Arabian Journal for Science and Engineering (2025) 50:4993–5005
https://doi.org/10.1007/s13369-024-09556-8

RESEARCH ARTICLE-PETROLEUM ENGINEERING

An Accurate Critical Total Drawdown Prediction Model for Sand Production: Adaptive Neuro-fuzzy Inference System (ANFIS) Technique

Fahd Saeed Alakbari[1] · Syed Mohammad Mahmood[1,2] · Mysara Eissa Mohyaldinn[1,2] · Mohammed Abdalla Ayoub[3] · Ibnelwaleed A. Hussein[4,5] · Ali Samer Muhsan[6] · Abdullah Abduljabbar Salih[1] · Azza Hashim Abbas[7]

Received: 1 February 2024 / Accepted: 28 August 2024 / Published online: 23 September 2024

Abstract

Sand production causes many problems in the petroleum industry. The sand production is predicted to control it in the early stages. Therefore, accurate prediction of sand production has been considered substantial in achieving successful sand control. Critical total drawdown (CTD) can indicate the sand production. The main drawback of the previous studies in predicting CTD is their lack of accuracy. Thus, this study aims to develop an accurate CTD estimation prediction model employing a trend analysis and adaptive neuro-fuzzy inference system (ANFIS). The method is chosen because of its higher performance; the model is built based on 23 published datasets from the Adriatic Sea. The developed ANFIS model is evaluated using various methods, namely, trend analyses. Trend analyses are conducted to show the effects of the features on the CTD to present the physical behavior. The model's performance was also evaluated using statistical error analyses. In addition, the ANFIS and previously published models were assessed. The trend analyses show the correct relationship between all features and the CTD. In addition, the trend analyses for the previous models are discussed. The results show that the proposed ANFIS method outperforms published methods with an R of 0.9984 and an absolute average percentage relative error (AAPRE) of 4.293%.

Keywords Sand control · Machine learning · Adaptive neuro-fuzzy inference system technique · ANFIS · Artificial intelligence · Critical total drawdown

Abbreviations

ANFIS	Adaptive neuro-fuzzy inference system
CTD	Critical total drawdown
TD	Total drawdown
TVD	Total vertical depth
TT	Transit time
COH	Cohesive strength
EOVS	Overburden vertical stress
RSM	Response surface methodology
SVM	Support vector machine
ML	Machine learning
GRNN	Generalized regression neural network
ANN	Artificial neural network
MLR	Multiple linear regression

✉ Fahd Saeed Alakbari
fahd.akbari@utp.edu.my

✉ Ibnelwaleed A. Hussein
ihussein@qu.edu.qa

1 Center of Flow Assurance, Institute of Subsurface Resources, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

2 Petroleum Engineering Department, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

3 Chemical and Petroleum Engineering Department, United Arab Emirates University, Al Ain, United Arab Emirates

4 Gas Processing Center, College of Engineering, Qatar University, P. O. Box 2713, Doha, Qatar

5 Department of Chemical Engineering, College of Engineering, Qatar University, P.O. Box 2713, Doha, Qatar

6 Mechanical Engineering Department, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

7 School of Mining and Geosciences, Nazarbayev University, Nur Sultan 010000, Kazakhstan

GA-MLR	Genetic algorithm-evolved multiple linear regression
FL	Fuzzy logic
APRE	Average percent relative error
AAPRE	Average absolute percent relative error
RMSE	Root-mean-square error
SD	Standard deviation
MPa	Megapascal (pressure unit)
R	Correlation coefficient
$E_{\max.}$	Maximum absolute percent relative error
$E_{\min.}$	Minimum absolute percent relative error
MMP	Minimum miscibility pressure
Pb	Reservoir bubble-point pressure
TrndAnal	Trend analysis
PhyBhvr	Physical behavior
µs	Microsecond
SEA	Statistical error analyses

1 Introduction

Sand production creates many economic and technical issues during production from hydrocarbon reservoirs, and as such, it is a topic of high interest in the petroleum industry. The sand in the flow stream can erode the tubings, valves, and other production facilities, to mention a few. Frequent breakdowns and routine maintenance thus add to the cost due to the lost production time and the hardware cost. It is often a complex problem that needs customization based on the reservoir type [1].

Sand production happens when formation stress exceeds its strength, causing rock failure due to tectonic activities, overburden pressure, pore pressure, stress induced during drilling, and producing fluid's drag forces. The two major factors influencing sand production include fluid flow and rock strength [2].

Sand production is prevalent in unconsolidated or poorly consolidated formations due to their low compressive strength. Such rocks are globally found in younger, shallow sedimentary formations. Acidizing can further weaken the formations, increasing the risk of sand production [3].

Increasing the production rate of reservoir fluids increases the pressure drawdown between the reservoir and wellbore flowing pressures, increasing sand production. The reservoir hydrocarbon fluid production may cause pressure frictional loss, and frictional forces (because of potential and kinetic energy) could create stresses that exceed the formation's compressive strength. To control this, production wells often have a critical flow rate, below which sand production does not occur. This critical flow rate is determined by gradually increasing the production rate until sand production is detected. A choke valve can limit the production rate below the necessary level to manage sand production. However, this rate is sometimes significantly lower than the desired fluid production rate for the well [4].

As reservoir pressure reduces over time, increased adequate overburden pressure increases the stress on the sand, potentially crushing formation particles and causing sand production. The formation sand may be damaged when the effective stress surpasses the formation strength due to the reservoir rock's compaction from the decreased formation pore pressure [3].

The frictional force on sand particles is proportional to the fluid viscosity and flow rate. High viscous drag can cause sand production from heavy oil reservoirs even at low flow rates [3].

An increase in the water cut (a higher proportion of water) also increases sand production. Two mechanisms explain this behavior. First, particle-to-particle cohesiveness is provided via the surface tension of connate water surrounding each sand particle in water-wet sandstone formations. As water is produced, this surface tension reduces, decreasing cohesiveness and weakening the sand arch around the perforation, producing sand. The second mechanism influencing sand production is the reduced mobility of oil when the saturation of oil in the reservoir is decreased, thus reducing the channels available for oil flow. As this happens, increasing differential pressure is needed to maintain the desired hydrocarbon flow rates, potentially leading to sand production [3].

The total drawdown (TD) represents the difference in pressure between the reservoir and the flowing bottom-hole pressure. When the reservoir pressure is static, meaning that there is no fluid production, the pressure at the sand surface equals the static reservoir pressure. When fluid is produced, the reservoir is in a dynamic state. The pressure at the perforations of the wellbore (sand face) has to be lowered from the static reservoir pressure to allow the fluid to flow toward the wellbore. The TD captures the dynamic state, showing the differential between the static reservoir pressure and the flowing bottom-hole pressure. The TD shows how much pressure is lost because of the fluid flow from the reservoir toward the wellbore. After production stops, the flowing bottom-hole pressure increases as the reservoir pressure stabilizes and returns to its static state. This process is identified as pressure build-up. The TD at which the sand production begins is defined as the critical total drawdown (CTD) [5].

Some methods are applied to control sand production, namely, mechanical and chemical techniques. The mechanical methods include frac-packing and open-hole gravel packs. The chemical method includes injecting some chemicals around the wellbore to consolidate the sand formations. Therefore, the chemical techniques may reduce the permeability near the wellbore [6]. Due to the cost and decreasing permeability of the mechanical and chemical methods, early sand production prediction is required to control sand production successfully [7].

As discussed above, the CTD can effectively predict sand production. Some studies used data-driven approaches to determine CTD. Kanj and Abousleiman [8] used 31 wells of the Adriatic Sea and generalized regression neural network (GRNN) based on cohesive strength (COH) to calculate the CTD. Khamehchi et al. [7] used multiple linear regression (MLR) and genetic algorithm-evolved MLP (GA-MLR) to predict the CTD. They utilized total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS) parameters as inputs. Alakbari et al. [9] used the same four inputs to determine the CTD by applying fuzzy logic (FL). Alakbari et al. [10] found the CTD using different methods, namely, response surface methodology. They stated that their models are the most accurate compared to the previous studies [10].

The ANFIS method has some benefits, such as presenting a reputable performance. It has a better learning capability; it involves few adjustable parameters. Its networks display a well-structured knowledge representation because its structure permits parallel computation [11]. It has artificial neural networks (ANN) and FL in a single method to conduct the procedure superbly and produce a faster decision [12]. Due to its fewer components that are primed with variables connected to the problem, it can decrease the training time [13].

The ANFIS was utilized successfully in numerous engineering determinations. Ayoub et al. [14] established a technique to find a penetration rate employing the ANFIS. Sambo et al. [15] and Hamdi and Chenxi [16] created ANFIS models to establish water saturation and CO_2 minimum miscibility pressure (MMP). Ayoub et al. [17] utilized the ANFIS to determine the isothermal oil compatibility under bubble-point pressure. Kalam et al. [18] utilized the ANFIS to obtain the relative permeability, and Ayoub et al. [19] used the ANFIS to obtain minimum miscibility pressure. The result demonstrated that the ANFIS model surpassed other models [19]. Recent studies used the ANFIS models to determine the pressure–volume–temperature (PVT), namely, reservoir bubble-point pressure (Pb), a gas–oil ratio below the Pb, and oil formation volume factor [20–22]. The ANFIS was utilized to obtain porosity and permeability based on the well-log data [23]. The study proved that the ANFIS method could be a powerful tool for getting porosity and permeability [23].

As sand production mechanisms were discussed, many factors that affect sand production were highlighted. It was also surmised that sand production could be effectively predicted by CTD, which was described to be based on several factors, most notably the TVD, TT, COH, and EOVS. Thus, CTD was chosen in this study to predict sand production while considering the mentioned inputs.

Several models were applied to obtain the CTD; however, it was felt that a more robust and accurate model was still warranted for predicting the CTD. Thus, previously published models were explored to improve them. It was observed that the ANFIS method yields better performance; therefore, it was chosen to enhance CTD predictive ability. It was evaluated using different procedures, namely, trend analysis (TrndAnal), to show the physical behavior (PhyBhvr) or the effects between the features and the CTD. Statistical error analyses (SEA) were employed to represent the models' accuracy. Moreover, the proposed model was compared with the current studies.

2 Methodology of ANFIS Model Development

Here in this section, the significant steps taken for the ANFIS model development are briefly explained.

2.1 Data Description

The datasets were gathered from the literature [24], which included 23 gas wells from the Adriatic Sea. The parameters that were collected were TVD, TT, COH, and EOVS as inputs and CTD as output to predict sand production. We tried our best to increase the dataset size. This study contains one of the highest datasets among all studies in the literature that predict CTD based on TVD, TT, COH, and EOVS.

The ANFIS method has been shown to perform satisfactorily on small datasets [25–27]. That is one of the reasons an ANFIS model was developed in this study to predict the CTD. Other benefits of using the ANFIS model will be discussed in the following subsection: 2.2 ANFIS model.

The dataset ranges are displayed in Tables 1 and 2. The datasets were split 70% for training the proposed ANFIS model and 30% for testing the previous and proposed ANFIS models. The same testing datasets were used to evaluate the previous and proposed ANFIS models for a fair comparison.

2.2 Model Description

The ANFIS approach excels in rapid decision-making between parameters because it combines ANN and FL [12, 28]. Figure 1 demonstrates the ANFIS's structure, which has five layers: fuzzification, rule, normalization, defuzzification, and output layers. It is a hybrid learning rule, a supervised learning capability using a multilayer feedforward neural network [29, 30].

The ANFIS method has some benefits over other machine learning approaches. It is an amalgamation of ANN and Fl to make the system higher in reaching a faster decision

Table 1 The training dataset's statistical analysis

Parameter	TVD (m)	TT (µs/ft)	COH (MPa)	EOVS (MPa)	CTD measured (MPa)
Minimum	1070	85	0.539	10.88	0.314
Maximum	4548	170	5.22	80.71	43.97
Mean	2565	115	1.77	38.16	15.28
Median	2380	110	1.27	29.42	12.80
Range	3478	85	4.68	69.82	43.65
Skewness	0.187	0.940	1.23	0.398	0.600
Standard deviation	10.2	0.208	0.012	0.228	0.123

Table 2 The testing dataset's statistical analysis

Parameter	TVD (m)	TT (µs/ft)	COH (MPa)	EOVS (MPa)	CTD measured (MPa)
Minimum	1122	85	0.559	11.3	0.883
Maximum	4088	150	3.87	76.6	32.6
Mean	2618	116	1.81	41.0	16.3
Median	2660	108	1.62	41.9	16.2
Range	2966	65	3.31	65.3	31.7
Skewness	− 0.008	0.387	0.73	0.065	0.125
Standard deviation	10.569	0.229	0.01	0.265	0.119

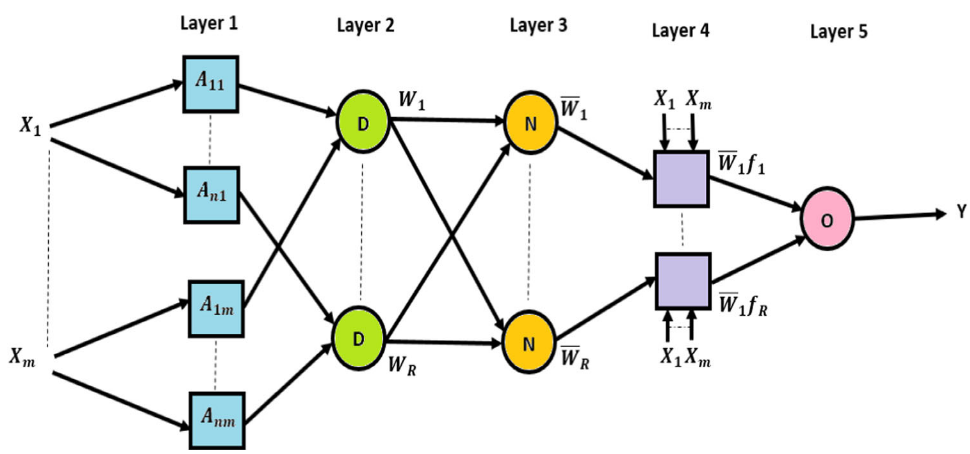

Fig. 1 The ANFIS's structure workflow

about the mapped relationship between the inputs and outputs [12]. It has fewer parameters compared to others. It has proven higher learning capabilities and conducts highly nonlinear mappings. It controls noisy data and finds classification and regression problems with missing input data [31]. The ANFIS structure is designed for parallel computation and offers a well-structured knowledge representation that integrates well with other design systems [11]. It can quickly decide the mapped connection between parameters [12].

The ANFIS model was created using MATLAB R2020a. The Gaussian curve membership is used to show the optimum result for predicting the CTD. The membership functions for all factors are determined based on domain information. The optimum parameters to accurately predict the CTD are presented in Table 3. The defuzzification methods were used in this study. The most critical parameter in the ANFIS model is the cluster center's range of influence, and its value in this study is 0.459. The number of fuzzy rules and training epoch are 10 and 24, respectively. The initial step size is 0.2461. The step size decrease and increase rates are 0.11 and 2, respectively. The optimum hyperparameters of the proposed ANFIS model were selected using the grid search method, which is a process that systematically finds the optimum hyperparameters through a specified subset of the hyperparameter space of

Table 3 The optimum hyperparameters of the proposed ANFIS model

Parameter	Description/value
Fuzzy structure	Sugeno type
Initial FIS for training	genfis2
Membership function type	Dsigmf
Cluster center's range of influence	0.459
Number of inputs	4
Number of outputs	1
Optimization method	Hybrid
Number of fuzzy rules	10
Training epoch number	24
Initial step size	0.2461
Step size decrease rate	0.11
Step size increase rate	2

the algorithm [32]. In this grid search method, each hyperparameter, such as the cluster center's range of influence, changes its types or values and keeps the other hyperparameters fixed. Using this approach, the ANFIS model's accuracy and the trend analysis were assessed. Finally, the optimum hyperparameters of the proposed ANFIS model were selected based on the correct trend analysis and highest accuracy.

2.3 Model Evaluation

The ANFIS model to predict the CTD was evaluated by applying the following procedure. Initially, the TrndAnal was conducted. After the parameters showed the correct PhyBhvr, statistical error analyses (SEA) were determined for the ANFIS and previously published models to show the models' accuracy. The SEA was discussed in the supporting information. Finally, all models were compared and ranked.

2.3.1 Trend Analysis (TrndAnal)

The trend analysis (TrndAnal) study focused on four input parameters: TVD, TT, COH, and EOVS. Each input, such as TVD, was varied between its minimum and maximum values, while other inputs, such as TT, COH, and EOVS, were kept constant in their measured values. Then, the changed parameters, such as TVD with the constant values of the other parameters, were used as inputs to predict the CTD, which is the output using the proposed ANFIS model. For TrndAnal, all independent variables, such as TVD, which changed between the minimum and maximum range, were plotted on the X-axis of a two-dimensional graph. The dependent variable (CTD) was plotted on the Y-axis of the same graph to show the relationships between each input parameter and the CTD to identify the PhyBhvr [17, 33, 34].

This study encountered difficulties obtaining the inputs, such as the TVD between the minimum and maximum, while keeping the other parameters constants, such as TT, COH, and EOVS, with the output CTD for these inputs. Due to these difficulties, the relations between the CTD and the input parameters were based on the physical relations from the literature.

3 Results and Discussion

3.1 Trend Analysis Results

Figure 2 shows the trends of TVD versus CTD for three published studies and our study. The trend was significantly unique in Kanj and Abousleiman's [8] study, such that the CTD stayed constant. At the same time, the TVD changed, perhaps because their established correlations were based only on the COH. Our proposed ANFIS model, on the other hand, follows the correct relationship between the TVD and CTD.

The shallow rocks are less consolidated than the deep ones [35]. As a result, lower TVD translates to lower CTD. When drilling the well to a target formation, the TVD of the well, like the well radius, remains constant and cannot be changed like the rate.

Figure 3 shows the TT trend. As shown in Fig. 3, increasing the TT decreases the target variable. However, Kanj and Abousleiman's [8] model demonstrates that the target variable constantly changes the TT because he did not use the TT as input to determine the CTD. Consequently, the ANFIS model trend followed the proper PhyBhvr; Fig. 3 demonstrates that the proposed model shows the correct PhyBhvr. The longer TT indicates that the sand is less consolidated [3]. Thus, raising the TT reduces the CTD.

Figure 4 presents the trend of the cohesive strength (COH) reported by four studies, all showing that the CTD increases with increasing COH. Kanj and Abousleiman's [8] model, however, has a CTD of − 2.57 MPa at 0.539 MPa of COH; thus, it is not recommended because it could predict a non-physical CTD. Figure 4 demonstrates that the COH trend of the ANFIS model has an accurate PhyBhvr. As shown in Fig. 4, raising the COH increases the CTD for the ANFIS model to demonstrate the correct PhyBhvr. The COH raises the degree of cementation [36]. A decrease in sand production can be due to the rising cementation degree of the rocks. Thus, raising the COH can raise the CTD.

Figure 5 displays the EOVS trend. Figure 5 reveals that the CTD declined by improving the EOVS. Nonetheless, Kanj and Abousleiman's [8] model demonstrates that CTD is steady with changing the EOVS, Fig. 5. Therefore, Kanj

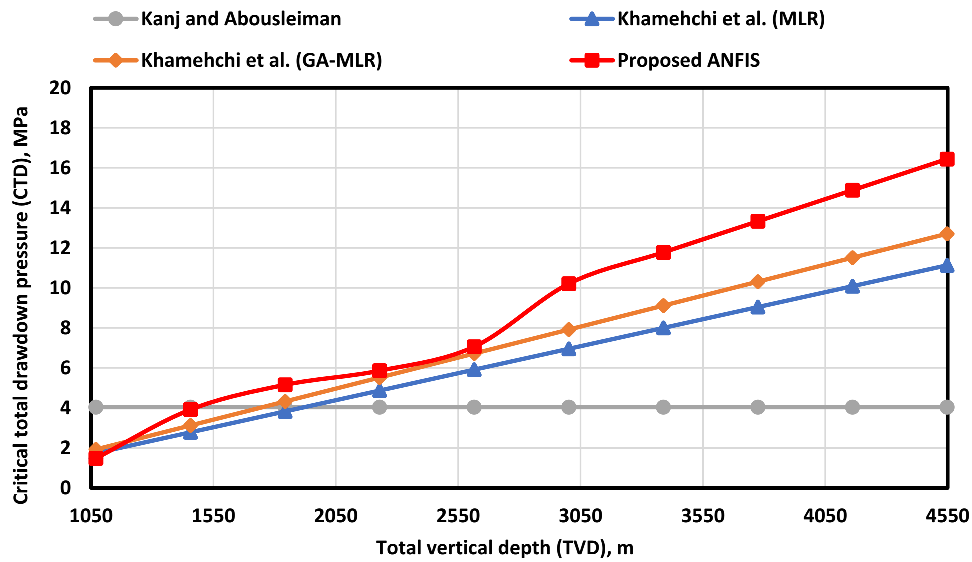

Fig. 2 Total vertical depth (TVD) TrndAnal of all models

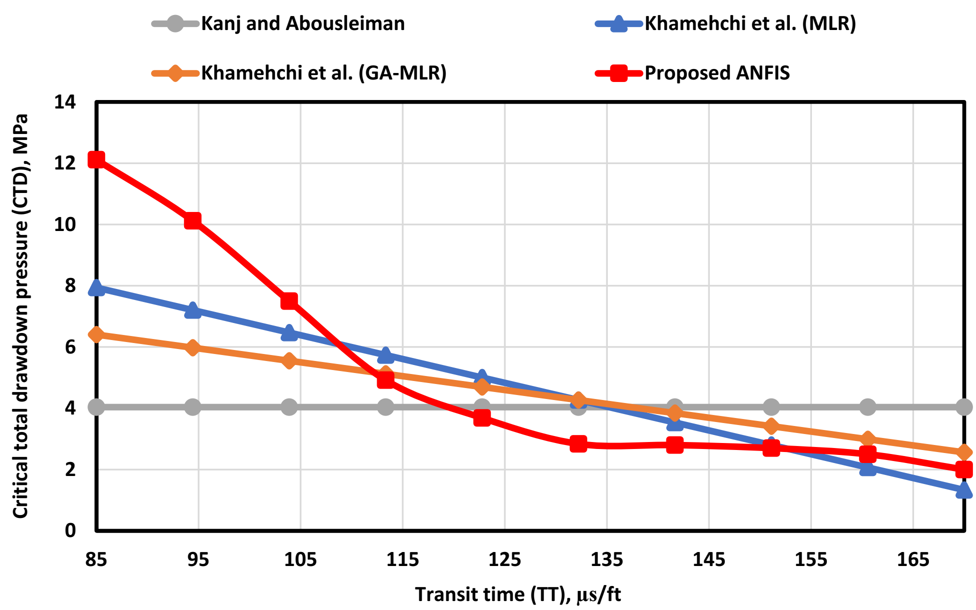

Fig. 3 TT TrndAnal of all models

and Abousleiman's [8] model indicates an incorrect relationship. As shown in Fig. 5, the ANFIS model trend displays that increasing the EOVS can increase the CTD to present the correct PhyBhvr.

In conclusion, all variables used as inputs to develop the ANFIS model have the correct trends to indicate the proper PhyBhvr. In addition, all previous models have the proper trends. However, Kanj and Abousleiman's [8] equation has incorrect TVD, TT, and EOVS trends.

3.2 The ANFIS Model with Other Model Comparisons

After the proposed ANFIS model showed the correct relationships between the inputs and output, cross-plotting analysis and statistical error analyses, such as AAPRE, APRE, RMSE, R, and SD, were performed to compare the proposed ANFIS model's accuracy with that of the previously published models.

3.2.1 Cross-plotting Analysis

The proposed model's cross-plotting to predict the CTD is shown in Figs. 6 and 7 for the training and testing datasets**.** As shown in Fig. 6**,** most data points of the measured values matched the predicted values to indicate the high accuracy for the training dataset. Most of the predicted points are close to the measured points for the testing dataset of the ANFIS model to show high accuracy in predicting the CTD with different data that are not used in the training model, Fig. 7. Figure 8 shows the cross-plotting comparison of all models.

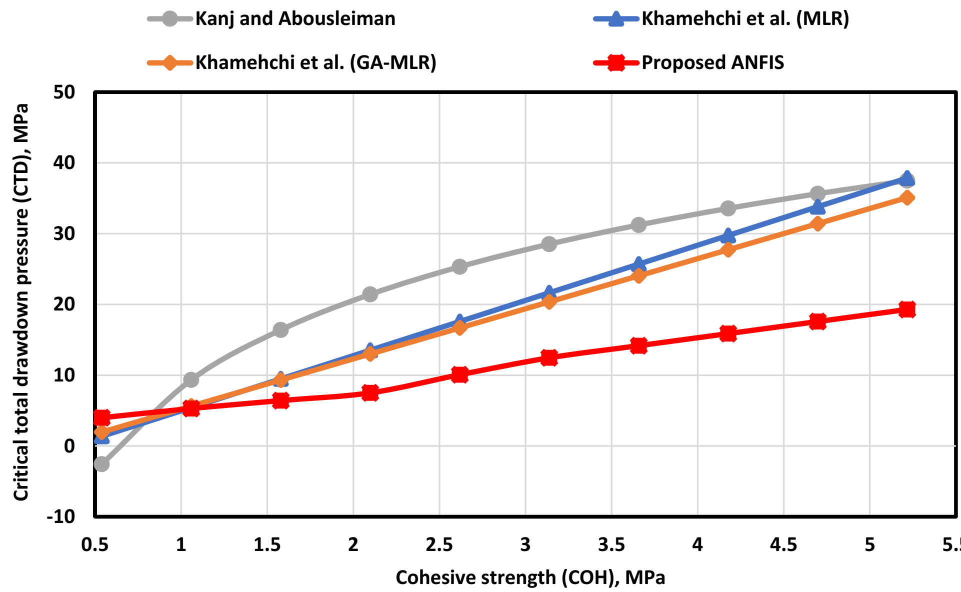

Fig. 4 COH TrndAnal of all models

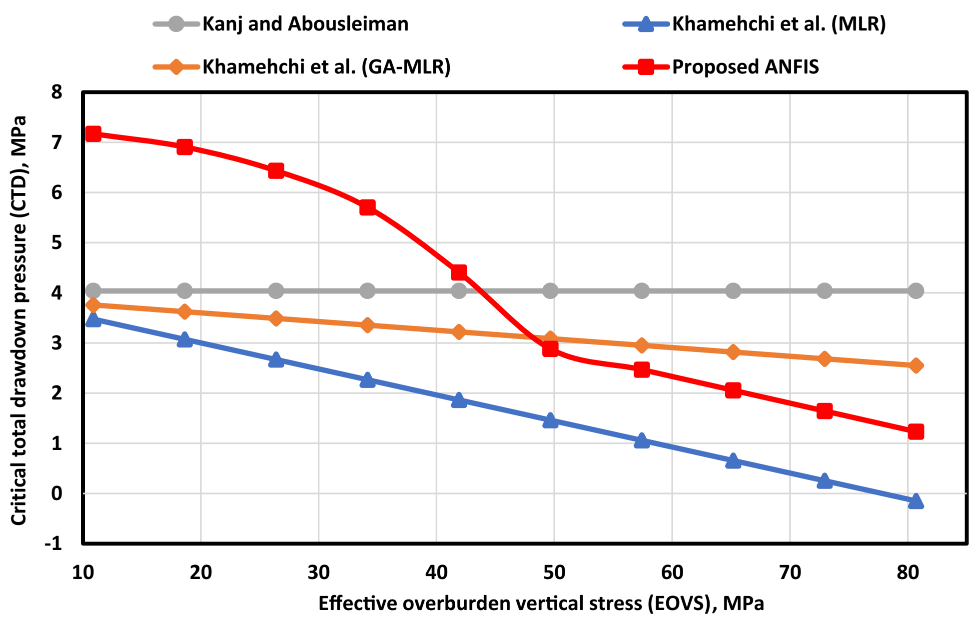

Fig. 5 EOVS TrndAnal of all models

As shown in Fig. 8, most data points of the proposed model are close to the straight line compared to other models, with some points far from the line to prove that the given model is the most accurate in predicting the CTD.

3.2.2 Statistical Error Analyses

After the TrndAnal was conducted, and all variables used as inputs to create the ANFIS model showed the proper trends, different SEA were performed to show the ANFIS model's accuracy. Furthermore, different SEA were also conducted for the previous models to compare them with the ANFIS model. The SEA equations are discussed in the Supporting Information. The ANFIS and previous models are ranked with the high R and low AAPRE to determine the best CTD model.

The *R* and AAPRE of the models were plotted as *y*-axis and *x*-axis to show their performance. As shown in Fig. 9, the ANFIS model has 0.9984 (*R*) and an AAPRE of 4.293%, which is the highest accurate model. Therefore, the ANFIS model surpasses all models in predicting CTD. Alakbari et al.'s [10] (SVM) method represents an AAPRE of 6.087% and an *R* of 0.997; Alakbari et al.'s [9] (FL) model has 0.9947 (*R*) and 8.647% (AAPRE). Alakbari et al.'s [10] (RSM) model has an AAPRE of 12.703% and 0.991 (*R*). Khamehchi et al.'s [7] (GA-MLR) method has 22.644% (AAPRE) and 0.983 (*R*). Khamehchi et al.'s [7] (MLR) has more than 30% (AAPRE).

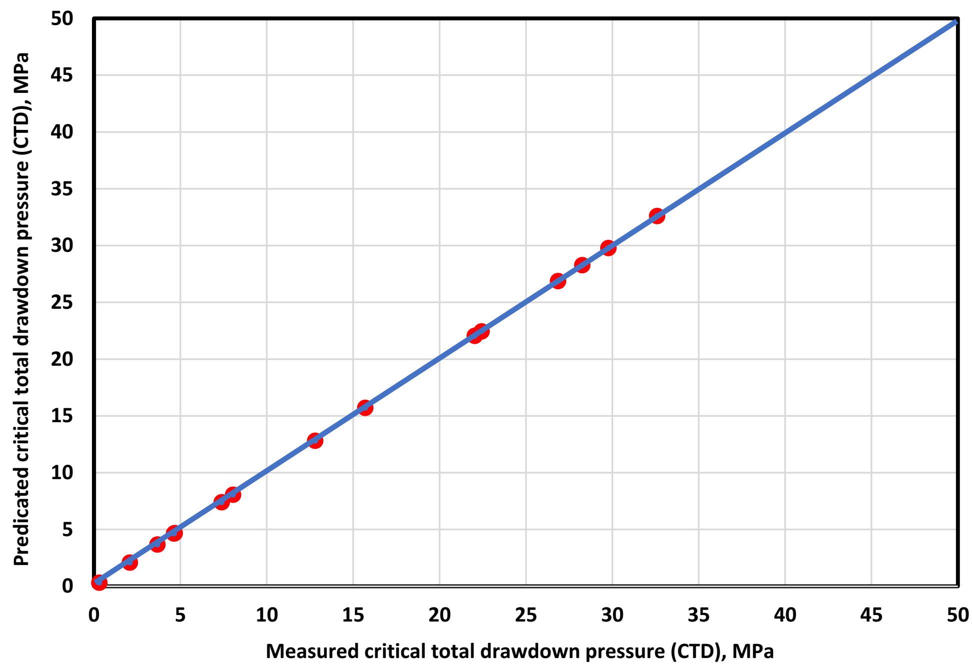

Fig. 6 Training ANFIS model's cross-plotting

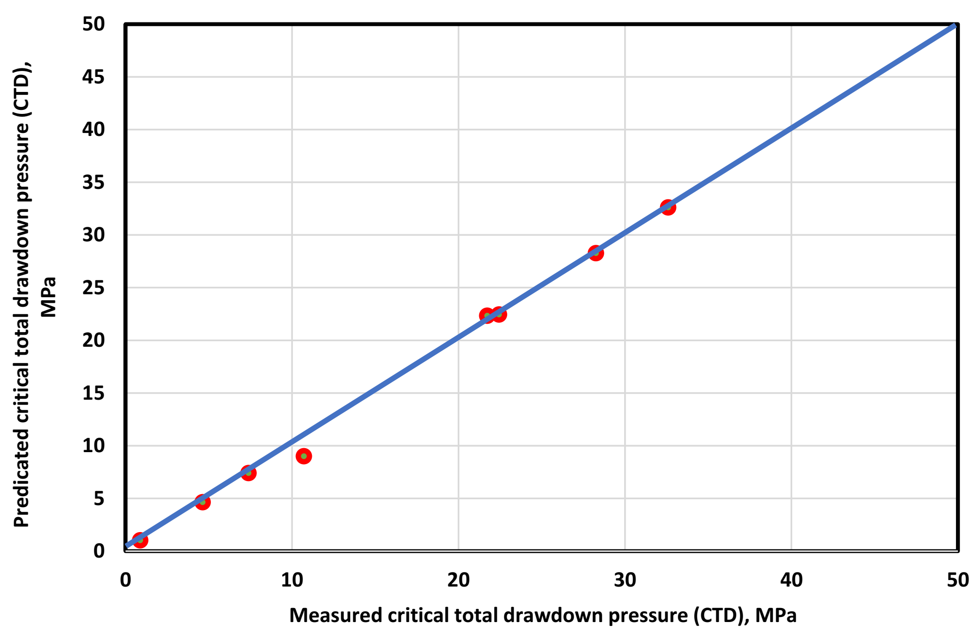

Fig. 7 Testing ANFIS model's cross-plotting

The prediction performance of all models was compared. Table 4 shows the SEA for all models. The models are ranked based on the low AAPRE and high *R*, Table 4. The ANFIS model has the lowest APRE, RMSE, and SD of − 2.52%, 0.0048 MPa, and 0.030 MPa, respectively, compared to all models. Alakbari et al.'s [10] (SVM) model has APRE, RMSE, and SD of 0.261%, 0.005 MPa, and 0.031 MPa, respectively. Alakbari et al.'s [9] (FL) model is the third-rank model with APRE, RMSE, and SD of 4.903%,0.014 MPa, and 0.082 MPa, respectively. Alakbari et al.'s [10] (RSM) model is the fourth rank. These SEA indicate that the proposed ANFIS model surpasses all current models with the highest accuracy in determining the CTD.

The previous models' performances were compared with the ANFIS approach, as shown in Figs. 10 and 11. SEA was presented in these figures to show the performances of all models with the proposed ANFIS model in a clear comparison visualization. As discussed, the ANFIS model has the lowest RMSE and SD and the highest *R* compared to other models (Fig. 10). The ANFIS model also has the lowest AAPRE (Fig. 11).

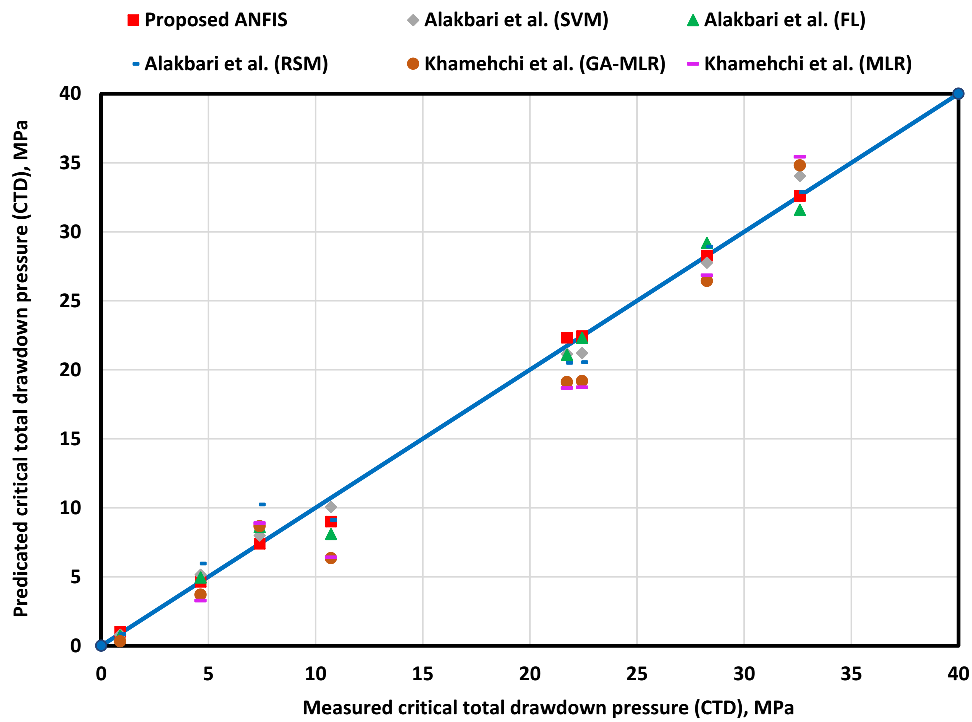

Fig. 8 All models cross-plot comparison

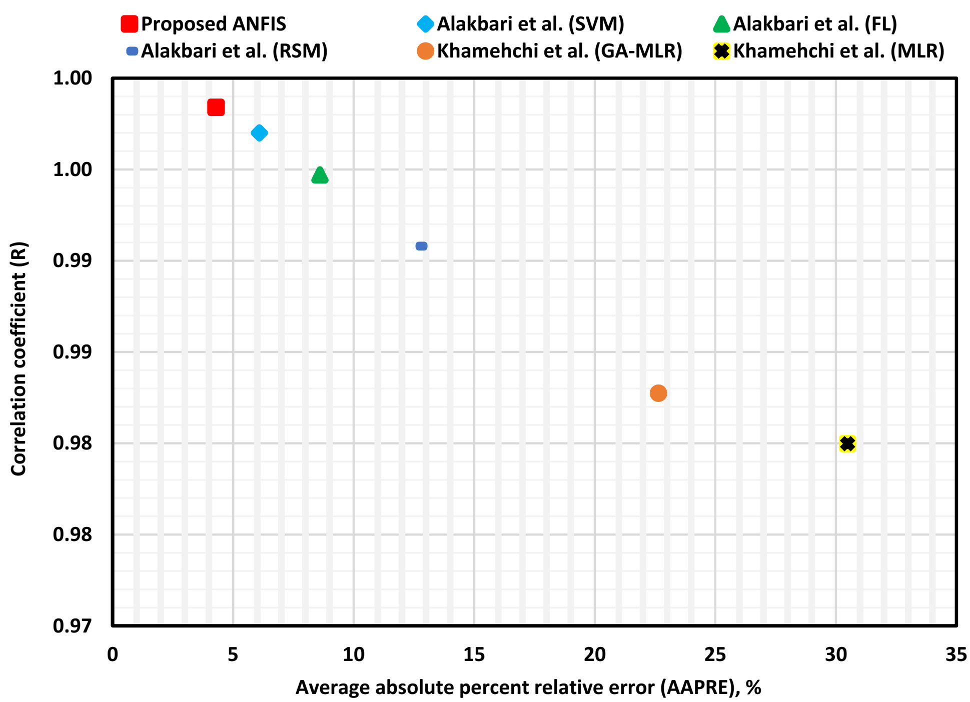

Fig. 9 *R* and AARE (%) comparison of all models

3.2.3 The Reasons Proposed Model Surpassed Other Models

Multiple linear regression (MLR) and the MLR evolved through the genetic algorithm (GA) assume linear relationships between the input and output variables. This is a significant limitation when dealing with real-world data since the relationships are often nonlinear. RSM uses low-order polynomials, such as quadratic, to model the relationships between input and output variables. RSM is sufficient for simple or moderately nonlinear systems but fails to capture more complex, higher-order nonlinear interactions. Therefore, the MLR, the GA-MLR, and the RSM show less accuracy than other methods.

Table 4 SEA comparison of all models

Model rank	Model	APRE (%)	AAPRE (%)	Emax. (%)	Emin. (%)	RMSE (MPa)	SD (MPa)	R
1	Proposed ANFIS model	− 0.252	4.3	16.0	0.00001	0.0048	0.030	0.9984
2	Alakbari et al. [10] (SVM)	0.261	6.1	10.8	1.8	0.005	0.031	0.9970
3	Alakbari et al. [9] (FL)	4.9	8.6	24.5	2.9	0.014	0.082	0.9947
4	Alakbari et al. [10] (RSM)	− 4.8	12.7	38.5	0.830	0.033	0.138	0.9910
5	Khamehchi et al. [7] (GA-MLR)	16.7	22.6	63.0	6.466	0.085	0.196	0.9830
6	Khamehchi et al. [7] (MLR)	23.2	30.5	109.9	5.015	0.194	0.340	0.9800

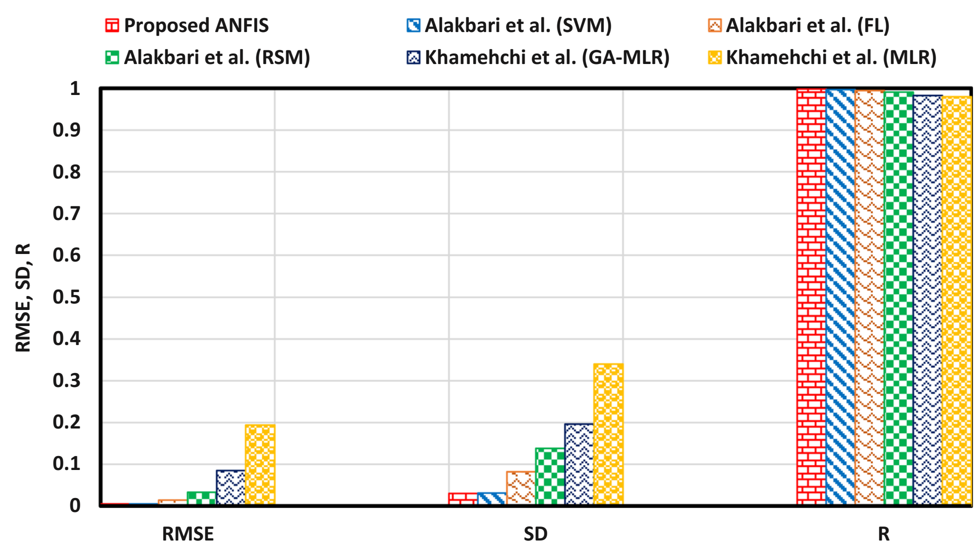

Fig. 10 RMSE, SD, and R of the models

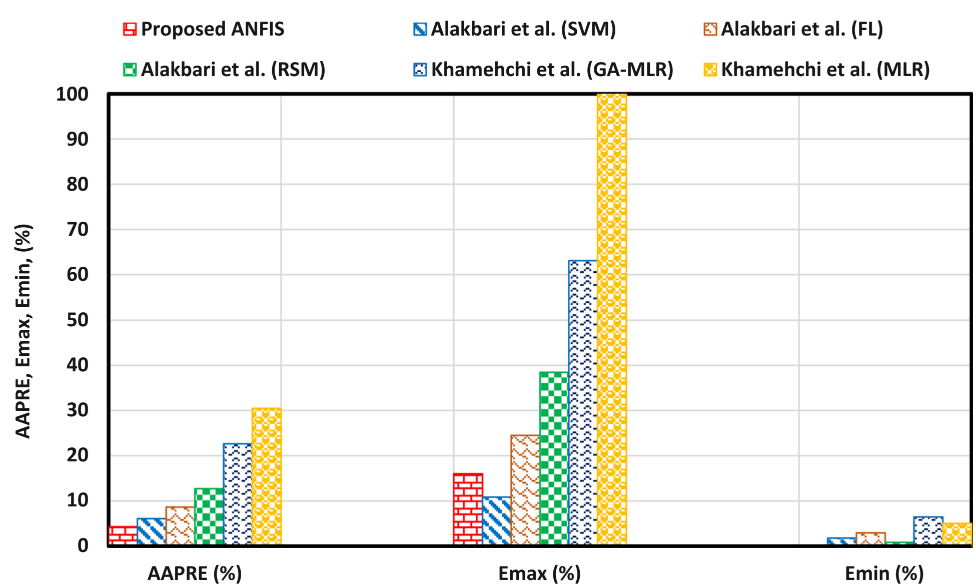

Fig. 11 AAPRE, E_{max}, and $E_{min.}$ of the models

The fuzzy logic (FL) systems without ANN support do not have the learning capabilities compared to the ANFIS, which is enhanced by the fuzzy systems with learning abilities, making ANFIS more powerful in modeling and predicting complex systems.

The SVM exceeds the RSM, GA-MLR, MLR, and FL, but the SVM was created based on a linear kernel function. The SVM requires a nonlinear kernel function to handle nonlinearity, which can complicate the model and increase computational costs. Thus, nonlinearity is avoided, and the SVM model is less accurate than the ANFIS.

The ANFIS model is described in the Methodology Sect. (2.2 Model Description), wherein the benefits of the ANFIS model are highlighted. To summarize, the ANFIS model combines the benefits of ANN and FL to speed up the decision-making process of the mapped relationship between the inputs and outputs [12]. It has been shown to have higher learning capabilities and the ability to conduct highly nonlinear mappings. It can quickly decide the mapped connection between parameters [12]. The ANFIS manages noisy data and solves classification and regression problems with the missing input data [31]. In addition, the ANFIS technique works with small data [25–27].

In addition to the benefits of ANFIS methods, this study benefitted from using one of the largest datasets available in the literature. The results show that the proposed model is more accurate in predicting the CTD than the best MLR, GA-MLR, RSM, FL, and SVM methods published in the literature.

4 Conclusions

An ANFIS model was developed to accurately predict the critical total drawdown pressure (CTD). The developed ANFIS model was assessed using several methods, such as SEA. The proposed ANFIS model was compared with the published models to determine the CTD. This study highlights the following:

- The TrndAnal proves that the proposed model has the correct relationship between all features and the target variable: CTD. In addition, the TrndAnal for the previous models was discussed.
- The SEA shows that the ANFIS model can accurately determine the CTD with a correlation coefficient (R) of 0.9984, standard deviation (SD) of 0.030 MPa, APRE of -0.252%, AAPRE of 4.293%, and RMSE of 0.0048 MPa.
- The ANFIS and previously published models were ranked based on the high R and low AAPRE. The first rank model is the ANFIS model, with 4.293% (AAPRE) and 0.9984 (R). Alakbari et al.'s [10] (SVM) mode comes after the ANFIS model with AAPRE and R of 6.087% and 0.9970, respectively. Alakbari et al.'s [9] (FL) model is the third rank with AAPRE and R of 8.647% and 0.9947, respectively. Alakbari et al.'s [10] (RSM) equation is the fourth rank and has AAPRE and R of 12.703% and 0.9910. The SEA shows that the ANFIS model surpasses all models with the highest accuracy to determine the CTD.
- The model reliability was improved by randomizing the data to ensure that each dataset did not memorize the pattern and to prevent the model from over-fitting.
- The cross-plotting demonstrates that most data points of predicted values for the ANFIS model matched the measured values to show the high accuracy of the proposed ANFIS model. Furthermore, cross-plotting for all previously published models was conducted to compare them with the ANFIS model.
- This research provides a cost-effective tool for CTD prediction in accurate and fast determinations to know sand production at the early stages and prevent it, which costs millions of dollars in the petroleum industry.

5 Benefits, Limitations, and Future Works

The proposed ANFIS model does have some limitations. It was created based on the total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS) in the ranges (1070–4548) m, (85–170) μs/ft, (0.539–5.22) MPa, and (10.88–80.71) MPa, respectively.

The proposed ANFIS model is based on the highest amount of data with a wide range of inputs compared to previous studies in the literature. Moreover, the proposed ANFIS model to predict CTD was assessed using different methods to show its robust CTD prediction. The relationships between inputs and output were determined by trend analysis to prove the physical behavior. The proposed ANFIS model is the best model to accurately predict the CTD compared to all models in the literature. Thus, the advantages of the proposed ANFIS model in predicting the CTD significantly surpass its limitations.

As discussed in the introduction, many factors may influence sand production, such as tectonic activities, overburden pressure, pore pressure, stress induced during drilling, fluid drag force, fluid flow, and rock strength [2]. The data-driven models can solve many complex field problems. This study can be extended to consider other factors that affect sand production if data for these parameters can be collected. Future studies can also explore other machine learning methods to predict the CTD.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13369-024-09556-8.

Acknowledgements The authors sincerely thank the Yayasan Universiti Teknologi PETRONAS (YUTP) (Cost Centre: 015LC0-451) at Universiti Teknologi PETRONAS for supporting this research. Open Access funding provided by the Qatar National Library.

Funding Open Access funding provided by the Qatar National Library. This work was supported by Universiti Teknologi Petronas, Grant no.: 015LC0-451, Syed Mohammad Mahmood.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Subbiah, S.K.; Samsuri, A.; Mohamad-Hussein, A.; Jaafar, M.Z.; Chen, Y.R.; Kumar, R.R.: Root cause of sand production and methodologies for prediction. Petroleum **7**, 263–271 (2021)
2. Vincent, O.N.; Abiola, S.O.; Felix, O.O.; Ajienka, J.A.: Sanding in oil well reservoir completions. In: SPE Nigeria Annual International Conference and Exhibition, pp. 1–10. Society of petroleum engineers (2012)
3. Mahmud, H.B.; Leong, V.H.; Lestariono, Y.: Sand production: a smart control framework for risk mitigation. Petroleum **6**, 1–13 (2020)
4. Cook, J.M.; Bradford, I.D.R.; Plumb, R.A.: A study of the physical mechanisms of sanding and application to sand production prediction. In: The European Petroleum Conference, pp. 473–480. Society of petroleum engineers (1994)
5. Veeken, C.A.M.; Davies, D.R.; Kenter, C.J.; Kooijman, A.P.: Sand production prediction review: developing an integrated approach. In: SPE annual technical conference and exhibition, pp. 335–346. Society of petroleum engineers (1991)
6. Alakbari, F.S.; Mohyaldinn, M.E.; Muhsan, A.S.; Hasan, N.; Ganat, T.: Chemical sand consolidation: from polymers to nanoparticles. Polymers (Basel) **12**, 1069 (2020). https://doi.org/10.3390/polym12051069
7. Khamehchi, E.; Kivi, I.R.; Akbari, M.: A novel approach to sand production prediction using artificial intelligence. J. Pet. Sci. Eng. **123**, 147–154 (2014)
8. Kanj, M.Y.; Abousleiman, Y.: Realistic sanding predictions: a neural approach. In: SPE Annual Technical Conference and Exhibition, pp. 1–10. Society of petroleum engineers (1999)
9. Alakbari, F.S.; Mohyaldinn, M.E.; Ayoub, M.A.; Muhsan, A.S.; Hussein, I.A.: A robust fuzzy logic-based model for predicting the critical total drawdown in sand production in oil and gas wells. PLoS ONE **16**, e0250466 (2021)
10. Alakbari, F.S.; Mohyaldinn, M.E.; Ayoub, M.A.; Muhsan, A.S.; Abdulkadir, S.J.; Hussein, I.A.; Salih, A.A.: Prediction of critical total drawdown in sand production from gas wells: machine learning approach. Can. J. Chem. Eng. **101**(5), 2493–2509 (2023)
11. Isanta Navarro, R.: Study of a neural network-based system for stability augmentation of an airplane. MSc thesis, Universitat Politecnica de Catalunya, Spain (2013)
12. Tahmasebi, P.; Hezarkhani, A.: A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. Comput. Geosci. **42**, 18–27 (2012)
13. Maguire, L.P.; Roche, B.; McGinnity, T.M.; McDaid, L.J.: Predicting a chaotic time series using a fuzzy neural network. Inf. Sci. (N Y). **112**, 125–136 (1998)
14. Ayoub, M.; Shien, G.; Diab, D.; Ahmed, Q.: Modeling of drilling rate of penetration using adaptive neuro-fuzzy inference system. Int. J. Appl. Eng. Res. **12**, 12880–12891 (2017)
15. Sambo, C.H.; Hermana, M.; Babasari, A.; Janjuhah, H.T.; Ghosh, D.P.: Application of artificial intelligence methods for predicting water saturation from new seismic attributes. In: Offshore Technology Conference, pp. 1–8. Society of Petroleum Engineers (2018)
16. Hamdi, Z.; Chenxi, D.: Accurate prediction of CO_2 minimum miscibility pressure using adaptive neuro-fuzzy inference systems. In: SPE Gas & Oil Technology Showcase and Conference, pp. 1–17. Society of Petroleum Engineers (2019)
17. Ayoub, M.A.; Zainal, S.N.; Elhaj, M.E.; Ku Ishak, K.E.H.; Ahmed, Q.: Revisiting the coefficient of isothermal oil compressibility below bubble point pressure and formulation of a new model using adaptive neuro-fuzzy inference system technique. In: International Petroleum Technology Conference, pp. 1–14. Society of Petroleum Engineers (2020)
18. Kalam, S.; Khan, M.; Khan, R.A.; Alam, M.M.; Sadeed, A.; Mahmoud, M.; Abu-Khamsin, S.A.: Advancing relative permeability estimation through data-driven modeling. In: Abu Dhabi International Petroleum Exhibition & Conference, pp. 1–21. Society of Petroleum Engineers (2020)
19. Ayoub, M.A.; Mohyaldinn, M.E.; Manalo, A.; Hassan, A.; Ahmed, Q.A.: A new model for predicting minimum miscibility pressure (MMP) in reservoir-oil/injection gas mixtures using Adaptive Neuro Fuzzy Inference System. In: Advances in Material Sciences and Engineering, pp. 527–545. Springer (2020)
20. Alakbari, F.S.; Mohyaldinn, M.E.; Ayoub, M.A.; Muhsan, A.S.; Hussein, I.A.: A reservoir bubble point pressure prediction model using the Adaptive Neuro-Fuzzy Inference System (ANFIS) technique with trend analysis. PLoS ONE **17**, e0272790 (2022)
21. Ayoub Mohammed, M.A.; Alakbari, F.S.; Nathan, C.P.; Mohyaldinn, M.E.: Determination of the gas-oil ratio below the bubble point pressure using the Adaptive Neuro-Fuzzy Inference System (ANFIS). ACS Omega **7**(23), 19735–19742 (2022)
22. Alakbari, F.S.; Mohyaldinn, M.E.; Ayoub, M.A.; Muhsan, A.S.; Hussein, I.A.: Development of oil formation volume factor model using Adaptive Neuro-Fuzzy Inference Systems ANFIS. In: SPE/IATMI Asia Pacific oil & gas conference and exhibition, pp. 1–23. Society of Petroleum Engineers (2021)
23. Aïfa, T.; Baouche, R.; Baddari, K.: Neuro-fuzzy system to predict permeability and porosity from well log data: a case study of Hassi R′ Mel gas field, Algeria. J. Pet. Sci. Eng. **123**, 217–229 (2014)
24. Moricca, G.; Ripa, G.; Sanfilippo, F.; Santarelli, F.J.: Basin scale rock mechanics: field observations of sand production. In: Rock Mechanics in Petroleum Engineering, pp. 317–328. Society of Petroleum Engineers (1994)
25. Tabaraki, R.; Khodabakhshi, M.: Performance comparison of wavelet neural network and adaptive neuro-fuzzy inference system with small data sets. J. Mol. Graph. Model. **100**, 107698 (2020)

26. Al-Ali, A.; Elharrouss, O.; Qidwai, U.; Al-Maaddeed, S.: ANFIS-Net for automatic detection of COVID-19. Sci. Rep. **11**, 17318 (2021)
27. Yang, J.; Shang, C.; Li, Y.; Li, F.; Shen, L.; Shen, Q.: Constructing anfis with sparse data through group-based rule interpolation: an evolutionary approach. IEEE Trans. Fuzzy Syst. **30**, 893–907 (2021)
28. Jafarzade, N.; Kisi, O.; Yousefi, M.; Baziar, M.; Oskoei, V.; Marufi, N.; Mohammadi, A.A.: Viability of two adaptive fuzzy systems based on fuzzy c means and subtractive clustering methods for modeling Cadmium in groundwater resources. Heliyon **9**(8), 1–13 (2023)
29. Jang, J.-S.: ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. **23**, 665–685 (1993)
30. Jang, J.-S.; Sun, C.-T.: Neuro-fuzzy modeling and control. Proc. IEEE **83**, 378–406 (1995)
31. Bamisaye, A.; Ige, A.R.; Adegoke, K.A.; Adegoke, I.A.; Bamidele, M.O.; Alli, Y.A.; Adeleke, O.; Idowu, M.A.: Amaranthus hybridus waste solid biofuel: comparative and machine learning studies. RSC Adv. **14**, 11541–11556 (2024)
32. Bergstra, J.; Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13 (2012)
33. Al-Shammasi, A.A.: Bubble point pressure and oil formation volume factor correlations. In: SPE Middle East Oil and Gas Show and Conference, pp. 1–11. Society of Petroleum Engineers (1999)
34. Osman, E.A.; Ayoub, M.A.; Aggour, M.A.: Artificial neural network model for predicting bottomhole flowing pressure in vertical multiphase flow. In: SPE middle east oil and gas show and conference, pp. 1–11. Society of Petroleum Engineers (2005)
35. Ahad, N.A.; Jami, M.; Tyson, S.: A review of experimental studies on sand screen selection for unconsolidated sandstone reservoirs. J. Pet. Explor. Prod. Technol. **10**, 1–14 (2020)
36. Aadnoy, B.; Looyeh, R.: Petroleum rock mechanics: drilling operations and well design, 2nd edn. Gulf Professional Publishing (2019)