ORIGINAL ARTICLE

Unconfined compressive strength (UCS) prediction in real-time while drilling using artificial intelligence tools

Ahmed Gowida[1] · Salaheldin Elkatatny[1,2] · Hany Gamal[1]

Abstract

Unconfined compressive strength (UCS) is a major mechanical parameter of the rock which has an essential role in developing geomechanical models. It can be estimated directly by lab testing of retrieved core samples or from well log data. These methods are very expensive and require huge efforts and time. Therefore, there is a need to develop a new technique for predicting UCS values in real-time. In this study, three artificial intelligence (AI) models were developed using artificial intelligence tools; artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM) to predict UCS of the downhole formations while drilling based on real-time recording of the drilling mechanical parameters. These parameters include rate of penetration (ROP), mud pumping rate (GPM), stand-pipe pressure (SPP), rotary speed in revolution per minute (RPM), torque (T), and weight on bit (WOB). A dataset of 1771 points from a Middle Eastern field was used to build the developed models: for training and testing processes. A new UCS correlation was developed based on the optimized AI model. Another set of data (2175 data points unseen by the model) was used to validate the model and the developed UCS correlation. The developed ANN-model outperformed the ANFIS- and SVM-models with a correlation coefficient (R-value) of 0.99 and an average absolute percentage error (AAPE) of 3.48% between the predicted and actual UCS values. The new UCS correlation outperformed the available correlations for UCS prediction and it was able to predict the UCS with AAPE of 4.2% compared to the actual UCS values.

Abbreviations

AAPE	Average absolute percentage error
AI	Artificial Intelligence
ANN	Artificial neural network
ANFIS	Adaptive network-based fuzzy interference system
SVM	Support vector machine
R	Correlation coefficient
R^2	Coefficient of determination
ROP	Rate of penetration
WOB	Weight on bit
RPM	Rotating speed in revolution per minute
GPM	Gallon per minute
SPP	Standpipe pressure
T	Torque
Fitnet Function	Fitting neural network
Newdtdnn	Create distributed time delay neural network
newnarx	Create feedforward backpropagation network with feedback from output to input
newelm	Create Elman backpropagation network
newfftd	Create feedforward input-delay backpropagation network
newff	Create feedforward backpropagation network
newlrn	Layer-Recurrent Network
tansig	Hyperbolic tangent sigmoid transfer function
logsig	Log-sigmoid transfer function
hardlims	Hard-limit transfer function
trainbr	Bayesian regularization

✉ Salaheldin Elkatatny
elkatatny@kfupm.edu.sa

Ahmed Gowida
g201708730@kfupm.edu.sa

Hany Gamal
g201706870@kfupm.edu.sa

1 College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals, Box: 5049, Dhahran 31261, Saudi Arabia

2 Petroleum Department, Cairo University, Giza 12613, Egypt

purelin	Linear transfer function
softmax	Softmax transfer function
tribas	Triangular basis transfer function
trainlm	Levenberg–Marquardt backpropagation
trainbfg	BFGS quasi-Newton backpropagation
traingdx	Gradient descent with momentum and adaptive learning rule backpropagation
trainoss	One step secant backpropagation

Table 1 Published correlations to predict UCS in MPa

Equation number	Correlation	References
(1)	$UCS = \left(\frac{7682}{\Delta t}\right)^{1.82}$	[6]
(2)	$UCS = 10\left(2.44 + \frac{109.14}{\Delta t}\right)$	[7]
(3)	$UCS = 143.8 \exp^{(6.95\phi)}$	[8]
(4)	$UCS = \left(\frac{80204}{\Delta t}\right)^{1.285}$	[9]
(5)	$UCS = 7600 e^{-0.064\Delta t_c}$	[10]
(6)	$UCS = 292.047 e^{-9.541\phi}$	[11]
(7)	$UCS = 570.808 e^{-0.031\Delta t_c}$	

1 Introduction

Unconfined compressive strength (UCS) is an essential mechanical rock-parameter for building geomechanical models. It defines the maximum compressive stress limit that the rock can endure without failure when uniaxial loading is applied under confinement conditions [1]. Rock mechanics is generally theoretical and applied science that describes the mechanical behavior of the rock when different stresses are applied. Rock failure is one of the petroleum-related features of rock mechanics which is the reason behind many downhole problems such as solids production and wellbore instability issues [2]. The availability of UCS data of the downhole formations has a high degree of importance in optimizing the drilling operation performance, bit hydraulics, determining the proper mud weight while drilling, and avoiding wellbore instability issues [3].

Rock mechanical properties can be determined directly using experimental tests conducted on retrieved core samples representing the actual stress-state conditions and mechanical properties of the rocks downhole. UCS can be estimated in the laboratory using different tests such as uniaxial compressive strength test, triaxial compressive strength test, scratch test, Schmidt hammer test, and point load test. These laboratory measurements are considered the most accurate and reliable method for determining such properties [4]. However, since retrieving core samples representing the sections of interest is very costly and requires a lot of time, core-testing data cannot yield a continuous profile of UCS along the drilled wellbore [5]. Besides, representative core samples are not always available and their number is often restricted. Therefore, indirect methods were developed to fill the missing gaps using derived correlations between the rock mechanical properties and petrophysical well-log data.

Many studies in the literature have reported different correlations (Table 1) to predict UCS of different lithologies using logging data based on nonlinear regression techniques [6–11]. Most of these models are based on sonic transit time (Δt) and formation porosity ($\emptyset$) for estimating UCS. Moreover, the application of artificial intelligence

tools has been introduced to correlate core-testing UCS data with different well-log data such as bulk density and sonic log data [12–14]. However, such well-log data are not always available while drilling the wellbore, as they are often obtained using a wireline logging technique which is usually performed after drilling the wellbore to avoid the harsh drilling environment [15]. Besides, the acquisition of logging data while drilling is a costly process and has several operational limitations and requires many corrections as well. This created the main motivation behind this study which is developing a new technique through which UCS values can be estimated in real-time other than using the well-log data. Additionally, the substantial value of obtaining the UCS data during drilling addresses the increasing need for such data in real-time. This is because it is very crucial to optimize the drilling operation, reduce the non-productive time (NPT), prevent the collapse of downhole formations, and, hence, avoid many wellbore instability problems by choosing the proper mud weight.

Therefore, the main objective of this paper is to introduce a novel technique for predicting the UCS values of the downhole formations while drilling. This involved developing new models using different AI tools to predict UCS depending on the mechanical drilling parameters instead of well log data. These mechanical drilling parameters are usually available while drilling; include rate of penetration (ROP), mud pumping rate (GPM), stand-pipe pressure (SPP), drillstring rotating-speed (RPM), drilling torque (T), and weight on bit (WOB). Unlike the logging data, such drilling data are always available during the drilling operation at a low cost. Accordingly, the developed AI-based models would help estimate UCS of the downhole formations while drilling in a time- and cost-effective way.

2 Materials and methods

In this study, new AI models were developed to estimate UCS values of the downhole formations while drilling using three AI tools: the artificial neural network (ANN),

adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM). The developed models used the mechanical drilling parameters as feeding inputs to predict the output; formations' UCS-values.

2.1 Data description

A dataset of 1771 data points was collected representing a vertical section (Well A) within a field in the Eastern province of Saudi Arabia (Middle East Region). The obtained dataset included mechanical drilling data and the corresponding UCS values of the drilled formations. These drilling parameters (ROP, GPM, SPP, RPM, T, and WOB) are normally recorded at the surface while drilling using accurate real-time sensors to monitor the drilling performance. The corresponding UCS data were extracted from the geomechanical model of the area under study: The aforementioned model was developed based on actual core samples lab-testing data.

2.2 Dataset statistical analysis

The dataset used was statistically analyzed to describe its ranges and distributions before the model development. Table 2 lists different statistical parameters for the inputs (ROP, GPM, SPP, RPM, T, and WOB) and the output parameter (UCS). The ranges of the input and output parameters are ROP from 35.18 to 108.34 ft/hr, GPM from 645.30 to 854.01 gallon-per-minute, SPP from 2140 to 3076 psi, RPM from 77.94 to 162.53 revolution-per-minute, T from 4.29 to 10.68 kft.Ib, WOB from 1.54 to 25.48 kIb, and UCS from 1030.02 to 17378.06 psi. The dataset used was found to cover wide ranges with satisfactory distribution as inferred from the statistical analysis. This would boost the prediction performance of the proposed model to describe the phenomena more efficiently.

2.3 Data processing and filtration

The prediction accuracy of the AI-based models is significantly affected by the quality of the data used while developing the model. Therefore, the data were pre-processed and filtered using statistical analysis and engineering sense based on the literature. A specially designed MATLAB program was used to remove unreasonable values like zeros and negative values in addition to any missing points. Then, outliers were removed using box and whisker plot, in which the top whisker represents the upper limit of the data and the bottom whisker represents the lower limit of the data [16]. Any value beyond these limits was considered an outlier then removed. These limits were determined using the statistical parameters listed in Table 2.

2.4 Study of input(s)/output relationship strength

To identify the strength of the relationship between the output (UCS) and the input parameters, the Pearson correlation coefficient (R) with a mathematical formula expressed in Eq. 8 was used to identify how strongly two parameters are linearly related to each other [17]. Its value ranges from -1 to $+1$. A strong direct linear-relation is indicated with an R-value of $+1$. On the contrary, the R-value of -1 shows a strong inverse linear-relation between these two variables. While the R-value of zero indicates no linear relationship exists between the two study variables. Figure 1 shows the relative importance between the output parameter (UCS) and the input parameters individually in terms of R-value. UCS had R-values of 0.15, 0.14, 0.12, 0.07, 0.02 and -0.18 with ROP, SPP, T, WOB, GPM, and RPM, respectively.

$$R = \frac{k \sum xy - (\sum x)(\sum y)}{\sqrt{k(\sum x^2) - (\sum y)^2}\sqrt{k(\sum b^2) - (\sum b)^2}} \tag{8}$$

where; R represents the correlation coefficient between the two variables x and y.

Table 2 Statistical analysis describing the ranges and distribution of the used data

Statistical parameter	WOB (klb)	T (kft.Ib)	SPP, psi	RPM	ROP (ft/hr)	GPM	UCS (psi)
Minimum	1.54	4.29	2140.20	77.94	27.41	645.30	1030.02
Maximum	25.48	10.68	3076.00	162.53	119.57	854.01	17378.06
Range	23.94	6.39	935.80	84.60	92.16	208.71	16348.04
Mean	12.13	7.48	2604.42	128.67	66.03	729.26	6193.38
Median	11.45	7.06	2611.89	134.10	70.10	700.11	5240.81
Mode	12.13	6.86	2802.09	159.25	108.35	843.09	10985
Standard deviation	7.22	1.66	210.77	14.11	19.16	72.38	3782.81
Skewness	0.44	0.31	$-$ 0.32	$-$ 1.06	$-$ 0.05	$-$ 0.03	0.45

Fig. 1 Relationship strength between the output (UCS) and the input parameters in terms of correlation coefficient (R)
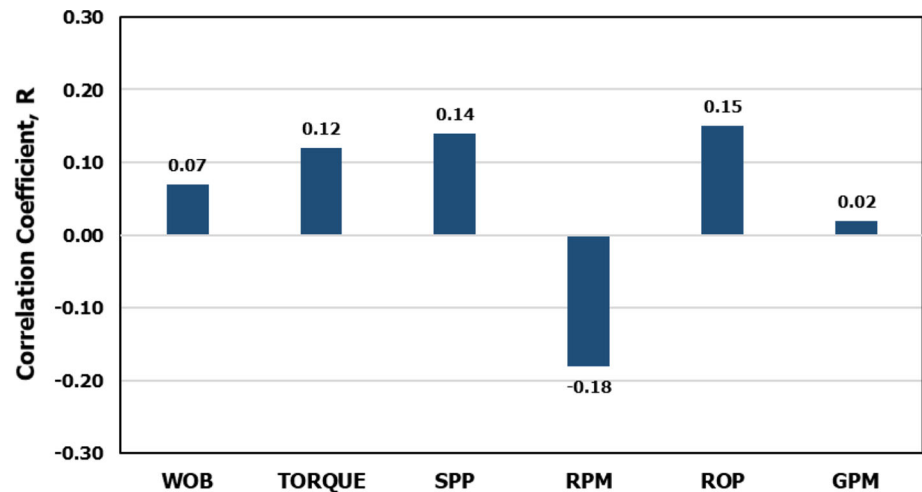
2.5 Application of artificial intelligence tools

2.5.1 Artificial neural network (ANN)

ANN is an artificial intelligence tool which has recently got a remarkable reputation for modeling complex engineering phenomenon and complicated problems in a significant efficient way. Recently, ANN has been used in many applications in petroleum engineering fields. Many studies in the literature introduced the application of ANN in the area of formation evaluation and rock characterization [18–24]. A typical neural network consists of three types of layers; input layer, hidden layer, and output layer [25]. These layers are connected by a set of weights and biases which are tuned during the optimization process of the network to control the prediction performance of the network [26]. The connection between the network components mimics that of the biological neural system in handling and processing the data [27]. The network is usually trained with different learning algorithms to optimize the network and to control the processing of the neurons [28]. These neurons are considered the elementary elements from which any neural network is constructed [29].

2.5.1.1 Neural network application to predict UCS
In this study, a new model was developed using the obtained dataset to predict UCS while drilling using ANN. A specially designed MATLAB program was used to feed the network with the input parameters to estimate UCS. Figure 2 shows a schematic diagram for the developed network architecture. The model development passes through three main stages; these are the training process, the testing process, and the validation process. First, the obtained dataset was randomly stratified into two sets; training and testing set for building the model. Splitting the data was in
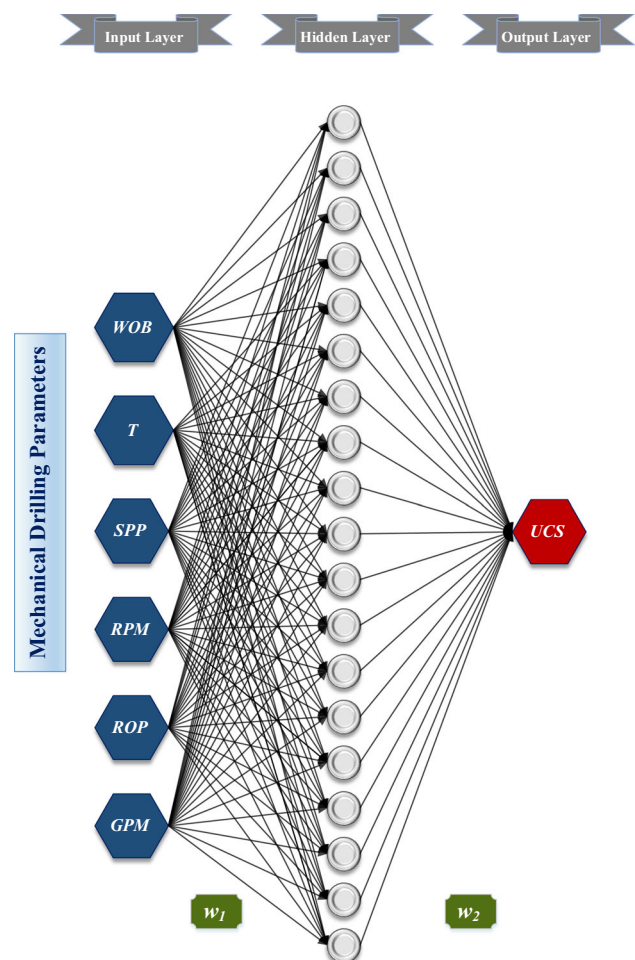

Fig. 2 The typical schematic architecture of the developed network

the way that the testing set was within the same range as the training set. During the training process, the network was trained with the data to tune its parameters to achieve the highest possible accuracy. Then, the results were evaluated and compared to the actual output (UCS) values

based on two main evaluating factors; these were correlation coefficient (R), and average absolute percentage error (AAPE). The formulas used for estimating these factors (R and AAPE) are stated in Eqs. 8 and 9. Thereafter, the developed network was tested and cross-validated using hidden testing set to check the performance of the developed network while using unseen data.

$$AAPE = \frac{\sum \left| \frac{UCS_{measured} - UCS_{predicted}}{UCS_{measured}} \right| \times 100}{n} \quad (9)$$

where; n is the number of the data points used.

2.5.1.2 ANN-model optimization
The accuracy of the prediction process depends mainly on the tuning process of the network parameters. These parameters include:

- Data splitting ratio
- Number of hidden layers
- Number of neurons in each hidden layer
- Network function
- Type of transfer functions
- Type of training algorithm
- Learning rate

The developed network was trained using the stochastic gradient descent optimization algorithm which is considered the most popular optimization algorithm in AI applications. It depends on updating the network parameters iteratively in the gradient direction of the objective function. Through every update, the algorithm helps the model gradually converge to reach the optimal value of the objective function. It involves taking random values of the model hyperparameters and adjusting them iteratively based on the available options to reduce the loss function eventually. Through each iteration, the model is guided to update the model hyperparameters to reduce the error of the next iteration using the backpropagation technique.

The developed MATLAB code was designed to generate different combinations of the aforementioned parameters with their available options listed in Table 3. The network was run with several trials and each trial included using one of the generated combinations of ANN parameters. During each trial, the results were tested, and the prediction performance was evaluated based on the comparison between the actual UCS values and the predicted ones, indicated from the evaluating factors (R, and AAPE). The target of the optimization process was to identify the ANN parameters' combination which would result in the highest prediction accuracy. This would be inferred from the highest R-value between the predicated and actual UCS values accompanied by the lowest AAPE between them. The optimized network parameters were:

Table 3 Summary of the tested ranges of the network parameters during the optimization process

| Network parameter	Tested options			
Splitting ratio	70/30	80/20		90/10
No. of hidden layer(s)	1–3			
No. of neurons	5–30			
Network function	fitnet	newdtdnn		newnarx
	newelm	newfftd	newff	newlrn
Transfer function (s)	tansig		purelin	
	logsig		softmax	
	hardlims		tribas	
Training algorithm (s)	trainbr		trainlm	
	trainbfg	traingdx		trainoss
Learning rate	0.01–0.9			

- Data splitting ratio: 70/30 ratio for training and testing respectively
- Number of hidden layers: single hidden layer
- Number of neurons in each hidden layer: 30 neurons
- Network function: fitnet
- Type of transfer functions: tan-sigmoidal function (tansig)
- Type of training algorithm: Bayesian Regularization backpropagation (trainbr)
- Learning rate: 0.12

The proposed model was developed by training the network with 70% of the dataset (1240 points) while 30% of the data (531 points) was used for testing the performance of the model. The results of the prediction process showed that the predicted UCS values significantly matched with the actual values. This obvious match is indicated from the high R-value which reached 0.99 between the predicted and actual UCS values for both the training and testing processes as shown in the cross-plots (Fig. 3a and b). These cross-plots inferred that the predicted UCS is so close to the actual ones as the plotted data considerably coincide with the forty-five-line with a correlation novelty coefficient exceeded 0.99 between the actual and predicted UCS values. This confirms the high accuracy of the prediction of The UCS values using the developed ANN-model.

2.5.2 Adaptive neuro-fuzzy inference system (ANFIS)

Fuzzy Logic (FL) systems are structured based on fuzzy techniques that resemble human reasoning during information processing. Its technique emulates the way of human decision making involving all intermediate possibilities. Adaptive fuzzy systems depend on the automatic

Fig. 3 Crossplots between the predicted and actual UCS values for **a** training process, and **b** testing process, using the developed ANN-model. This shows how the predicted values were so close to the actual ones

(a)
(b)
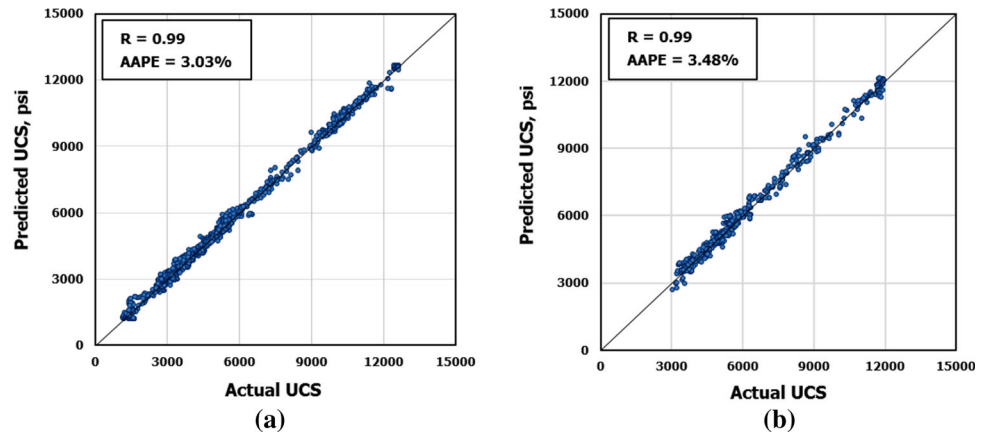
online synthesis and tuning of fuzzy controller parameters using online data to continually "learn" the fuzzy controller. Recently, different adaptive fuzzy systems are used to simulate complex, and nonlinear systems; adaptive neuro-fuzzy inference system (ANFIS), Event-triggered adaptive fuzzy systems, etc. [30–32]. From the functionality perspective, there are almost no constraints on the node functions of the ANFIS network except piecewise differentiability. While from the structure perspective, the only limitation of network configuration is that it should be a feedforward network. According to these minimal restrictions, ANFIS applications are immediate and immense in various areas [33].

Additionally, many studies in the literature introduced ANFIS as a reliable predictive tool with several applications in the petroleum industry [34–37]. Therefore, ANFIS was selected in this study to build the predictive model of UCS. ANFIS is an integrated supervised learning algorithm for processing data that combines the concepts of neural networks and fuzzy logic utilizing a fuzzy inference system [38, 39]. Such a combination between neural networks and fuzzy logic provides an advantage of easy translation of the final system into a set of if–then rules, and the fuzzy system can be viewed as a neural network structure with knowledge distributed throughout connection strengths. It follows Takagi–Sugeno inference system which employs conventional Boolean logic (i.e., zeros and ones) while analyzing the systems using a set of fuzzy IF–THEN rules [33].

ANFIS uses a hybrid learning algorithm to identify parameters of Sugeno-type fuzzy inference systems. This learning algorithm uses the least-squares method and the backpropagation gradient descent method for training membership function parameters of the Fuzzy Inference System (FIS) to imitate a certain training dataset. The learning algorithm works in two main steps. First, in the forward phase, consequent parameters identify the least-squares estimate. In the backward phase, the error signals represented by the derivatives of the squared error concerning each node output, propagate in the backward direction into the input layer. During the second step, the parameters are updated by the gradient descent algorithm. A combination of the gradient descent algorithm and a least-squares algorithm is used as an effective tool for investigating the optimal parameters with much faster convergence. This is because it helps reduce the search space dimensions used by the backpropagation method [40].

2.5.2.1 ANFIS application to predict UCS The same filtered dataset was used to build the ANFIS-model. Data were split randomly using MATLAB for selecting the best partitioning ratio that resulted in the highest prediction accuracy. Two different algorithms were tested while building the model; these are Grid partitioning; genfis-1, and subtractive clustering; genfis-2. No reasonable results were observed when genfis-1 type was applied, however, subtractive clustering type (Sugeno-Fis type) yielded much better results with accepted accuracy. The developed ANFIS-model was optimized by testing different cluster radii and evaluating the prediction accuracy in terms of R-value and AAPE as listed in Table 4. The optimized ANFIS-model of Sugeno-Fis type with a clustering radius of 0.3 and splitting ratio of 70/30 for training and testing, respectively, resulted in acceptable prediction accuracy with an R-value exceeds 0.97 between the actual and predicted UCS values compared to the other tested options. Figure 4a and b shows a high-degree matching between the predicted UCS values using the optimized ANFIS-model and the corresponding actual ones with AAPE of 7.85% for the testing process.

2.5.3 Support vector machine (SVM)

SVM is one of the AI tools that can handle highly complex problems for both regression and classification purposes. It

Table 4 Optimizing the ANFIS-model with different cluster radii

Cluster Radius	R-value (Train)	R-value (Test)	AAPE (Train), %	AAPE (Test), %
0.3	0.98	0.97	7.75	7.85
0.5	0.93	0.94	17.53	15.91
0.7	0.92	0.92	19.73	17.56
0.9	0.84	0.85	27.89	25.61

Fig. 4 Cross-plots between the predicted and actual UCS values for **a** training process, and **b** testing process, using the developed ANFIS-model
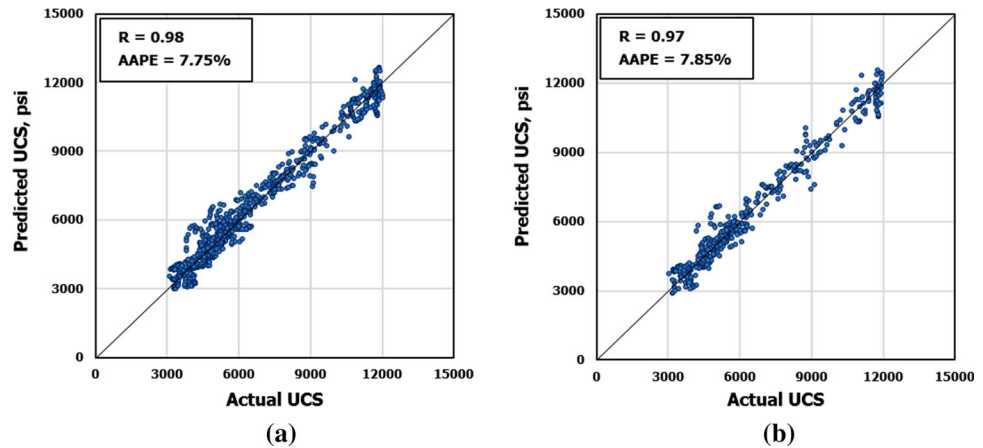
provides a larger space for training the data features in hyperplanes by transforming the data into space with a higher degree of dimensions [41]. SVM uses a statistical learning algorithm for solving quadratic-programming problems by minimizing generalization errors to reach the optimized solution [42]. SVM has been introduced as an effective predicting and classifier tool in the oil and gas industry through many studies in the literature [43–45].

2.5.3.1 SVM application to predict UCS Another model was developed using the same dataset using SVM to predict UCS of the drilled formations from the drilling parameters (ROP, SPP, GPM, ROP, T, and RPM). Splitting the data into a 70/30 ratio for training and testing the model, respectively, the model was tested using two kernel functions; polynomial and Gaussian functions, with different tuning parameters; kernel option, epsilon, lambda, verbose, and C-value). Thereafter, the model performance was evaluated in terms of R-value and AAPE between the predicted and actual UCS-values. The SVM-model with the optimized parameters listed in Table 5, yielded to predicting the UCS values with an R-value of 0.94 and AAPE of 6.5% when they are compared with the actual values during the testing process. The results show that the predicted UCS values using the SVM model agreed in a satisfactory degree with the actual ones as shown in the cross-plots; Fig. 5a and b.

The results of the testing process of the three developed models, listed in Table 6, show that the ANN-model is more accurate with AAPE of 3% between actual and

Table 5 Optimization of the SVM-model tuning parameters

Parameter	Optimized value
Kernel function	Gaussian
Kernel option	5
Epsilon	1.00E−05
Lambda	1.00E−05
Verbose	1
C-value	3000

predicted UCS values, and outperforms the two other models when tested using an unseen dataset. This observation is also confirmed by the significant match between the actual and predicted UCS values from the ANN model when a continuous profile therebetween is plotted along with the section understudy for both the training and testing processes (see Fig. 6a and b). Thus, it is more recommended to use the ANN-model as a robust tool for predicting the UCS of the downhole formations from the drilling mechanical parameters while drilling.

3 New correlation development for estimating UCS

The ANN-model was converted from a black-box into a white-box model by extracting an empirical correlation that mimics the developed network. The derived correlation

Fig. 5 Cross-plots between the predicted and actual UCS values for **a** training process, and **b** testing process, using the developed SVM-model
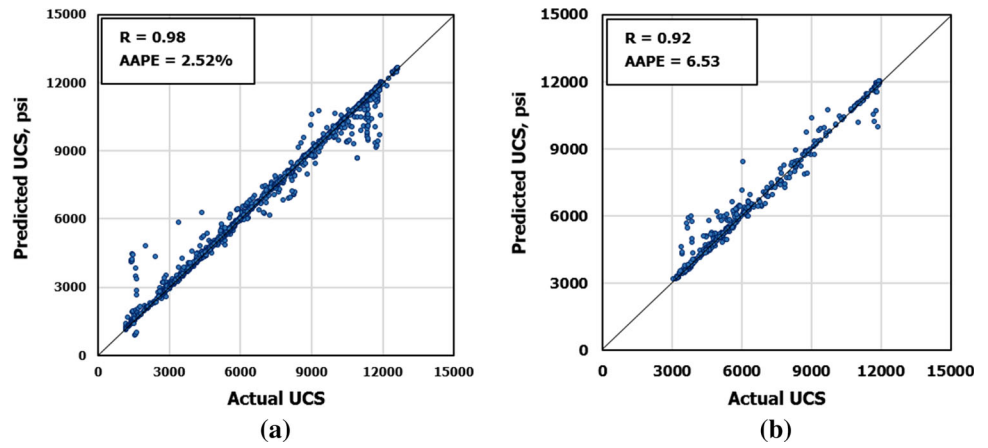

(a)

(b)

Table 6 Comparison of the prediction performance of the three developed models

AI tool	R-value (Train)	R-value (Test)	AAPE (Train), %	AAPE (Test), %
ANN	0.99	0.99	3.03	3.48
ANFIS	0.98	0.97	7.75	7.85
SVM	0.98	0.92	2.52	6.53

Fig. 6 Predicted UCS versus actual UCS for **a** training process, and **b** testing process, using the developed ANN-model
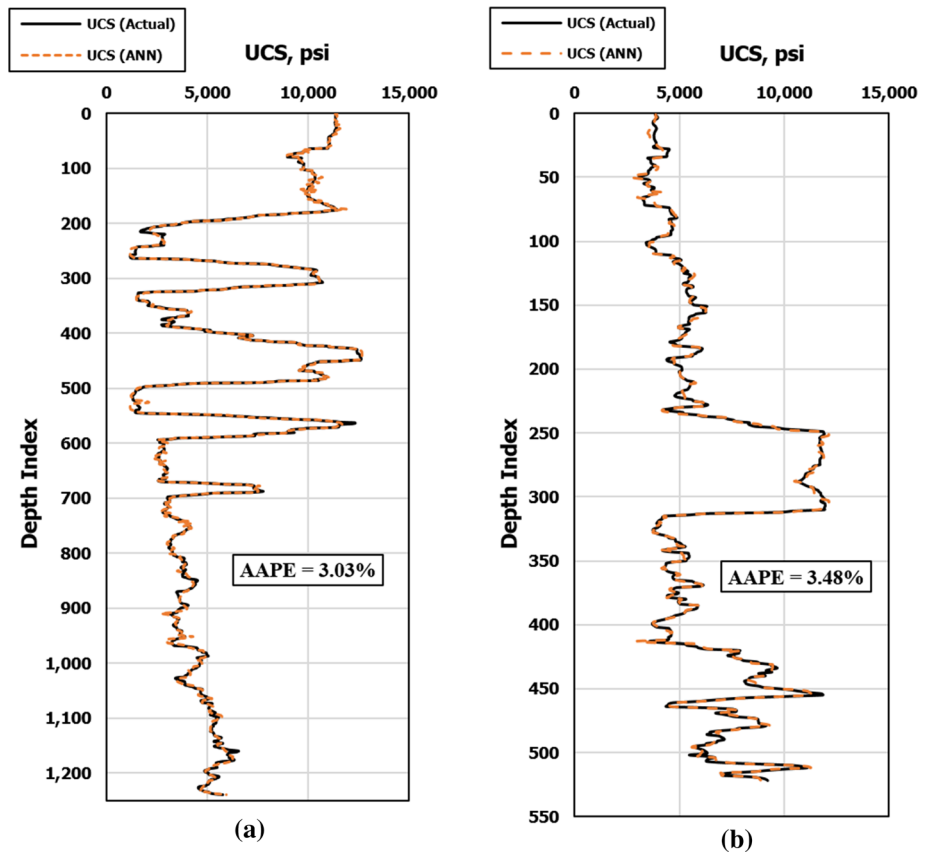

(a)

(b)

Table 7 Weights and biases of the developed ANN-based correlation

Neuron index (i)	w_1						w_2	b_1	b_2
	$w_{1_{i,1}}$	$w_{1_{i,2}}$	$w_{1_{i,3}}$	$w_{1_{i,4}}$	$w_{1_{i,5}}$	$w_{1_{i,6}}$			
1	− 1.889	1.230	− 0.861	0.646	− 1.194	2.556	4.958	− 1.638	7.474
2	11.907	− 11.266	0.561	1.617	− 1.424	− 5.727	1.269	2.832	
3	− 1.322	− 0.938	1.070	− 5.527	− 0.797	2.902	3.221	3.705	
4	− 4.292	2.893	− 2.868	4.519	− 6.530	0.308	2.405	2.754	
5	− 2.189	− 5.351	4.680	− 10.042	− 2.917	0.416	− 0.500	6.145	
6	4.306	− 3.375	− 6.420	− 3.917	3.207	0.265	2.078	− 3.121	
7	2.666	− 6.311	− 2.892	− 1.257	− 0.483	− 4.036	3.203	− 0.076	
8	− 3.030	0.239	2.324	2.228	2.098	− 1.231	− 1.806	− 0.009	
9	− 6.309	6.524	− 0.098	5.616	1.683	− 2.462	− 2.484	1.041	
10	− 8.449	3.170	− 3.687	− 1.040	− 3.802	− 0.766	− 0.979	− 1.547	
11	− 0.587	0.836	2.650	0.554	0.004	1.893	3.867	− 0.472	
12	− 4.746	− 0.182	− 0.622	2.822	− 2.892	− 1.077	4.710	4.262	
13	− 1.464	− 0.937	− 0.788	2.178	− 1.048	− 1.921	6.158	0.341	
14	1.223	1.506	0.546	− 0.068	2.420	− 1.123	5.293	− 1.502	
15	− 5.561	8.856	0.012	− 1.790	3.302	3.320	2.125	− 1.015	
16	− 5.832	3.825	4.823	1.110	1.027	− 2.951	1.154	-1.235	
17	− 0.667	0.621	− 2.345	− 3.388	− 2.869	5.157	− 4.217	− 0.217	
18	3.750	− 9.610	− 0.439	− 0.754	3.781	− 4.160	− 2.113	− 1.718	
19	− 5.108	5.871	− 0.343	6.678	4.217	5.320	− 0.861	1.115	
20	2.507	− 0.106	− 6.443	− 7.317	7.858	1.834	− 0.843	− 0.115	
21	− 2.107	− 2.828	− 4.849	1.117	6.791	− 4.217	0.728	− 0.519	
22	− 2.651	2.028	1.636	− 4.181	5.938	1.497	1.464	− 1.488	
23	2.412	− 0.008	− 1.950	− 3.774	2.204	1.937	3.913	− 2.369	
24	2.864	− 5.711	− 0.256	− 2.595	2.538	− 2.429	1.694	− 1.543	
25	2.042	− 2.085	0.456	− 4.835	0.825	− 2.308	− 2.891	− 0.193	
26	6.663	3.848	3.432	0.335	3.462	5.616	1.935	− 2.993	
27	− 2.696	3.981	2.151	5.919	− 2.138	1.477	− 2.155	− 5.736	
28	− 1.913	2.223	0.383	− 1.427	− 3.111	− 0.162	7.828	− 3.977	
29	0.814	6.396	− 0.359	− 2.239	0.011	5.454	− 1.378	4.639	
30	0.560	− 2.198	2.053	3.772	0.653	− 2.164	− 2.710	− 2.616	

includes the weights and biases that control the connections between input/hidden layers and hidden/output layers as described in Eq. 10. Table 7 lists the weights and biases of the developed correlation.

$$\text{UCS} = \left[\sum_{i=1}^{N} w_{2_i} \left(\frac{2}{1 + e^{-2\left(w_{1_{i,1}}\text{WOB}_n + w_{1_{i,2}}T_n + w_{1_{i,3}}\text{SPP}_n + w_{1_{i,4}}\text{RPM}_n + w_{1_{i,5}}\text{ROP}_n + w_{1_{i,6}}\text{GPM}_n + b_{1_i}\right)} - 1 \right) \right] + b_2$$
(10)

where, UCS_n is the normalized UCS value, N is the number of neurons in the hidden layer, w_1 is the weights between input/hidden layers, w_2 is the weights between hidden/output layers, b_1 is the biases associated with the hidden layer, and b_2 is the bias associated with the output layer.

To use the developed correlation, the input parameters should be normalized between − 1 and +1 as shown in Eq. 11.

$$Y = 2 \times \left(\frac{X - X_{\min}}{X_{\max} - X_{\min}} \right) - 1$$
(11)

where Y is the normalized input value, X is the input parameter value, $X_{\min}$ is the minimum value of the input dataset, and $X_{\max}$ is the maximum value of the input.

Thereafter, the normalized value of the output (UCS_n) can be calculated using Eq. 10 with the weights and biases listed in Table 7. Finally, UCS_n should be de-normalized to get the UCS value using Eq. 12.

$$UCS = 8174.02 \times UCS_n + 9201.04 \qquad (12)$$

4 Model validation

4.1 Phase # 1

To evaluate the performance of the developed ANN-based correlation, it was tested using the validation dataset. This dataset included 2175 points comprising drilling parameters and the corresponding UCS values of formation interval of 2175 ft. These data were collected from another well (Well B) from the same field under study. It should be noted that the validation dataset was not used while training and testing the model. The results showed an obvious agreement between the predicted and actual UCS values indicated from the R-value which exceeded 0.98 and AAPE of 4.2%. Figure 7 shows the predicted UCS values versus the actual ones along with the depth interval. The results of the validation process demonstrated the high accuracy of
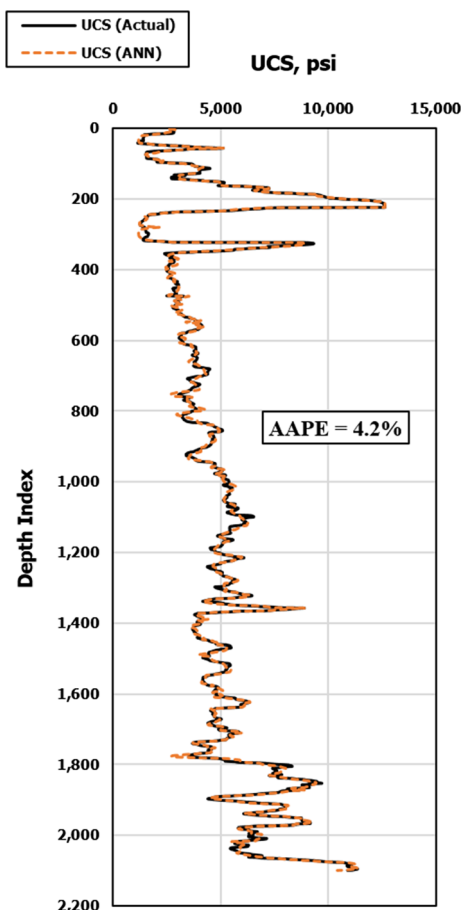
the developed ANN model in estimating the UCS values of the drilled formations while drilling.

4.2 Phase # 2

In this phase, the developed correlation was validated by comparing its performance with previously published correlations. A dataset of 2175 points from Well B was used for testing the developed ANN-correlation against some of the previously published correlations listed in Table 1. The results confirmed the robustness of the developed ANN-based correlation for predicting the UCS values with high accuracy in terms of the lowest AAPE (4.16%) between the predicted and measured UCS values, compared to the published correlations as shown in Fig. 8. Additionally, the results of the ANN-based correlation were plotted in a cross plot versus those obtained from Eq. 1 (see Fig. 9) which resulted in the lowest AAPE (24.71%) among the other published correlations. Figure 9 shows the significant deviation of Eq. 1 results from the forty-five line which indicates that its results are far from the actual UCS values. On the other hand, the results of the developed ANN-based correlation significantly coincided with the forty-five line to confirm the high accuracy of its prediction process.

5 Conclusions

This study presents a novel technique for predicting UCS values of the downhole formations while drilling in a time- and cost-effective way. This involved developing new AI-based models to predict UCS based on the low-cost mechanical drilling parameters other than using the high-cost well-log data. Besides, the usual availability of the mechanical data during the drilling operation makes it possible to predict the UCS values on a real-time basis using the developed correlations. In this study, three
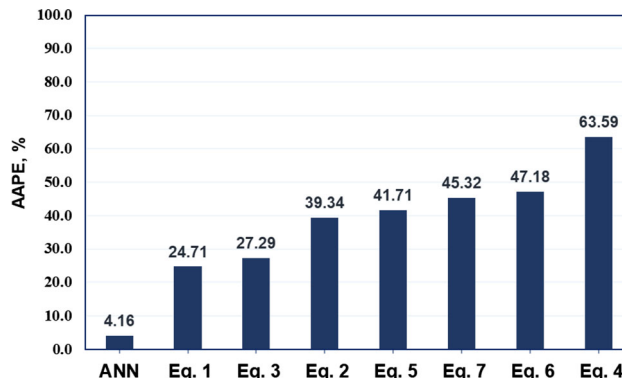
Fig. 7 Predicted UCS versus actual UCS for the validation process using data from Well B

Fig. 8 Comparison of the prediction accuracy of the developed ANN-based correlation against the preciously published correlations in terms of AAPE, %
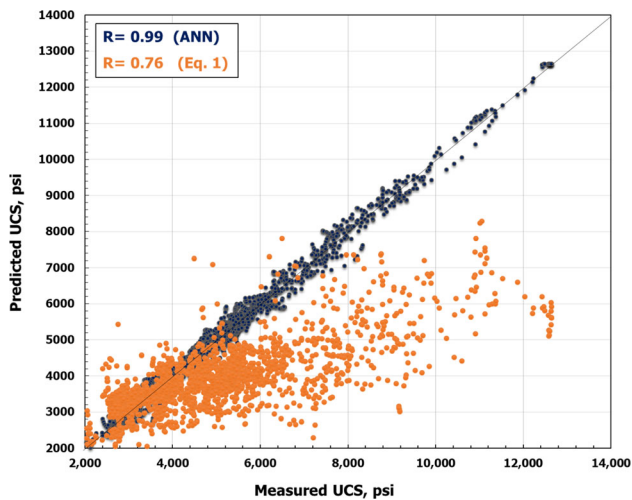
Fig. 9 UCS prediction using the developed ANN-based correlation versus the published correlations

different AI tools (ANN, ANFIS, and SVM) were applied to develop the proposed models using the mechanical drilling parameters (WOB, T, ROP, SPP, RPM, and GPM) as inputs for these models. The models were developed using 1771 actual data from a Middle Eastern field to be optimized. The ANN-model outperformed ANFIS- and SVM-models with a considerable match between the predicated and actual UCS values inferred from the high R-value of 0.99 and AAPE of 3.48%. An empirical correlation was then extracted to convert the model into a white-box model. The results of the validation process of the ANN-based correlation confirmed its robustness to predict the UCS values of the drilled formations while drilling with high accuracy. This was indicated from the significant match between the predicted UCS values and the actual ones with R-value exceeded 0.98 and AAPE of 4.2%.

The performance of the developed models is guaranteed whenever they are applied using a dataset within the same range of the data used for training the AI model. For other fields with different data ranges and geologic features, it is recommended to update the models with these data to update the network weights to guarantee a viable prediction process with high accuracy.

Compliance with ethical standards

References

1. Chau KT, Wong RHC (1996) Uniaxial compressive strength and point load strength of rocks. Int J Rock Mech Min Sci Geomech Abstracts 33(2):183–188. https://doi.org/10.1016/0148-9062(95)00056-9
2. Fjar E, Holt RM, Raaen AM, Horsrud P (2008) Petroleum related rock mechanics. Elsevier, Amsterdam
3. Shi X, Meng Y, Li G, Li J, Tao Z, Wei S (2015) Confined compressive strength model of rock for drilling optimization. Petroleum 1(1):40–45.
4. Liu H (2017) Rock mechanics. In: Principles and applications of well logging. Springer, Berlin, Heidelberg. pp. 237–269. https://doi.org/10.1007/978-3-662-54977-3
5. Abdulraheem A, Ahmed M, Vantala A, Parvez T (2009) Prediction of rock mechanical parameters for hydrocarbon reservoirs using different artificial intelligence techniques. In: SPE Saudi Arabia section technical symposium. Society of Petroleum Engineers., https://doi.org/10.2118/126094-MS
6. Militzer H, Stoll R (1973) Einige Beitrageder geophysics zur primadatenerfassung im Bergbau. Neue Bergbautechnik, pp 21–25
7. Golubev AA, Rabinovich G (1976) Resultaty primeneia appartury akusticeskogo karotasa dlja predeleina proconstych svoistv gornych porod na mestorosdeniaach tverdych isjopaemych. Prikl. Geofiz. Moskva 73:109–116
8. Chang C (2004) Empirical rock strength logging in boreholes penetrating sedimentary formations. J. Eng. Geol. 7(3):174–183
9. Mostofi M, Rahimzadeh H, Shahbazi K (2011) The development of a new sonic correlation for UCS estimation from drilling data. Pet Sci Technol 29(7):728–734. https://doi.org/10.1080/10916460903452025
10. Nabaei M, Shahbazi K (2012) A new approach for predrilling the unconfined rock compressive strength prediction. Pet Sci Technol 30(4):350–359. https://doi.org/10.1080/10916461003752546
11. Amani A, Shahbazi K (2013) Prediction of rock strength using drilling data and sonic logs. Int J Comput Appl 81(2):5–10. https://doi.org/10.5120/13982-1986
12. Asadi A (2017) Application of artificial neural networks in prediction of uniaxial compressive strength of rocks using well logs and drilling data. Proc Eng 191:279–286
13. Tariq Z, Abdulraheem A, Mahmoud M, Elkatatny S, Ali AZ, Al-Shehri D, Belayneh MW (2019) A new look into the prediction of static Young's modulus and unconfined compressive strength of carbonate using artificial intelligence tools. Pet Geosci 25(4):389–399
14. Hassanvand M, Moradi S, Fattahi M, Zargar G, Kamari M (2018) Estimation of rock uniaxial compressive strength for an Iranian carbonate oil reservoir: Modeling vs artificial neural network application. Pet Res 3(4):336–345. https://doi.org/10.1016/j.ptlrs.2018.08.004
15. Amani A, Shahbazi K (2013) Prediction of rock strength using drilling data and sonic logs. Int J Comput Appl 81(2):7–10
16. Jackson CE, Heysse DR (1994) Improving formation evaluation by resolving differences between lwd and wireline log data. In: SPE annual technical conference and exhibition. Society of Petroleum Engineers
17. Dawson R (2011) How significant is a boxplot outlier? J Stat Educ 19
18. Zhou H, Deng Z, Xia Y, Fu M (2016) A new sampling method in particle filter based on Pearson correlation coefficient. Neurocomputing 216:208–215. https://doi.org/10.1016/j.neucom.2016.07.036
19. Tariq Z, Elkatatny S, Mahmoud M, Ali AZ, Abdulraheem A (2017) A new technique to develop rock strength correlation

using artificial intelligence tools. In SPE reservoir characterization and simulation conference and exhibition; Society of Petroleum Engineers: Calgary, AB, Canada. https://doi.org/10.2118/186062-MS

20. Elkatatny S, Tariq Z, Mahmoud M, Abdulraheem A, Mohamed I (2018) An integrated approach for estimating static Young's modulus using artificial intelligence tools. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3344-1

21. Gowida A, Moussa T, Elkatatny S, Ali A (2019) A hybrid artificial intelligence model to predict the elastic behavior of sandstone rocks. Sustainability 11(19):5283

22. Anifowose FA, Labadin J, Abdulraheem A (2017) Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. J PetSci Eng 151:480–487. https://doi.org/10.3390/su11195283

23. Jamshidi E, Arabjamaloei R, Hashemi A, Ekramzadeh MA, Amani M (2013) Real-time Estimation of Elastic Properties of Formation Rocks Based on Drilling Data by Using an Artificial Neural Network. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 35(4):337–351

24. Zorlu K, Gokceoglu C, Ocakoglu F, Nefeslioglu HA, Acikalin SJEG (2008) Prediction of uniaxial compressive strength of sandstones using petrography-based models. Eng Geol 96(3–4):141–158

25. Cevik A, Sezer EA, Cabalar AF, Gokceoglu C (2011) Modeling of the uniaxial compressive strength of some clay-bearing rocks using neural network. Appl Soft Comput 11:2587–2594. https://doi.org/10.1016/j.asoc.2010.10.008

26. Lippman RP, Lippman RP (1987) An intrduction to computing with neural nets. In: Mag A (ed.) IEEE ASSP Magazine. IEEE. pp. 4–22, https://doi.org/10.1109/MASSP.1987.1165576

27. Hinton GE, Osindero S, Teh Y-W (2006) A Fast Learning Algorithm for Deep Belief Nets. Neural Comput 18:1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

28. Graves A, Liwicki M, Fernandez S, Bertolami R, Bunke H, Schmidhuber J (2009) A Novel Connectionist System for Unconstrained Handwriting Recognition. IEEE Trans Pattern Anal Mach Intell 31:855–868. https://doi.org/10.1109/TPAMI.2008.137

29. Nakamoto P (2017) Neural Networks and Deep Learning: Deep Learning Explained to Your Granny a Visual Introduction for Beginners Who Want to Make Their Own Deep Learning Neural Network (Machine Learning); CreateSpace Independent Publishing Platform: Scotts Valley. CA, USA

30. Jang JS, Sun CT (1995) Neuro-fuzzy modeling and control. Proc IEEE 83(3):378–406

31. Wang T, Ma M, Jianbin Q, Gao H (2020) Event-triggered adaptive fuzzy tracking control for pure-feedback stochastic nonlinear systems with multiple constraints. IEEE Trans Fuzzy Syst. https://doi.org/10.1109/TSMC.2020.3013744

32. Qiu J, Ma M, Wang T (2020) Event-triggered adaptive fuzzy fault-tolerant control for stochastic nonlinear systems via command filtering. IEEE Trans Syst Man Cybern Syst 2:1493–1499. https://doi.org/10.1109/FUZZY.1996.552396

33. Jang JS (1996) Input selection for ANFIS learning. In: Proceedings of IEEE 5th international fuzzy systems (Vol. 2, pp. 1493—1499). https://doi.org/10.1109/fuzzy.1996.552396

34. Aghli Ghasem, Moussavi-Harami Reza, Mortazavi Saiedollah, Mohammadian Roohangiz (2019) Evaluation of new method for estimation of fracture parameters using conventional petrophysical logs and ANFIS in the carbonate heterogeneous reservoirs. J Petrol Sci Eng 172:1092–1102

35. Gowida A, Elkatatny S, Al-Afnan S, Abdulraheem A (2020) New computational artificial intelligence models for generating synthetic formation bulk density logs while drilling. Sustainability 12(2):686. https://doi.org/10.3390/su12020686

36. Ja'fari A, Kadkhodaie-Ilkhchi A, Sharghi Y, Ghaedi M (2014) Integration of adaptive neuro-fuzzy inference system, neural networks and geostatistical methods for fracture density modeling. Oil & Gas Science and Technology-Revue d'IFP Energies nouvelles 69(7):1143–1154. https://doi.org/10.2516/ogst/2012055

37. Anifowose FA, Labadin J, Abdulraheem A (2013) Prediction of petroleum reservoir properties using different versions of adaptive neuro-fuzzy inference system hybrid models. Int. J. Comput. Inf. Syst. Ind. Manage. Appl 5:413–426

38. Vilela LRM, Oluyemi G, Petrovski A (2019) A fuzzy inference system applied to value of information assessment for oil and gas industry. Decision Making: Applications in Management and Engineering 2(2):1–18

39. Stojcic M, Stjepanovic A, Stjepanovic Djordje (2019) ANFIS model for the prediction of generated electricity of photovoltaic modules. Decision Making: Applications in Management and Engineering 2(1):35–48

40. Walia N, Singh H, Sharma A (2015) ANFIS: Adaptive neuro-fuzzy inference system-a survey. Int J Comput Appl 123(13):515–535

41. Xuegong Z (2000) Statistical learning theory and support vector machines. Acta Automatica Sinica 26(1):32–42

42. Xuegong Z (2000) Introduction to statistical learning theory and support vector machines. Acta Automatica Sinica 26(1):32–42

43. Gholami R, Shahraki AR, Jamali Paghaleh M (2012) Prediction of hydrocarbon reservoirs permeability using support vector machine. Math Probl Eng. https://doi.org/10.1155/2012/670723

44. Rostami H, Khaksar Manshad A (2014) A new support vector machine and artificial neural networks for prediction of stuck pipe in drilling of oil fields. J Energy Technol 136(2):67. https://doi.org/10.1115/1.4026917

45. Zhong YH, Li R (2009) Application of principal component analysis and least square support vector machine to lithology identification. Well Logging Technol 33(5):425–429

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.