intelligent white box technique. *Petroleum* **2018**, *4*, 178−186.

(13) Elkatatny, S.; Al-AbdulJabbar, A.; Mahmoud, A. A. New Robust Model to Estimate the Formation Tops in Real-Time Using Artificial Neural Networks (ANN). *Petrophysics* **2019**, *60*, 825−837.

(14) Alsaihati, A.; Elkatatny, S.; Mahmoud, A. A.; Abdulraheem, A. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities While Drilling Horizontal Wells, With Real Case Study. *J. Energy Resour. Technol* **2020**, *143*, No. 043201.

(15) Magana-Mora, A.; Affleck, M.; Ibrahim, M.; Makowski, G.; Kapoor, H.; Otalvora, W.; et al. Well Control Space Out: A Deep-Learning Approach for the Optimization of Drilling Safety Operations. *IEEE Access* **2021**, *9*, 76479−76492.

(16) Luu, T.; Bomidi, J.; Magana-Mora, A.; Alawi, A.; Zhan, Z. In *Bi-Directional Long Short-Term Memory Variational Autoencoder for Real-Time Bit-Wear Estimation*, Presented at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, virtual, Oct 12−14, 2021.

(17) Alsaihati, A.; Alotaibi, B. Determining Severity of Lateral and Torsional Downhole Vibrations While Drilling Surface Holes Using Three Machine Learning Techniques. *SPE J.* **2022**, 1−11.

(18) Jeirani, Z.; Mohebbi, A. Estimating the initial pressure, permeability and skin factor of oil reservoirs using artificial neural networks. *J. Pet. Sci. Eng.* **2006**, *50*, 11−20.

(19) Moazzeni, A. R.; Nabaei, M.; Jegarluei, S. G. In *Prediction of Lost Circulation Using Virtual Intelligence in One of Iranian Oilfields*, Presented at the Nigeria Annual International Conference and Exhibition, Tinapa, Calabar, Nigeria, July 31−Aug 7, 2010.

(20) Jahanbakhshi, R.; Keshavarzi, R.; Jalili, S. Artificial neural network-based prediction and geomechanical analysis of lost circulation in naturally fractured reservoirs: a case study. *Eur. J. Environ. Civ. Eng.* **2014**, *18*, 320−335.

(21) Jahanbakhshi, R.; Keshavarzi, R. Quantitative and qualitative analysis of lost circulation in natural and induced fractured formations: the integration of operational conditions and geomechanical parameters. *Eur. J. Environ. Civ. Eng.* **2015**, *19*, 418−444.

(22) Toreifi, H.; Rostami, H.; Manshad, A. K. New method for prediction and solving the problem of drilling fluid loss using modular neural network and particle swarm optimization algorithm. *J. Pet. Explor. Prod. Technol.* **2014**, *4*, 371−379.

(23) Manshad, A.; Rostami, H.; Niknafs, H.; Mohammadi, A. H. *Integrated Lost Circulation Prediction in Oil Field Drilling Operation. Heavy Oil: Characteristics, Production and Emerging Technologies*. Nova Science Publishers, Inc., New York, 2017.

(24) Behnoud far, P.; Hosseini, P. Estimation of lost circulation amount occurs during under balanced drilling using drilling data and neural network. *Egypt. J. Pet.* **2017**, *26*, 627−634.

(25) Solomon, O.; Adewale, D.; Anyanwu, C. In *Fracture Width Prediction and Loss Prevention Material Sizing in Depleted Formations Using Artificial Intelligence*, Lagos, Nigeria, July 31−Aug 2, 2017.

(26) Li, Z.; Chen, M.; Jin, Y. et al. In *Study on Intelligent Prediction for Risk Level of Lost Circulation While Drilling Based on Machine Learning*, Presented at the 52nd US Rock Mechanics/Geomechanics Symposium, Seattle, Washington, USA, June 17−20, 2018.

(27) Abbas, A. K.; Al-haideri, N.; Bashikh, A. Implementing artificial neural networks and support vector machines to predict lost circulation. *Egypt. J. Pet.* **2019**, *28*, 339−347.

(28) Abbas, A. K.; Bashikh, A.; Abbas, H.; Mohammed, H. Intelligent decisions to stop or mitigate lost circulation based on machine learning. *Energy* **2019**, *183*, 1104−1113.

(29) Shi, X.; Zhou, Y.; Zhao, Q. et al. In *A New Method to Detect Influx and Loss During Drilling Based on Machine Learning*, Presented at the International Petroleum Technology Conference, Beijing, China, March 26−28, 2019.

(30) Ahmed, A.; Elkatatny, S.; Ali, A.; Abughaban, M.; Abdulraheem, A. Application of Artificial Intelligence Techniques in Predicting the Lost Circulation Zones Using Drilling Sensors. *J. Sens.* **2020**, *2020*, 1−18.

(31) Hou, X.; Yang, J.; Yin, Q. et al. In *Lost Circulation Prediction in South China Sea Using Machine Learning and Big Data Technology*, Presented at the Offshore Technology Conference, Houston, Texas, USA, May 4−7, 2020.

(32) Gupta, S.; Gupta, A. Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review. *Procedia Comput. Sci.* **2019**, *161*, 466−474.

(33) Brown, R. *Statistical Forecasting for Inventory Control*; McGraw-Hill Book Company: New York, 1959.

(34) Winters, P. R. Forecasting Sales by Exponentially Weighted Moving Averages. *Manage. Sci.* **1960**, *6*, 324−342.

(35) Holt, C. Author's retrospective on 'Forecasting seasonal and trends by exponentially weighted moving averages. *Int. J. Forecast.* **2004**, *20*, 11−13.

(36) Batina, L.; Gierlichs, B.; Prouff, E.; Rivain, M.; Standaert, F.; Veyrat-Charvillon, N. Mutual Information Analysis: A Comprehensive Study. *J. Cryptol.* **2011**, *24*, 269−291.

(37) Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley: New York, N.Y, 1991.

(38) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273−297.

(39) Vapnik, V. *Statistical Learning Theory*; Wiley: New York, 1998.

(40) Drucker, C.; Burges, L.; Kaufman, A.; Smola, A.; Vapnik, V. Support Vector Regression Machines. In *Advances in Neural Information Processing Systems* Mozer, M.; Jordan, M.; Petsche, T., Eds.; MIT Press: Cambridge, MA, 1997; Vol. *9*, pp 155−161.

(41) Hegde, C.; Gray, K. Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *J. Nat. Gas Sci. Eng.* **2017**, *40*, 327−335.

(42) Efron, B. *The Jackknife, the Bootstrap and the Other Resampling Plans*; Society for industrial and Applied Mathematics: Philadelphia, PA, 1982.

(43) Abu Alfeilat, H. A.; Hassanat, A.; Lasassmeh, O.; Tarawneh, A.; Alhasanat, M.; Eyal Salman, H.; Prasath, V. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data* **2019**, *7*, 221−248.

(44) Wang, L. Research and Implementation of Machine Learning Classifier Based on KNN. *IOP Conf. Ser.: Mater. Sci. Eng.* **2019**, *677*, No. 052038.

(45) Coomans, D.; Massart, D. Alternative k-nearest neighbor rules in supervised pattern recognition. *Anal. Chim. Acta* **1982**, *138*, 153−165.

(46) Rousseeuw, P.; Leroy, A. Robust Regression and Outlier Detection. 1987.

(47) Mousavi, S. P.; Atashrouz, S.; Nait Amar, M.; Hemmati-Sarapardeh, A.; Mohaddespour, A.; Mosavi, A. Viscosity of Ionic Liquids: Application of the Eyring's Theory and a Committee Machine Intelligent System. *Molecules* **2021**, *26*, 156.

(48) Menad, N. A.; Noureddine, Z.; Hemmati-Sarapardeh, A.; Shamshirband, S. Modeling temperature-based oil-water relative permeability by integrating advanced intelligent models with grey wolf optimization: Application to thermal enhanced oil recovery processes. *Fuel* **2019**, *242*, 649−663.