Received: 13 December 2021 | Revised: 10 April 2022 | Accepted: 14 June 2022

DOI: 10.1002/cjce.24640

RESEARCH ARTICLE

THE CANADIAN JOURNAL OF CHEMICAL ENGINEERING

Prediction of critical total drawdown in sand production from gas wells: Machine learning approach

Fahd Saeed Alakbari[1] | **Mysara Eissa Mohyaldinn**[1] | **Mohammed Abdalla Ayoub**[1] | **Ali Samer Muhsan**[2] | **Said Jadid Abdulkadir**[3] | **Ibnelwaleed A. Hussein**[4,5] | **Abdullah Abduljabbar Salih**[1]

[1]Petroleum Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Malaysia

[2]Mechanical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Malaysia

[3]Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Malaysia

[4]Gas Processing Centre, College of Engineering, Qatar University, Doha, Qatar

[5]Department of Chemical Engineering, College of Engineering, Qatar University, Doha, Qatar

Correspondence
Mysara Eissa Mohyaldinn, Petroleum Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak, Malaysia.
Email: mysara.eissa@utp.edu.my

Ibnelwaleed A. Hussein, Gas Processing Center, College of Engineering, Qatar University, Doha, Qatar and Department of Chemical Engineering, College of Engineering, Qatar University, Doha, Qatar.
Email: ihussein@qu.edu.qa

Funding information
Yayasan Universiti Teknologi PETRONAS (YUTP FRG), Grant/Award Number: 015LC0-428

Abstract

Sand production is a critical issue in petroleum wells. The critical total drawdown (CTD) is an essential indicator of the onset of sand production. Although some models are available for CTD prediction, most of them are proven to lack accuracy or use commercial software. Furthermore, the previous correlations have not studied the trend analysis to verify the correct relationships between the parameters. Therefore, this study aims to build accurate and robust models for predicting CTD using response surface methodology (RSM) and support vector machine (SVM). The RSM is utilized to obtain the equation without using any software. The SVM model is an alternative method to predict the CTD with higher accuracy. This study used 23 datasets to develop the proposed models. The CTD is a strong function of the total vertical depth, cohesive strength, effective overburden vertical stress, and transit time with correlation coefficients (R) of 0.968, 0.963, 0.918, and −0.813. Different statistical methods, that is, analysis of variance (ANOVA), F-statistics test, fit statistics, and diagnostics plots, have shown that the RSM correlation has high accuracy and is more robust than correlations reported in the literature. Moreover, trend analysis has proven that the proposed models ideally follow the correct trend. The RSM correlation decreased the average absolute percent relative error (AAPRE) by 12.7% compared to all published correlations' AAPRE of 22.6%–30.4%. The SVM model has shown the lowest AAPRE of 6.1%, with the highest R of 0.995. The effects of all independent variables on the CTD are displayed in three-dimensional plots and showed significant interactions.

KEYWORDS
critical total drawdown, machine learning, sand control, sand management

1 | INTRODUCTION

Around 70% of the world's oil and gas wells are produced from weakly consolidated reservoirs.[1,2] Sand production causes numerous issues like equipment erosion and plugging; consequently, sand production leads to a drop in oil recovery, high maintenance cost, and safety, environmental, and health concerns.[2-5] Nevertheless, early prediction of sand production is highly recommended to ensure successful sand control management strategies, according to Khamehchi et al.[6]

Prediction of sand production from oil and gas wells has been addressed through numerical, analytical, and empirical methods. Numerical techniques apply the finite element method (FEM) or finite difference method to account for three-dimensional (3D) stresses, complex material behaviour like plasticity, and fluid flow. Fan et al.[7] used mechanical analysis on a potentially collapsed pore throat system to quantify sand production using a pressure-gradient-based sand-failure-criterion. They state that the failure criterion can be applied to characterize the sand production in cold heavy-oil production with the sand process.[7] However, the numerical methods have some drawbacks since they are time-consuming with high complexity. Furthermore, to build a numerical prediction, the key required input parameters are difficult to find and need advanced laboratory measurements of fluid, mechanical, and petrophysical rock properties.[8] On the other hand, analytical methods have disadvantages, like assuming symmetrical geometry and boundary conditions such as ignoring stress anisotropy. Consequently, ignoring the fundamental effect of stress anisotropy on sanding, a method may not explain the sanding risk related to the borehole orientation. In general, assumptions or approximations without actual data render the models unreliable and inaccurate despite their complexity.[8]

Alternatively, empirical approaches use well data and field observations to predict sand production. These sand prediction methods correlate the well data on sand production and field operation parameters. This technique is classified into one, two, and multi-parameter correlations according to Khamehchi et al.[6] For instance, Tixier et al.[9] used acoustic log data to obtain the shear modulus ratio to bulk compressibility to predict sand production. When the ratio is higher than 0.8×10^{12} psi^2, there is low sand production, whereas when the ratio is less than 0.7×10^{12} psi^2, there is a high probability of sand production.[9] Veeken et al.[10] applied a two-parameter model: the depleted reservoir pressure and drawdown pressure to indicate sand risk. Increasing the number of parameters involved in a model will improve sand prediction modelling.[10]

Response surface methodology (RSM) has been applied in many fields to create models and acquire a relationship between the parameters. The optimal condition for the synthesis of mesoporous carbon is obtained using the RSM.[11] The RSM was used to find the reaction condition optimization for the single-walled carbon nanotubes synthesis.[12] Nam et al.[13] used RSM to determine drilling torques, edge radii, and thrust forces as a function of drill diameter, spindle speed, nanofluid weight concentration, and feed rate. The obtained regression models were able to identify the parameters affecting the drilling performance. Salehnezhad et al.[14] investigated and optimized the rheological behaviour of drilling fluid composed of ZnO nanoparticles and starch by applying RSM. Ishak and Ayoub[15] used RSM to define oil removal performance using an anionic polymer. They stated that their model has R^2 of 94.9% and AAPE of 12.2%.[15] Umar et al.[16] utilized RSM for forecasting the viscosity of petroleum emulsions with an R^2 of 0.8716. They proved that the aging time of an emulsion has the highest impact on its viscosity, and the viscosity of the emulsion improves with time.[16] Alakbari et al.[17] developed RSM to predict the apparent and plastic viscosities of water-based drilling fluids. Their model was developed based on temperature, bentonite, nanosilica, and nanoclay concentrations, and pressure. The apparent and plastic viscosities have an R^2 of 0.978 and 0.961, respectively.[17] Alhajabdalla et al.[18] applied RSM to investigate the stability of fibrous dispersion used in drilling and completion operations. They showed that their model has an R^2 of 0.91–0.99.[18] They proved that polymer concentration is the primary factor influencing fibrous suspension stability.[18] Zhang et al.[19] utilized RSM to forecast profile control by clay particles for a polymer flooded reservoir and stated that the results agree very well with the numerical simulation. The RSM developed statistical models for cuttings transport in non-Newtonian drilling fluids.[20]

Many researchers have recently applied innovative modelling techniques in petroleum engineering calculations, namely machine learning (ML). Zendehboudi et al. and Sircar et al.[21,22] represented comprehensive reviews of the machine learning methods in petroleum engineering applications. They explained the ML techniques and revealed their benefits and limitations.[21,22] The support vector machine (SVM) method is one of the ML methods successfully applied in many applications. An SVM algorithm was used to solve non-linear pattern recognition problems of high generalization capability, few tuning parameters, and small sample size.[23] Kamari et al.[24] applied the least-square-support-vector machine to predict the unloading pressure gradient region in different oil production rates and tubing sizes. Their model has an average absolute relative deviation and squared

correlation coefficient of 1.084% and 0.9994.[24] Olatunji and Micheal[25] used a SVM classification method to detect the sand production in the Niger Delta reservoirs. They used fluid, rock, geotechnical, and other data to build their model and stated that their model could accurately forecast sand production.[25] SVM models have been used for determining the bubble point pressure and oil formation volume factor and have proved the flexibility and reliability of the model, according to El-Sebakhy.[26] Al-Azani et al.[27] used an SVM model to obtain the hole cleaning efficiency,[27,28] while Cheng et al.[29] applied SVM to detect sand deposition in the pipeline. An SVM model was also applied to predict sandstone thickness using the geological and geophysical characterization data (Li et al.[30]). Yasin et al.[31] utilized SVM to find the spatial distribution of porosity using seismic and log data.

Some researchers used multi parameters, that is, total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS), to determine the critical total drawdown (CTD) for detecting sand production. Total drawdown (TD) is the difference between the original static reservoir pressure and the bottom hole flowing pressure. This parameter is determined at the onset of sanding for problematic wells and is known as the CTD.[32] Some empirical correlations were used to predict the CTD as an indicator of the onset of sand production. Kanj and Abousleiman[32] used artificial neural networks (ANNs) to predict the CTD for 31 wells of the Adriatic Sea. Khamehchi et al.[6] applied multiple linear regression (MLR) with the regression Equation (1) and a genetic algorithm evolved MLP (GA-MLR) to determine the CTD for 23 wells of the Adriatic Sea. Alakbari et al.[33] applied the fuzzy logic-based model using TVD, TT, COH, and EOVS as inputs to predict the CTD. Table 1 summarizes the previous correlations and models used to determine the CTD and their used input parameters.

$$Y = b_1X_1 + b_2X_2 + \ldots + b_nX_n + C \quad (1)$$

However, the previously published correlations and models are shown to lack accuracy or use sophisticated commercial software. Alakbari et al.[33] utilized commercial software to predict the CTD. Kanj and Abousleiman[32] and Khamehchi et al.[6] showed the correlations to predict the CTD with more than 20% average absolute percent relative error (AAPRE). Alakbari et al.[33] stated that the fuzzy logic-based model could predict the CTD with 8.6% AAPRE; however, there is a need to improve the CTD accuracy, and exploration of other techniques to predict the CTD is required. In addition, the published correlations have not conducted the trend analysis to indicate the proper relationships between the input and output parameters to prove the correct physical behaviour.

TABLE 1 Inputs' previous correlations and models

Correlation/model	Used input parameters
Kanj and Abousleiman[32]	COH.
Khamehchi et al.[6] (MLR)	TVD, TT, COH, and EOVS.
Khamehchi et al.[6] (GA-MLR)	TVD, TT, COH, and EOVS.
Alakbari et al.[33]	TVD, TT, COH, and EOVS.

Abbreviations: COH, cohesive strength; EOVS, effective overburden vertical stress; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression; TT, transit time; TVD, total vertical depth.

Therefore, this study aims to apply RSM and SVM to develop more accurate CTD prediction models than the models that exist in the literature. The RSM correlation and SVM model were developed using data from 23 gas wells of the Adriatic Sea. The data parameters include CTD as output and TVD, TT, COH, and EOVS as input parameters. The RSM correlation and SVM models used in this study have been statistically analyzed and validated using analysis of variance (ANOVA), *F*-statistics test, fit statistics, diagnostic plots, and trend analysis to show their reliability and accuracy. In addition, some statistical error analyses, such as AAPRE, maximum absolute percent relative error ($E_{\max.}$), minimum absolute percent relative error ($E_{\min.}$), root mean square error (RMSE), standard deviation (SD), and R have been conducted to describe, validate, and compare the RSM correlation and SVM model with the existing models. The SVM is the highest accuracy model to predict the CTD compared to all models; however, the use of RSM was found to have some benefits compared to the SVM. These benefits include providing a direct equation, as no software is needed to predict the CTD. The RSM also helps identify interaction by plotting the 3D surface response of the CTD model used to assess the interactive relationships between (TVD, TT, COH, and EOVS) inputs and the CTD output. Moreover, the RSM correlation outperformed the studied equations used to predict the CTD. The RSM and SVM methods are used for the first time in this study to predict CTD. By applying RSM and SVM techniques, the study aims to develop accurate and robust models to predict the CTD, an essential indicator of the onset of sand production. Such a model will help control costly sand production issues and may be consequential.

2 | METHODOLOGY

This research was performed by following the phases, as shown in Figure 1. First, the datasets were collected from

the literature.[34] Then, the collected datasets were divided into two subdivisions, that is, training or developing and testing or validation datasets. After that, the RSM and SVM models were developed to determine the CTD. The developed models were evaluated by applying different methods, that is, ANOVA, *F*-statistics test, fit statistics, diagnostics plots, and trend analysis, to indicate that the models are robust and show the proper relationships between the inputs and outputs to prove the correct physical behaviour. Then, cross plots and various statistical error analyses, namely *R*, SD, APRE, AAPRE, and RMSE, were conducted to indicate the accuracy of the proposed models and the previous models. Finally, different correlations and the proposed models were compared, and the highest accuracy model was selected to predict the CTD.

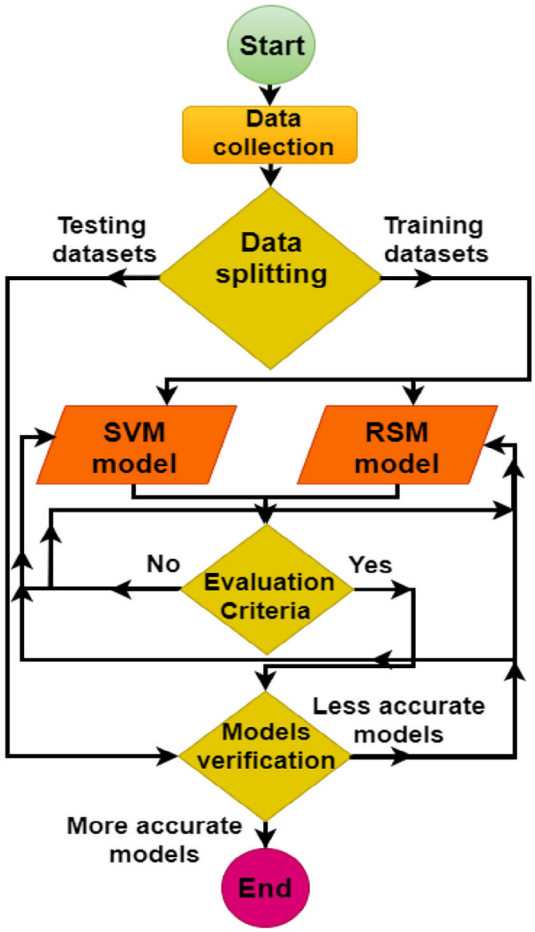

FIGURE 1 Flowchart of the proposed response surface methodology (RSM) and support vector machine (SVM) models

2.1 | Data collection and description

The current study data are obtained from 23 gas wells from the Adriatic Sea collected from the literature.[34] The data were divided into two sections; the first section contains 15 wells used to develop the RSM correlation and train and validate the SVM model. The second section has eight wells to test the RSM correlation, SVM model, and existing correlations to indicate that the proposed RSM and SVM models are more robust. In addition, the testing dataset can be used to make a fair comparison between all models by using the same dataset. Statistical analyses, namely, minimum, maximum, mean, and range of the collected data, are demonstrated in Tables 2 and 3.

2.2 | Leave-one-out cross-validation

Leave-one-out cross-validation is one of the cross-validation methods in which each dataset can be individually assigned to the testing set and the rest to the training set. The number of folds equals the number of datasets.[35] The application of leave-one-out cross-validation is proper in small datasets.[36] Furthermore, the leave-one-out cross-validation is less biased than other methods because the training set is $n - 1$ size.[37] In addition, the leave-one-out cross-validation was conducted to overcome problems such as over-fitting and under-fitting. Therefore, the leave-one-out cross-validation is applied. In this study, 15 wells were used for training and leave-one-out cross-validation. Hence, the

TABLE 2 Statistical analysis of the data used to develop the response surface methodology correlation training and validating the support vector machine model

Parameter	TVD (m)	TT (μsec/ft)	COH (MPa)	EOVS (MPa)	CTD (MPa)
Minimum	1070	85.00	0.539	10.88	0.314
Maximum	4548	170.0	5.217	80.71	43.97
Mean	2564	115.0	1.775	38.16	15.28
Median	2380	110.0	1.275	29.42	12.81
Range	3478	85.00	4.678	69.82	43.66
Skewness	0.187	0.940	1.234	0.398	0.600
SD	10.23	0.208	0.012	0.228	0.123

Abbreviations: CTD, critical total drawdown; COH, cohesive strength; EOVS, effective overburden vertical stress; TT, transit time; TVD, total vertical depth.

TABLE 3 Statistical analysis of the data used for testing the previous and response surface methodology correlations and the support vector machine model

Parameter	TVD (m)	TT (μsec/ft)	COH (MPa)	EOVS (MPa)	CTD (MPa)
Minimum	1122	85.00	0.559	11.28	0.883
Maximum	4088	150.0	3.874	76.59	32.60
Mean	2618	116.5	1.812	41.02	16.27
Median	2660	108.5	1.618	41.87	16.22
Range	2966	65.00	3.315	65.31	31.71
Skewness	−0.008	0.387	0.728	0.065	0.125
SD	10.57	0.229	0.012	0.265	0.119

Abbreviations: CTD, critical total drawdown; COH, cohesive strength; EOVS, effective overburden vertical stress; TT, transit time; TVD, total vertical depth.

average of all the 15 wells' validation results was performed to indicate the performance of the SVM model for the validation datasets. Finally, the datasets that were not used for training and validating the SVM model (eight datasets that are independently assigned) were used to test the SVM model.

2.3 | CTD model applying RSM

RSM is a common method utilized for conducting the design of experiments (DOE). RSM is used as a statistical and mathematical tool for optimizing industrial processes or chemical reactions and is generally used for experimental design.[38,39] RSM uses a mathematical model to analyze the independent variables. RSM includes DOE and a regression analysis technique that provides a relationship between independent and dependent parameters in a mathematical model that analyzes the independent variables.[40] This relationship could be expressed in the following Equation (2). The RSM can generate a mathematical model that defines the overall process.[41] In addition, the RSM can be used to save time and cost by decreasing the number of trials in the design of experiments.[42]

$$\eta = f(x_1, x_2, \ldots x_n) + \varepsilon \tag{2}$$

where η is the response in this study, represented by selected CTD; f is the unknown function of response; $x_1, x_2, \ldots x_n$ are the independent variables in this study, represented by total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS); n is the number of the independent variables; and ε is the statistical error.

RSM uses a low-order polynomial equation to obtain a predetermined region of the independent variables, which are later analyzed to locate the optimum values of the independent variables to identify the response.[43] Central composite design (CCD), historical data design (custom design), D-optimal designs, and Box–Behnken–Behnken are the most commonly used forms in the RSM.[44] CCD is a design approach for continuous experimentation; it allows for testing a lack of fit when an adequate number of experimental values are provided.[45] The custom designs are applied when the process involves changes to the experiment that cannot be adjusted by a standard design and data has already been gathered.[46] The historical data design was applied in this research.

In this study, the RSM method has been utilized to obtain relationships between the following parameters: TVD, TT, COH, and EOVS, and the CTD response. The CTD model was obtained by a statistical association between independent parameters and dependent parameters (responses). Design-Expert software and Statistica software have developed new empirical correlations for predicting the CTD as a function of TVD, TT, COH, and EOVS. The Design-Expert and Statistica software have recently been used with the regression method to obtain the best match for the data.[47] Regression analysis is utilized to obtain the relationships between the features and the responses to acquire a mathematical model that satisfies the relationship between test factors and objective functions.[48] The regression method is applied to investigate the behaviour of the response surface.[49] An RSM mathematical correlation can be generated by utilizing a polynomial function fitted by the least square method.[50] Generally, the response surface model is determined by the following equation:

$$Y = b_o + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=1}^{k} b_{ij} x_i x_j + e \tag{3}$$

where Y is the response, x_i and x_j are the variables (i and j from 1 to k), b_o is the model intercept coefficient, b_j, b_{jj}, and b_{ij} are the interaction coefficients of the linear, quadratic, and second-order terms, respectively, k is the number of the features ($k = 5$ in this work), and e is the error.[51]

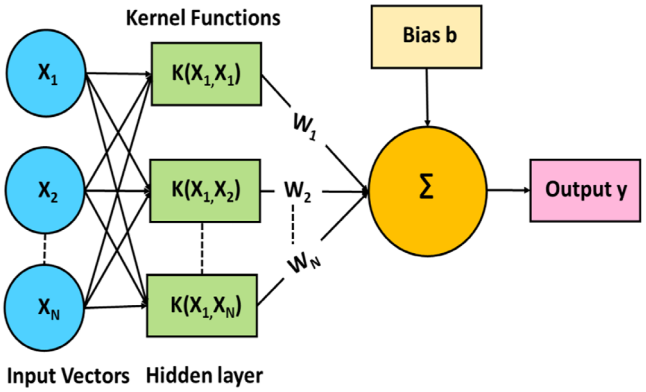

FIGURE 2 Support vector machine (SVM) architecture

2.4 | CTD model using an SVM

SVM is a supervised learning technique derived from statistical learning theory for regression and classification.[52] SVM can be used to determine a hyperplane that puts the data into two classes.[53] As illustrated in Figure 2, the SVM structure comprises input vectors, hidden layers (kernel functions), the w and b (the adjustable model parameters), and the output y. SVM has advantages like non-convergence to local minima, accurate generalization, and predictive ability for small and non-linear datasets.[54,55] The SVM is a constrained non-linear optimization problem. Therefore, the SVM can provide a unique solution in many applications.[21] Moreover, the SVM incorporates the structural risk minimization (SRM) principle by applying a maximum margin approach. The SVM model has been proposed to decrease its computational cost while enhancing its generalization ability.[56] Richhariya and Tanveer[56] used a small-sized rectangular kernel matrix. The decreased kernel-based approach leads to a computationally efficient model of Universum-based SVM. Hence, this removes the overhead of the higher computation cost of Universum based algorithms.[56] The SVM technique is robust due to its ability to control noisy data.[21,57] Furthermore, the SVM can learn the feature space dimensionality independently. Thus, SVM provides a good generalization even in the presence of many features.[58] The SVM can classify two linearly conceivable classes of the data, utilizing a linear hyperplane with a maximum margin from both classes.

The SVM model in this research was developed using MATLAB R2020b. The SVM model includes the idea of the kernel function. Some kernel functions are used in the SVM, namely linear, polynomial, radial basis function (RBF), sigmoid, and Gaussian kernel functions.[59] Selecting the kernel function and optimized SVM model parameters are essential for improving the SVM model's performance.[60] In this study, the linear kernel function was chosen because it provides the highest accuracy results compared to other kernel functions, that is, polynomial, RBF, and sigmoid kernel functions. The linear kernel function can be applied when the data is linearly separable and can be separated utilizing a single line.[61] The datasets are not projected onto higher dimensions when the linear kernel is applied. Therefore, the linear kernel is the inner product of x_1 and x_2 with the c constant. The linear kernel function can be determined by using Equation (4). The advantage of the linear kernel is that it is the simplest of all the kernel functions and has only the c term.[62]

$$k(x_1, x_2) = x_1{}^T . x_2 + c \tag{4}$$

Table 4 shows the optimized SVM model parameters used to generate the CTD. The epsilon parameter plays a crucial role in enhancing the SVM performance and is 0.03308 in this study. The solver used in this study is SMO, and the box constraint is 10.831. Additionally, the kernel scale and standardize are 3.2800 and accurate, respectively. The optimized parameters in Table 4 are selected to provide the best proposed SVM model that can accurately and robustly determine the CTD.

Before applying the SVM model, the data were normalized between one and minus one by using (mapminmax) function in MATLAB or utilizing the following:

$$Y = (Y_{\max} - Y_{\min}) \times \frac{X - X_{\min}}{X_{\max} - X_{\min}} + Y_{\min} \tag{5}$$

where Y is the parameter in the normalized form, $Y_{\max}$ is the maximum value of the normalized form (1), $Y_{\min}$ is the minimum value of the normalized form (-1), X is the parameter to be normalized, $X_{\min}$ is the maximum parameter, and $X_{\max}$ is the minimum parameter.

After normalization, the dataset is divided into two subparts, 15 wells for the training and validation of the SVM model and eight wells for testing the model, as shown in Tables 2 and 3, respectively.

The correlation coefficient (R) is calculated to evaluate and determine the importance of each input parameter (TVD, TT, COH, and EOVS) to the output CTD. The CTD is a strong function of the TVD, COH, and EOVS, where the correlation coefficients were 0.968, 0.963, and 0.918, respectively. In addition, it is a strong function in the opposite direction of the TT, where the R was -0.813, as shown in Figure 3. A negative R implies that the variables are inversely related.[63]

The following stages were used to evaluate and validate the proposed SVM model and compare it with the

TABLE 4 Specifications of the support vector machine model

Parameter	Description/value
Inputs	4
Outputs	1
Box constraint	10.831
Kernel scale	3.2800
Kernel function	Linear
Epsilon	0.033 08
Standardize	True
Solver	Sequential minimal optimization
Bias	−0.3374

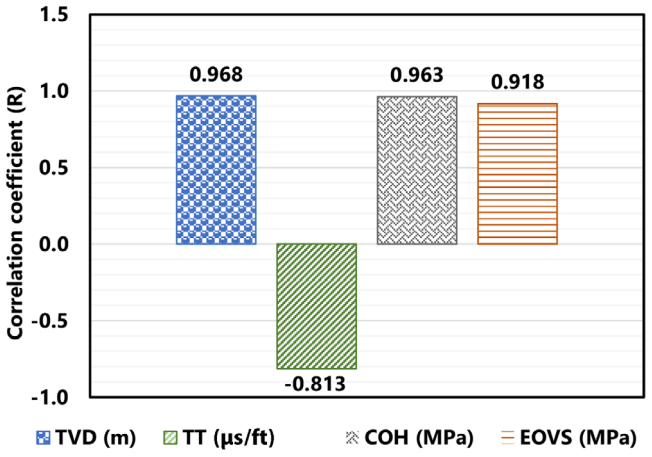

FIGURE 3 Relative importance of input parameters with critical total drawdown (CTD). COH, cohesive strength; EOVS, effective overburden vertical stress; TT, transit time; TVD, total vertical depth

other models: Initially, the SVM model was checked, applying trend analysis to ensure the model is stable and simulates the physical process trends. Second, the proposed SVM performance was shown by using cross plots and statistical error analysis, that is, *R*, SD, APRE, AAPRE, and RMSE. Finally, the proposed SVM performance was compared with the RSM correlation and the current correlations and models in the literature.

3 | RESULTS AND DISCUSSION

3.1 | Mathematical RSM correlation

The RSM correlation resulted in an empirical relationship between the response CTD. The features are shown in the following equation:

$$Y = 1.521\,680\,784\,615 + 0.007\,103\,677\,883\,743\,7 \times X_1 - 0.072\,458\,817\,372\,904 \times X_2 + 4.615\,005\,616\,577\,1 \times X_3 - 0.661\,282\,992\,199\,99 \times X_4 + 5.239\,182\,293\,300\,3e-05 \times X_1 \times X_4 + 0.003\,814\,386\,245\,673 \times X_2 \times X_4 \quad (6)$$

where Y is CTD (MPa), X_1 is TVD (m), X_2 is TT (micro s/ft), X_3 is COH (MPa), and X_4 is EOVS (MPa).

3.2 | Verification of the RSM and SVM models

The proposed RSM and SVM models were tested by using eight unseen wells data points that were not used to develop the RSM model and to train and validate the SVM model. Statistical error analyses, namely *R* and AAPRE, have been performed to evaluate the proposed RSM and SVM models. The equations to determine the *R*, AAPRE, RMSE, and other statistical error analyses are presented in (the Supporting Information file: Appendix S1). As shown in Table 5, the proposed RSM and SVM models have low AAPRE, RMSE, and fSD, and a high *R*. Therefore, according to Table 5, the proposed RSM and SVM models' accuracies are high.

The cross plots in Figures 4 and 5 show that the RSM correlation and SVM model have high accuracy and high confidence level models. The RSM correlation and SVM model were built in this research to determine whether the CTD was acceptable based on the results found. Comparing the two models, the SVM model has higher accuracy than the RSM correlation, as displayed in Table 5.

3.3 | Evaluation criteria

The developed RSM correlation for CTD has been statistically analyzed and validated by ANOVA, *F*-statistics test, fit statistics, and diagnostics plots.

3.3.1 | ANOVA for the CTD model

ANOVA is applied to determine whether the quadratic model represents the significance of the experimental data at a 95% confidence interval.[64] ANOVA is utilized to obtain the influence of features (TVD, TT, COH, and EOVS) on the response CTD in the regression study. Finally, the ANOVA was used for the statistical analyses implemented to detect the features (TVD, TT, COH, and EOVS) on the response CTD.[10,65]

From the ANOVA analysis shown in Table 6, the resulting *p*-values indicate that the model is significant at

TABLE 5 Statistical error analysis of the response surface methodology (RSM) and support vector machine (SVM) models using testing datasets

Model	APRE (%)	AAPRE (%)	$E_{max.}$ (%)	$E_{min.}$ (%)	RMSE (MPa)	SD (MPa)	R
Proposed SVM	0.261	6.087	10.86	1.783	0.005	0.031	0.997
Proposed RSM	−4.828	12.70	38.47	0.829	0.033	0.138	0.991

Abbreviations: AAPRE, average absolute percent relative error; APRE, average percent relative error; RMSE, root mean square error.

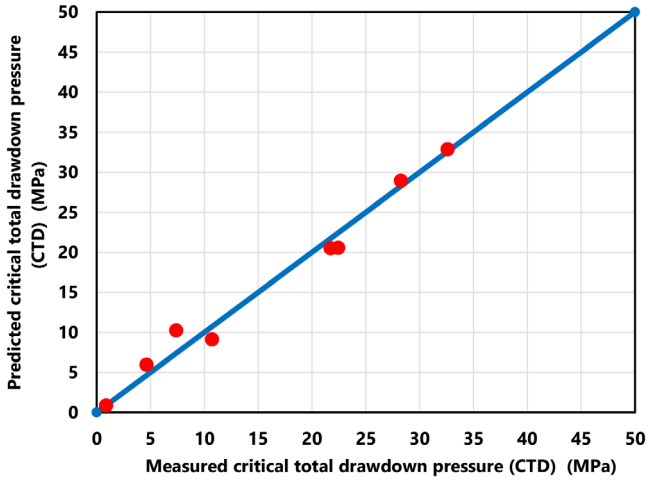

FIGURE 4 Cross-plot of the response surface methodology (RSM) correlation for the testing datasets. CTD, critical total drawdown

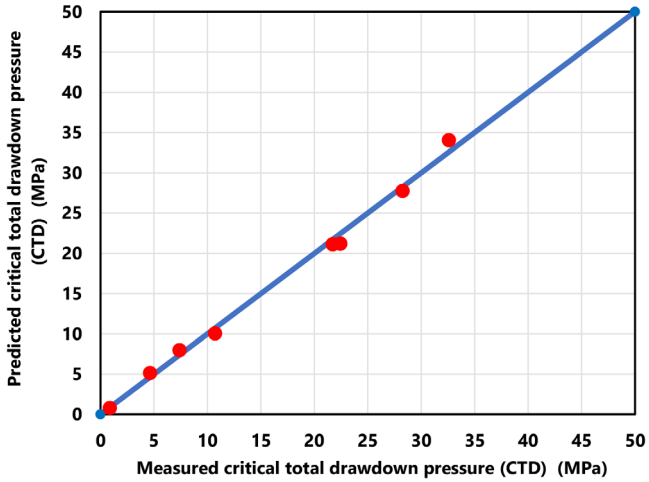

FIGURE 5 Cross-plot of the support vector machine (SVM) model for the testing datasets. CTD, critical total drawdown

a confidence level of 95%. As displayed in Table 6, the model F-value of 217.86 and the low probability (model p-value) of (<0.0001) show the significance of the CTD model. The p-value of less than 0.0500 reveals the model marking for 95% confidence intervals. However, if the p-value is greater than 0.100, a model is insignificant.[66,67] Thus, the model is significant; the model is enhanced to acquire a high accuracy and more reliability. The insignificant terms of factors are required for the RSM model and thus cannot be eliminated.[55]

3.3.2 | F-statistic test

F-test is a statistical test that signifies the similarity of variances of two samples to check if the null hypothesis is governed or not. The null hypothesis term refers to the no relationship between two measured phenomena or no association among groups. Consequently, the null hypothesis must be rejected or disproved to acquire a relationship between two variables. In statistical analysis, if the null hypothesis is rejected, the observed data are significant, and only then can the model be accepted. In contrast, the observed data were insignificant if the null hypothesis was accepted, and no relationship existed between the variables.[68] The F-distribution test is used to reject the null hypothesis. The F-distribution test can be applied by using F-calculated and F-statistic values. The F-calculated value is the F-value for the model, that is, 217.9 in the ANOVA analysis (Table 6). F-statistic values are found from F-statistic Tables (see the sample of F-statistic Tables in Table S1 in Appendix S1) by utilizing df for the model and residual, that is, 6 and 7, respectively, in the ANOVA analysis (Table 6) and α (confidence levels), that is, 0.1, 0.05, 0.01. For example, the F-statistic value of df for the model and residual, that is, 6 and 7 and $\alpha = 0.1$ denote $F_{6,7,0.1}$. The $F_{6,7,0.1}$ can be found from F-statistic Tables (Table S1 in Appendix S1) using df for model (d1) = 6, df for residual (d2) = 7 and $\alpha = 0.1$ and equals 2.827. $F_{6,7,0.05} = 3.865$ and $F_{6,7,0.01} = 7.191$ are also obtained from the F-statistic Tables for $\alpha = 0.05$ and 0.01. To reject the null hypothesis by using the F-distribution test, the F-calculated value must be greater than the F-statistic value. As shown in Table 7, all F-calculated values are higher than the F-statistic values. Therefore, the null hypothesis for the CTD model is rejected. The F-statistic test indicated that the CTD model is acceptable.

TABLE 6 Analysis of variance (ANOVA) for the critical total drawdown (CTD)

Source	Sum of squares	df	Mean square	*F*-value	*p*-value	
Model	2573	6	428.9	217.9	<0.0001	Significant
A-TVD (m)	3.480	1	3.480	1.770	0.2251	
B-TT (μsec/ft)	1.870	1	1.870	0.949	0.3624	
C-COH (MPa)	1.820	1	1.820	0.923	0.3688	
D-EOVS (MPa)	0.195	1	0.195	0.099	0.7621	
AD	0.254	1	0.254	0.129	0.7300	
BD	1.980	1	1.980	1.010	0.3488	
Residual	13.78	7	1.970			
Cor total	2587	13				

Abbreviations: COH, cohesive strength; df, degree of freedom; Cor total, the amount of variation around the mean of the observations; EOVS, effective overburden vertical stress; F, Fisher statistical value; TT, transit time; TVD, total vertical depth.

TABLE 7 *F*-distribution test for critical total drawdown (CTD)

$F_{\text{df model, df residual, }\alpha}$	*F*-statistic value < *F*-calculated value
$F_{6,7,0.1}$	2.827 < 217.9
$F_{6,7,0.05}$	3.865 < 217.9
$F_{6,7,0.01}$	7.191 < 217.9

3.3.3 | Fit statistics

The RSM correlation was evaluated using fit statistics analysis like R^2 and predicted R^2, as shown in Table 8. The R^2 value of 0.995 and the predicted R^2 value of 0.978 for the CTD model revealed a better correspondence between the expected and measurement results. The R should be at a minimum of 0.8 for a suitable fit model.[69] The R value attained for the CTD model was higher than 0.8, which indicates that the empirical model may not explain the total dissimilarity (only 2.2%). The difference between actual R^2 values and R^2 predicted values is less than 0.1, as shown in Table 8, indicating a good model correlation. Table 8 also demonstrates that the adequate precision value for the CTD model is higher than 4, which reveals good model discrimination.[67]

3.3.4 | Diagnostics plots

The diagnostic plots have been used to examine the adequacy of the model. In Design-Expert 12 software, there are some plots, such as the Box-Cox, normal, and predicted versus actual plots. Diagnostics plots are used to ensure the chosen CTD model can deliver an adequate approximation of the real system. The diagnostics plots include Box-Cox plot, the normal plot, predicted versus actual plots, and residual versus predicted plots. These plots are discussed in the Appendix S1.

3.4 | The SVM model validation

The 15 wells were applied to train and validate the SVM model using the leave-one-out cross-validation. As shown in Table 9, the SVM model has high accuracy with an R of 0.995. The SVM has a low error, with the AAPRE, APRE, RMSE, and SD of 6.060%, −1.341%, 0.005 MPa, and 0.026 MPa, respectively, as shown in Table 9. $E_{\text{max.}}$ and $E_{\text{min.}}$ for the proposed SVM model are 17.987% and 0.488%, respectively, as shown in Table 9. These statistical error analyses indicate that the proposed SVM model can be used to predict the CTD accurately.

3.5 | Trend analysis

Trend analysis is an integral part of this study. Although the trend analysis can test the model's robustness in the presence of uncertainty, it can provide comprehensive relationships between input and output variables in the models. Trend analysis can show unexpected relationships between input and output, highlighting the need to search for the models' errors. Besides, trend analysis can simplify the models by fixing model inputs that do not affect the outputs or identifying and removing redundant parts of the model structure.[70] Trend analysis is also applied to show significant connections between features and responses, guiding the development of robust models.[71]

In this study, the trend analysis checks the relationships between the inputs (TVD, TT, COH, and EOVS) and output CTD of the RSM correlation, SVM model,

TABLE 8 Validation properties of the response surface methodology (RSM) model

R^2	SD (MPa)	Mean (MPa)	Coefficient of variation (%)	Adjusted R^2	Predicted R^2	Adequate precision
0.995	1.403	16.37	8.568	0.990	0.978	44.36

TABLE 9 Statistical error analysis of the support vector machine (SVM) model for the validation datasets

APRE (%)	AAPRE (%)	$E_{max.}$ (%)	$E_{min.}$ (%)	RMSE (MPa)	SD (MPa)	R
−1.341	6.060	17.987	0.488	0.005	0.026	0.995

Abbreviations: AAPRE, average absolute percent relative error; APRE, average percent relative error; RMSE, root mean square error.

and the previous correlations to show correct trend physical behaviours. Graphs were plotted for the input parameters values (TVD, TT, COH, and EOVS) as the *x*-axis against the output CTD as the *y*-axis for the previous models and the RSM correlation and SVM model. The input parameters vary between the minimum and maximum values, while the other features are fixed at their constant mean values.[72–74]

Four input parameters (TVD, TT, COH, and EOVS) have been chosen for the trend analysis. The trend analyses were conducted for the proposed RSM correlation and SVM model and the previous correlations, that is, Kanj and Abousleiman's correlations,[32] Khamehchi et al.'s[6] multiple linear regression (MLR), and Khamehchi et al.'s[6] genetic algorithm evolved MLR (GA-MLR).

The trend of TVD has been plotted for the proposed RSM correlation and SVM model and previous models from the literature and displayed in Figure 6. Figure 6 shows that Kanj and Abousleiman's[32] correlation, established based on the COH only, showed that the CTD was independent of TVD. The RSM correlation and SVM model successfully follow the proper relationship between the TVD and the CTD. Ahad et al.[75] indicated that deeper rocks could be more consolidated. Alternatively, shallow formations could be poorly consolidated.[75] Consequently, increasing the TVD increases the CTD, indicating that sand formation is more consolidated. Therefore, sand production will be decreased by increasing the CTD and TVD.

The trend of the TT is displayed in Figure 7. The previous models show that the TT is inversely proportional to the CTD. Nevertheless, Kanj and Abousleiman's[32] correlation shows that the CTD is constant as the TT is not included. Consequently, the model did not represent the behaviour accurately. The RSM correlation and SVM model trend followed the correct trend of the TT, as shown in Figure 7, indicating that the RSM correlation and SVM model follow proper physical behaviour. The shorter TT implies that the sand is more consolidated.[76] Therefore, reducing TT will increase the CTD. Therefore, the sand production will be increased by increasing the TT.

The trends of the COH are shown in Figure 8. The COH is directly proportional to the CTD. Kanj and Abousleiman's[32] correlation followed the existing correlations, but CTD is negative (−2.57 MPa) when the COH is 0.539 MPa. Consequently, Kanj and Abousleiman's[32] correlation has not proven a proper trend for the CTD correlation. Figure 8 shows that the COH trend of the RSM correlation and SVM model agrees with the correct trend of the COH. The cohesive strength can increase the degree of cementation.[77] Increasing the cementation degree of sand grains can lead to a reduction in sand production. As a result, increasing the rock's cohesive strength increases the CTD. The sand production will be decreased by increasing the cohesive strength of the rocks.

Figure 9 shows the trends of EOVS. The CTD decreases the EOVS, as demonstrated by the RSM correlation, SVM model, and Khamehchi et al.'s[6] correlations. However, Kanj and Abousleiman's[32] correlation shows a horizontal line, indicating that their correlation did not include the EOVS parameter. The trend demonstrated by the current proposed RSM correlation and SVM model also follows the proper relationship between the EOVS and the CTD, as shown in Figure 9. In the literature, there is not any direct relationship between the EOVS and CTD explained. The overburden stress remains constant. Nevertheless, the effective overburden stress must increase when the pore pressure reduces.[77] The CTD decreases by decreasing the pore pressure.[78] Thus, increasing the EOVS will reduce the CTD. As a result, sand production increases by increasing the EOVS.

All the input parameters (TVD, TT, COH, and EOVS) of the RSM correlation and SVM model follow the proper trends compared with the existing models. Consequently, the current trend analysis represented that the proposed RSM correlation and SVM model follow the correct trends of the physical behaviours, unlike Kanj and Abousleiman's[32] correlation.

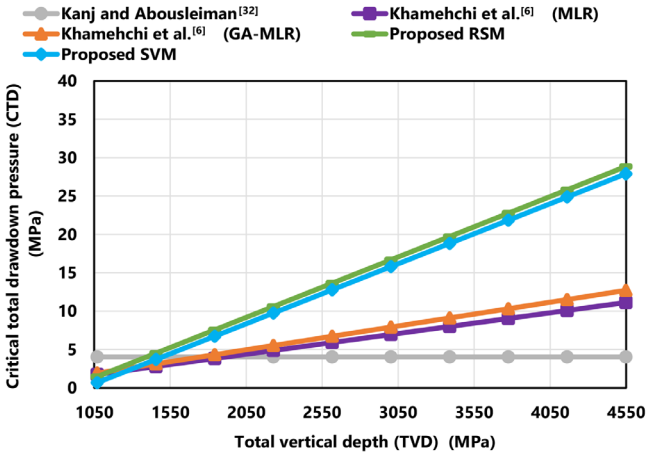

FIGURE 6 Total vertical depth (TVD) trend analysis of the response surface methodology (RSM) correlation, support vector machine (SVM) model, and previously published correlations. CTD, critical total drawdown; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression

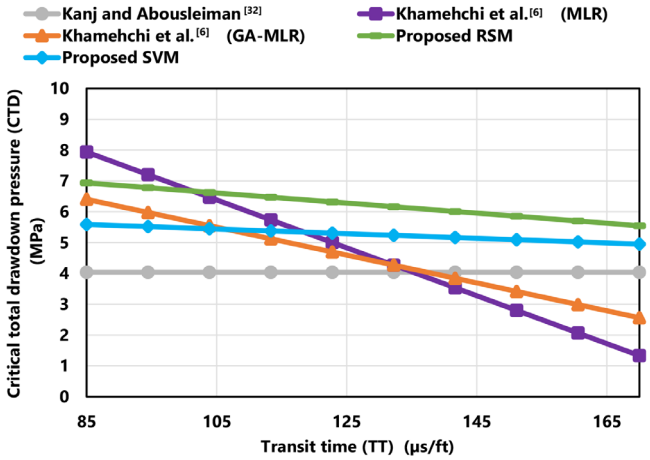

FIGURE 7 Transit time (TT) trend analysis of the response surface methodology (RSM) correlation, support vector machine (SVM) model, and previously published correlations. CTD, critical total drawdown; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression

3.6 | Comparison of RSM and SVM models with published models

Several statistical error analyses have been used to describe, validate, and compare the RSM correlation and SVM model with the previously published models. The statistical error analysis applied for the RSM correlation and SVM model and the previous models included R, SD, APRE, AAPRE, RMSE, E_{max}, and E_{min} (see the Supporting Information file: Appendix S1). In this study, the AAPRE and R are used as key indicators. Kanj and Abousleiman's[32] correlation does not follow the proper trend analyses because their correlation was built based on the COH only. Therefore, Kanj and Abousleiman's[32] correlation was not compared with the other correlations and models. On the other hand, the models and correlations in the same data ranges are compared in this study.

Figure 10 demonstrates a comparison between the AAPRE and R of the RSM correlation and SVM model with previous correlations and Alakbari et al.'s[33] model that follows the correct trend analysis. Alakbari et al.[33] stated that their fuzzy logic-based model follows the correct trend analysis and has an AAPRE of 8.6% and an R of 0.9947. The figure shows that the SVM model reveals the lowest AAPRE of 6.087% and the highest R of 0.997. Thus, the SVM model outperforms all existing models to predict the CTD. On the other hand, the RSM correlation has an AAPRE of 12.703% and an R of 0.991, making it come after the SVM and Alakbari et al.'s[33] fuzzy logic-based model; however, the RSM can be applied directly by using the equation, and there is no need to use any software to obtain the CTD. Therefore, the RSM correlation outperforms existing correlations where Khamehchi et al.'s[6] (GA-MLR) correlation resulted in an AAPRE of 22.644% and an R of 0.983. The other correlation by Khamehchi et al.[6] (MLR) showed the highest AAPRE of 30.485%.

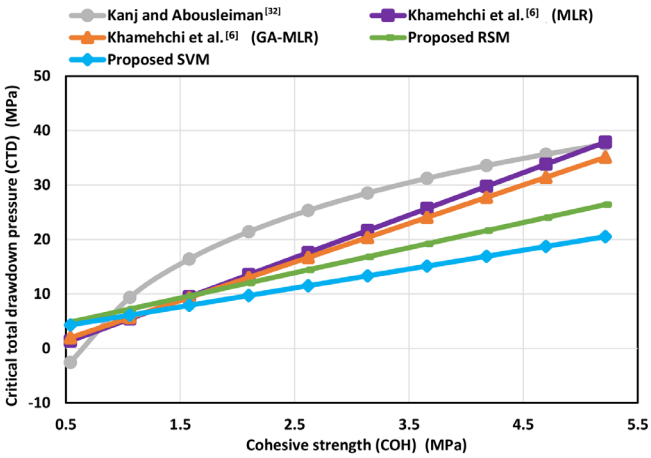

FIGURE 8 Cohesive strength (COH) trend analysis of the response surface methodology (RSM) correlation, support vector machine (SVM) model, and previously published correlations. CTD, critical total drawdown; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression

The prediction performance of the RSM correlation and SVM model was compared'against the previous correlations and model using the eight testing datasets for all models. Table 10 demonstrates the critical statistical parameters for evaluating the earlier correlations and model, the RSM correlation, and the SVM model. The RSM correlation, SVM model, and the existing correlations and model are ranked based on the AAPRE and

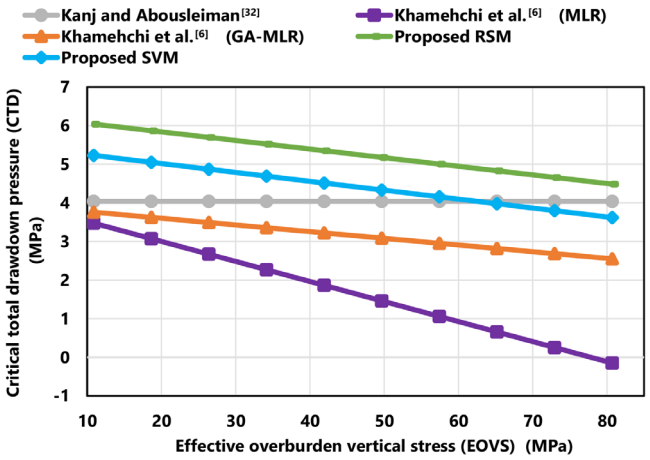

FIGURE 9 Effective overburden vertical stress (EOVS) trend analysis of the response surface methodology (RSM) correlation, support vector machine (SVM) model, and previously published correlations. CTD, critical total drawdown; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression

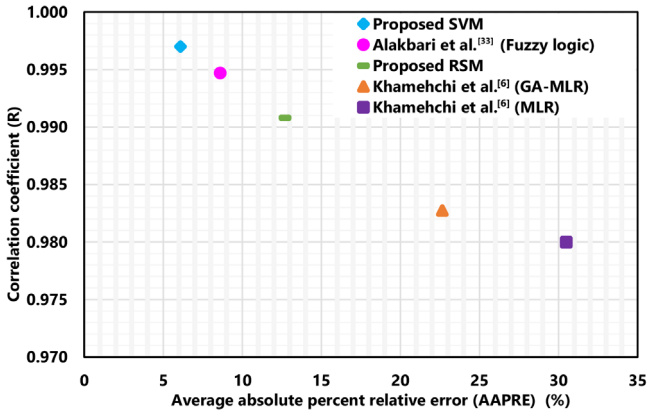

FIGURE 10 The correlation coefficient and average absolute percent relative error (AAPRE) (%) comparison of the response surface methodology (RSM) correlation and support vector machine (SVM) model with the previous correlations and model. GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression

(R), as shown in Table 10. The SVM model has the lowest APRE, RMSE, and SD compared to all models. The proposed RSM correlation is the third ranked compared to all models; however, the proposed RSM correlation outperformed all existing correlations that use direct equations. This critical comparison of all correlations and models, the RSM correlation, and the SVM model form a considerable means of evaluating the previous and RSM correlations and SVM model performance. These statistical error analyses indicate that the SVM model surpasses all published models and correlations considered in this study. In addition, the RSM correlation outperformed all existing correlations to predict the CTD. Therefore, both RSM and SVM methods can determine the CTD accurately compared to the current models and correlations.

3.7 | Effect of input variables on CTD output

The 3D surface response of the CTD model plots was utilized to assess the interactive relationships between features (TVD, TT, COH, and EOVS) and the output CTD, as shown in Figure 11A–F. The effect of each variable will be discussed in the following subsections.

3.7.1 | Effect of TVD and TT

Figure 11A displays the influence of TVD and TT on the CTD while keeping other parameters (COH and EOVS) constant. As the TVD increases, the CTD is increased (Figure 11A); the TT slightly affects the CTD (Figure 11A). This is because deeper rocks can be more consolidated. On the other hand, shallow formations can be weakly consolidated.[75] Therefore, increasing the depth will increase the CTD. This study confirms that sandstone cementation happens through a secondary geological process in which deeper formations can be tighter than shallow ones. Therefore, sand production is increased in shallow formations.[76]

3.7.2 | Effect of TVD and COH

The combined influence of TVD and COH on the CTD is shown in Figure 11B. Increasing the TVD and the COH increases the CTD, Figure 11B. According to Figure 11B, at low values of TVD, the higher the COH, the higher the CTD. Also, at low COH, increasing the TVD increases the CTD. Therefore, the TVD and the COH significantly affect the CTD, as displayed in Figure 11B. The COH increases the degree of cementation of the sand grains.[77] As a result, increasing the TVD and COH increases the sandstone cementation, and the rocks become more consolidated, so the CTD significantly increases and the sand production decreases.

3.7.3 | Effect of TVD and EOVS

The graphical illustration of the 3D surface response plot of the TVD and EOVS on the CTD is shown in Figure 11C. Figure 11C shows that an increase in the TVD and EOVS increases the CTD. Figure 11C indicates

TABLE 10 Statistical error analysis of the proposed critical total drawdown (CTD) models and the previous models

Model rank	Model	APRE (%)	AAPRE (%)	$E_{max.}$ (%)	$E_{min.}$ (%)	RMSE (MPa)	SD (MPa)	R
1	Proposed SVM	0.261	6.087	10.860	1.783	0.005	0.031	0.997
2	Alakbari et al.[33] (Fuzzy logic)	4.903	8.647	24.515	2.945	0.014	0.082	0.9947
3	Proposed RSM	−4.829	12.703	38.469	0.830	0.033	0.138	0.991
4	Khamehchi et al.[6] (GA-MLR)	16.663	22.644	63.075	6.466	0.085	0.196	0.983
5	Khamehchi et al.[6] (MLR)	23.257	30.485	109.962	5.015	0.194	0.340	0.980

Abbreviations: AAPRE, average absolute percent relative error; APRE, average percent relative error; GA-MLR, genetic algorithm evolved multiple linear regression; MLR, multiple linear regression; RMSE, root mean square error; RSM, response surface methodology; SVM, support vector machine.

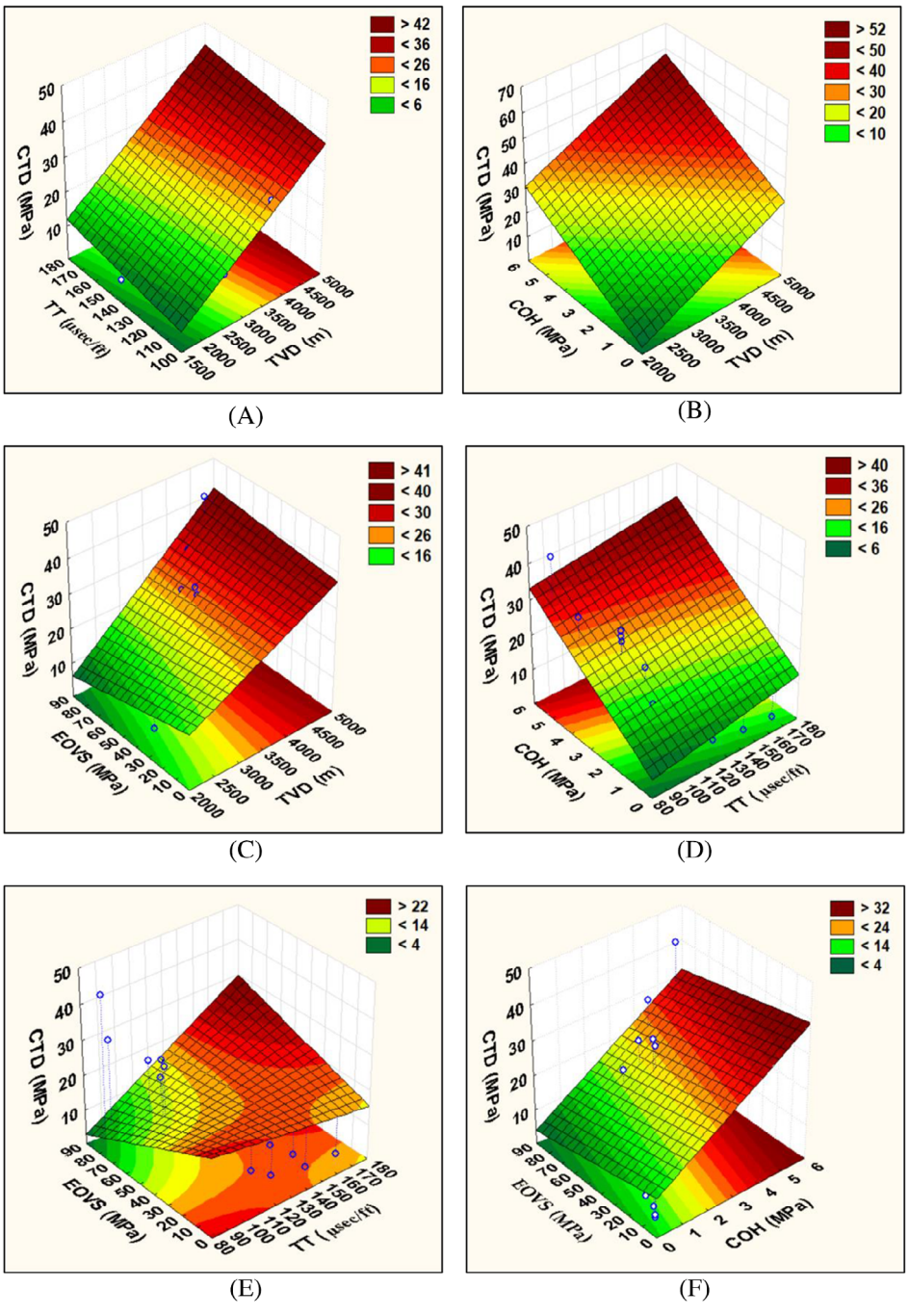

FIGURE 11 3D response surface plot of critical total drawdown (CTD) as a function of (A) total vertical depth (TVD) versus transit time (TT), (B) TVD versus cohesive strength (COH), (C) TVD versus effective overburden vertical stress (EOVS), (D) TT versus COH, (E) TT versus EOVS, and (F) COH versus (EOVS)

that the CTD increases by increasing TVD and decreases by increasing EOVS, as discussed early in the trend analysis.

3.7.4 | Effect of TT and COH

The effect of changing COH by varying the TT on the CTD and keeping other properties, TVD and EOVS, constant has been investigated. As shown in Figure 11D, it has been found that the CTD increases with increasing the COH. The cohesive strength raises the degree of cementation.[77] The increasing degree of cementation of the sand grains reduces sand production. Thus, enhancing the rock's cohesive strength improves the CTD. The TT also slightly affects the CTD, as shown in Figure 11A.

3.7.5 | Effect of TT and EOVS

The 3D response surface in Figure 11E demonstrates the combined effects of TT and EOVS on the CTD. For example, from Figure 11E, increasing the EOVS decreases the CTD at low TT, while at high TT, the EOVS increases with the CTD, as shown in Figure 11E. The effect of EOVS and TT on the CTD was explained in the trend analysis. At a high TT, the TT has more effect on the CTD than the EOVS to increase the CTD.

3.7.6 | Effect of COH and EOVS

Figure 11F illustrates the 3D response surface plot of the CTD dependency on COH and EOVS. Other parameters (i.e., TVD and TT) were kept constant in this case. Figure 11F shows the increase of COH with CTD. Therefore, the rock strength increases by increasing the COH,[79] and sand production decreases with the rise in COH. The increasing EOVS decreases the CTD, as it was clarified in the trend analysis study.

4 | CONCLUSIONS

The CTD is an essential indicator of the onset of sand production. Some correlations and models are used to determine the CTD in the literature. Nonetheless, the previous methods have limitations, such as a lack of accuracy and the use of specific software. Moreover, the published correlations did not include trend analysis to prove the correct relationships between the input parameters, that is, total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS) and output parameter, that is, CTD, to show the proper physical behaviour. Therefore, this study aims to build new, accurate, and robust models using RSM and SVM. The proposed RSM and SVM models were validated using different methods, that is, ANOVA, *F*-statistics test, fit statistics, and diagnostics plots to prove that the proposed models are more robust. The proposed models were also validated by applying the trend analysis to indicate the proper relationships between all input parameters and output. In addition, various statistical error analyses, such as AAPRE, maximum absolute percent relative error ($E_{\max.}$), minimum absolute percent relative error ($E_{\min.}$), root mean square error (RMSE), standard deviation (SD), and correlation coefficient (R) have been performed to describe, validate, and compare the RSM correlation and SVM model with the existing models. The conclusions can be highlighted as follows:

- The ANOVA, *F*-statistics test, fit statistics, and diagnostics plots showed that the RSM correlation is more reliable with high accuracy compared with the published correlations. The RSM correlation decreased the AAPRE from 22.644% to 12.703% compared to all existing correlations.
- The trend analysis of the proposed RSM and SVM models shows an accurate description of the CTD profile as a function of all the considered parameters (TVD, TT, COH, and EOVS). Thus, the predicted profile follows the actual trend, as expected from the physical relationship.
- SVM has some merits: non-convergence to local minima, accurate generalization, and predictive ability for small and non-linear datasets. The SVM model has higher accuracy in predicting CTD than previously published models and the RSM correlation. Furthermore, SVM has good generalization capabilities, preventing it from over-fitting.
- The following values confirm the accuracy of the SVM model; the highest correlation coefficient of 0.997, the lowest AAPRE of 6.087%, the lowest APRE of 0.261%, the lowest RMSE of 0.005, the lowest SD of 0.031, the lowest $E_{\min.}$ of 1.783%, and the lowest $E_{\max}$ of 10.860%. These values are the best compared with the RSM correlation and the published correlations and model.
- The effects of the independent variables (TVD, TT, COH, and EOVS) on the CTD are shown in 3D response surface plots. The interaction between variables in the system was statistically significant. According to the 3D response surface plots, as the TVD increases and the other variables are held constant, the CTD increases. Thus, deeper sediments are more

consolidated. On the other hand, shallow rocks are weakly consolidated formations. The increase in COH increases CTD as the COH increases, and the degree of cementation grows. Thus, the increase in COH decreases sand production. The decreasing TT indicates that the rock is hard and more consolidated. Therefore, sand production can be increased by reducing the CTD and increasing the TT. Increasing the EOVS will decrease the CTD. As a result, the sand production increases by increasing the EOVS.

4.1 | Limitations and recommendations

The proposed RSM and SVM models were built based on four input parameters, that is, TVD, TT, COH, and EOVS, to determine the CTD and at their given data ranges. Therefore, the proposed models will show high accuracy in their inputs' ranges. Nevertheless, the proposed models were proven to be more robust and accurate than all literature models. Considering other parameters that have effects on the CTD and increasing the datasets from different places are recommended to enhance the CTD model. In addition, other machine learning methods can be applied to determine the CTD.

NOMENCLATURE

Latin synonyms

R	correlation coefficient
$E_{max.}$	maximum absolute percent relative error
$E_{min.}$	minimum absolute percent relative error
Y	response
3D	three-dimensional
k	number of the features ($k = 5$ in this work)
x_i, x_j	variables (i and j from 1 to k)
b_o	model intercept coefficient
b_j, b_{jj}, and b_{ij}	interaction coefficients of the linear, quadratic, and second-order terms, respectively
e	error
w and b	adjustable model parameters
c	constant
df	degree of freedom
P	probability
F	Fisher statistical value
d1	df for a model from the ANOVA table
d2	df for residual from the ANOVA table

Greek synonyms

μsec	microsecond
α	confidence levels

Abbreviations

AAPRE	average absolute percent relative error
ANNs	artificial neural networks
ANOVA	analysis of variance
APRE	average percent relative error
CCD	central composite design
COH	cohesive strength
CTD	critical total drawdown
DOE	design of experiments
EOVS	overburden vertical stress
FEM	finite element method
GA-MLR	genetic algorithm evolved multiple linear regression
ML	machine learning
MLR	multiple linear regression
MPa	megapascal (pressure unit)
RBF	radial basis function
RMSE	root mean square error
RSM	response surface methodology
SD	standard deviation
SMO	sequential minimal optimization
SRM	structural risk minimization
SVM	support vector machine
TT	transit time
TVD	total vertical depth

AUTHOR CONTRIBUTIONS

Fahd Saeed Alakbari: Conceptualization; data curation; formal analysis; methodology; writing – original draft; writing – review and editing. **Mysara Eissa Mohyaldinn:** Conceptualization; investigation; methodology; project administration; resources; supervision; writing – original draft; writing – review and editing. **Mohammed Abdalla Ayoub:** Conceptualization; formal analysis; investigation; methodology; writing – original draft; writing – review and editing. **Ali Samer Muhsan:** Conceptualization; formal analysis; investigation; methodology; writing – original draft; writing – review and editing. **Said Jadid Abdulkadir:** Conceptualization; formal analysis; methodology; writing – original draft; writing – review and editing. **Ibnelwaleed A. Hussein:** Conceptualization; formal analysis; investigation; writing – original draft; writing – review and editing. **Abdullah Abduljabbar Salih:** Conceptualization; formal analysis; investigation; methodology; writing – original draft.

ACKNOWLEDGEMENTS

The authors would like to thank the Yayasan Universiti Teknologi PETRONAS (YUTP FRG Grant No: 015LC0-428) at Universiti Teknologi PETRONAS financial support. Open access funding is provided by the Qatar National Library.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cjce.24640.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Fahd Saeed Alakbari https://orcid.org/0000-0002-3227-698X
Ibnelwaleed A. Hussein https://orcid.org/0000-0002-6672-8649

REFERENCES

[1] N. Alireza, V. Hans, B. Hadi, I. Rafiqul, presented at the SPE Asia Pacific Oil and Gas Conf. and Exhibition, Jakarta, Indonesia, September **2003**.
[2] F. S. Alakbari, M. E. Mohyaldinn, A. S. Muhsan, N. Hasan, T. Ganat, *Polymers* **2020**, *12*, 1069.
[3] M. E. Mohyaldinn, M. C. Ismail, N. Hasan, *Advances in Material Sciences and Engineering*, Springer, Kuala Lumpur, Malaysia **2020**, p. 287.
[4] D. Kanesan, M. E. Mohyaldinn, N. I. Ismail, D. Chandran, C. J. Liang, *J. Nat. Gas Sci. Eng.* **2019**, *65*, 267.
[5] A. Alghurabi, M. Mohyaldinn, S. Jufar, O. Younis, A. Abduljabbar, M. Azuwan, *J. Nat. Gas Sci. Eng.* **2020**, *85*, 103706.
[6] E. Khamehchi, I. R. Kivi, M. Akbari, *J. Pet. Sci. Eng.* **2014**, *123*, 147.
[7] Z. Fan, D. Yang, X. Li, *SPE J.* **2019**, *24*, 988.
[8] A. K. Abbas, H. A. Baker, R. E. Flori, H. Al-Hafadhi, N. Al-Haideri, presented at 53rd US Rock Mechanics/Geomechanics Symp., New York City, NY, June **2019**.
[9] M. P. Tixier, G. W. Loveless, R. A. Anderson, *J. Pet. Technol.* **1975**, *27*, 283.
[10] C. A. M. Veeken, D. R. Davies, C. J. Kenter, A. P. Kooijman, presented at SPE Annual Technical Conf. and Exhibition, Dallas, TX, October **1991**.
[11] Darmadi, T. S. Y. Choong, T. G. Chuah, R. Yunus, Y. H. T. Yap, *Can. J. Chem. Eng.* **2009**, *87*, 591.
[12] W.-W. Liu, A. Aziz, S.-P. Chai, A. R. Mohamed, C.-T. Tye, *Can. J. Chem. Eng.* **2012**, *90*, 489.
[13] J. Nam, J. W. Kim, J. S. Kim, J. Lee, S. W. Lee, *Procedia Manufacturing* **2018**, *26*, 403.
[14] L. Salehnezhad, A. Heydari, M. Fattahi, *J. Mol. Liq.* **2019**, *276*, 417.
[15] K. E. H. K. Ishak, M. A. Ayoub, *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 2927.
[16] A. A. Umar, I. M. Saaid, A. A. Sulaimon, R. M. Pilus, *Journal of Applied Mathematics* **2020**, *2020*, 1.
[17] F. S. Alakbari, M. E. Mohyaldinn, M. A. Ayoub, A. S. Muhsan, A. Hassan, *Colloids Surf., A* **2021**, *616*, 126278.
[18] M. Alhajabdalla, H. Mahmoud, M. S. Nasser, I. A. Hussein, R. Ahmed, H. Karami, *ACS Omega* **2021**, *6*, 2513.
[19] A. Zhang, Z. Fan, L. Zhao, C. He, *J. Pet. Sci. Eng.* **2021**, *196*, 107690.
[20] A. M. Awad, I. A. Hussein, M. S. Nasser, S. A. Ghani, A. O. Mahgoub, *J. Pet. Sci. Eng.* **2021**, *208*, 109613.
[21] S. Zendehboudi, N. Rezaei, A. Lohi, *Appl. Energy* **2018**, *228*, 2539.
[22] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, H. Oza, *Pet. Res.* **2021**, *6*, 379.
[23] W. Ji, D. Liu, Y. Meng, Y. Xue, *Evolutionary Intelligence* **2020**, *14*, 1.
[24] A. Kamari, A. Bahadori, A. H. Mohammadi, S. Zendehboudi, *Pet. Sci. Technol.* **2014**, *32*, 2961.
[25] O. O. Olatunji, O. Micheal, presented at SPE Nigeria Annual Int. Conf. and Exhibition, Lagos, Nigeria, July **2017**.
[26] E. A. El-Sebakhy, *J. Pet. Sci. Eng.* **2009**, *64*, 25.
[27] K. Al-Azani, S. Elkatatny, A. Abdulraheem, M. Mahmoud, A. Ali, presented at SPE Kingdom of Saudi Arabia Annual Technical Symp. and Exhibition, Dammam, Saudi Arabia, April **2018**.
[28] K. Al-Azani, S. Elkatatny, A. Ali, E. Ramadan, A. Abdulraheem, *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 2769.
[29] H. Cheng, F. Wang, L. Huo, G. Song, *Structural Health Monitoring* **2020**, *19*, 2075.
[30] W. Li, D. Yue, L. Colombera, Y. Du, S. Zhang, R. Liu, W. Wang, *J. Pet. Sci. Eng.* **2021**, *196*, 107749.
[31] Q. Yasin, G. M. Sohail, P. Khalid, S. Baklouti, Q. Du, *J. Pet. Sci. Eng.* **2021**, *197*, 107975.
[32] M. Y. Kanj, Y. Abousleiman, presented at SPE Annual Technical Conf. and Exhibition, Houston, TX, October **1999**.
[33] F. S. Alakbari, M. E. Mohyaldinn, M. A. Ayoub, A. S. Muhsan, I. A. Hussein, *PLoS One* **2021**, *16*, e0250466.
[34] G. Moricca, G. Ripa, F. Sanfilippo, F. J. Santarelli, presented at Rock Mechanics in Petroleum Engineering, Delft, The Netherlands, August **1994**.
[35] A. M. Molinaro, R. Simon, R. M. Pfeiffer, *Bioinformatics* **2005**, *21*, 3301.
[36] G. C. Cawley, *The 2006 IEEE Int. Joint Conf. on Neural Network Proceedings*, IEEE, Piscataway, NJ **2006**.
[37] K. R. Rad, A. Maleki, *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **2020**, *82*, 965.
[38] R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, S. M. Kowalski, *Journal of Quality Technology* **2004**, *36*, 53.
[39] R. H. Myers, D. C. Montgomery, C. M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, Hoboken, NJ **2016**.
[40] D. C. Montgomery, E. A. Peck, G. G. Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, Hoboken, NJ, **2021**.
[41] S. P. J. N. Senanayake, F. Shahidi, *Food Chemistry* **2002**, *77*, 115.
[42] İ. H. Boyacı, *Biochem. Eng. J.* **2005**, *25*, 55.
[43] H. Ceylan, S. Kubilay, N. Aktas, N. Sahiner, *Bioresour. Technol.* **2008**, *99*, 2025.
[44] A. Asghar, A. Raman, A. Aziz, W. M. A. W. Daud, *Sci. World J.* **2014**, *2014*, 869120.
[45] I. Arslan-Alaton, G. Tureli, T. Olmez-Hanci, *J. Photochem. Photobiol., A* **2009**, *202*, 142.
[46] B. Haq, *Polymers* **2021**, *13*, 3269.

[47] E. R. Dawud, A. K. Shakya, *Arabian J. Chem.* **2019**, *12*, 718.
[48] L. Deng, C. S. Cai, *Journal of Bridge Engineering* **2010**, *15*, 553.
[49] D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, Hoboken, NJ **2017**.
[50] M. Y. Ong, S. Nomanbhay, F. Kusumo, R. M. H. Raja Shahruzzaman, A. H. Shamsuddin, *Energies* **2021**, *14*, 295.
[51] A. Kazemzadeh, F. Ein-Mozaffari, A. Lohi, L. Pakzad, *Can. J. Chem. Eng.* **2016**, *94*, 2394.
[52] C. Cortes, V. Vapnik, *Machine Learning* **1995**, *20*, 273.
[53] S. He, X. Liu, Y. Wang, S. Xu, J. Lu, C. Yang, S. Zhou, Y. Sun, W. Gui, W. Qin, *Can. J. Chem. Eng.* **2017**, *95*, 2357.
[54] K. O. Akande, T. O. Owolabi, S. Twaha, S. O. Olatunji, *IOSR Journal of Computer Engineering* **2014**, *16*, 88.
[55] L.-L. Li, X. Zhao, M.-L. Tseng, R. R. Tan, *J. Cleaner Prod.* **2020**, *242*, 118447.
[56] B. Richhariya, M. Tanveer, *Pattern Recognition* **2020**, *102*, 107150.
[57] K. Popli, V. Maries, A. Afacan, Q. Liu, V. Prasad, *Can. J. Chem. Eng.* **2018**, *96*, 1532.
[58] W.-H. Chen, S.-H. Hsu, H.-P. Shen, *Computers and Operations Research* **2005**, *32*, 2617.
[59] H. Chen, L. Xu, W. Ai, B. Lin, Q. Feng, K. Cai, *Sci. Total Environ.* **2020**, *714*, 136765.
[60] M. Amroune, T. Bouktir, I. Musirin, *Arabian J. Sci. Eng.* **2018**, *43*, 3023.
[61] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification*, Springer, Berlin **2016**, p. 207.
[62] M. Alida, M. Mustikasari, *Jurnal Online Informatika* **2020**, *5*, 53.
[63] H. Akoglu, *Turkish Journal of Emergency Medicine* **2018**, *18*, 91.
[64] P. Kirmizakis, C. Tsamoutsoglou, B. Kayan, D. Kalderis, *J. Environ. Manage.* **2014**, *146*, 9.
[65] F. Z. Derdour, M. Kezzar, L. Khochemane, *Powder Technol.* **2018**, *339*, 846.
[66] M. Zabeti, W. M. A. W. Daud, M. K. Aroua, *Fuel Process. Technol.* **2010**, *91*, 243.
[67] M. J. Anderson, P. J. Whitcomb, *Third Edition DOE Simplified: Practical Tools for Effective Experimentation*, CRC Press, Boca Raton, FL **2017**.
[68] B. S. Everitt, A. Skrondal, *The Cambridge Dictionary of Statistics*, 4th ed., Cambridge, UK **2010**.
[69] A. M. Joglekar, A. T. May, *Cereal Foods World* **1987**, *32*, 857.
[70] D. J. Pannell, *Agricultural Economics* **1997**, *16*, 139.
[71] M. C. Hill, C. R. Tiedeman, *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, John Wiley & Sons, Hoboken, NJ **2006**.
[72] A. A. Al-Shammasi, presented at Middle East Oil Show Conf., Bahrain, February **1999**.
[73] S. A. Osman, M. A. Ayoub, M. A. Aggour, presented at the SPE Middle East Oil and Gas Show and Conf., Kingdom of Bahrain, March **2005**.
[74] M. A. Ayoub, S. N. Zainal, M. E. Elhaj, K. Ishak, K. E. Hani, Q. Ahmed, presented at Int. Petroleum Technology Conf., Dhahran, Kingdom of Saudi Arabia, January **2020**.
[75] N. A. Ahad, M. Jami, S. Tyson, *J. Pet. Explor. Prod. Technol.* **2020**, *10*, 1675.
[76] H. Ben Mahmud, V. H. Leong, Y. Lestariono, *Petroleum* **2020**, *6*, 1.
[77] B. Aadnoy, R. Looyeh, *Petroleum Rock Mechanics: Drilling Operations and Well Design*, Gulf Professional Publishing, Amsterdam, The Netherlands **2019**.
[78] C. Yan, J. Deng, X. Lai, L. Hu, Z. Chen, *Indian Geotech. J.* **2014**, *44*, 101.
[79] R. Zhong, *Adv. Pet. Explor. Dev.* **2014**, *7*, 57.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: F. S. Alakbari, M. E. Mohyaldinn, M. A. Ayoub, A. S. Muhsan, S. J. Abdulkadir, I. A. Hussein, A. A. Salih, *Can. J. Chem. Eng.* **2023**, *101*(5), 2493. https://doi.org/10.1002/cjce.24640