



# A systematic review of data science and machine learning applications to the oil and gas industry

Zeeshan Tariq<sup>1</sup> · Murtada Saleh Aljawad<sup>1</sup> · Amjed Hasan<sup>1</sup> · Mobeen Murtaza<sup>1</sup> · Emad Mohammed<sup>1</sup> · Ammar El-Husseiny<sup>1</sup> · Sulaiman A. Alarifi<sup>1</sup> · Mohamed Mahmoud<sup>1</sup> · Abdulazeez Abdurraheem<sup>1</sup>

Received: 5 July 2021 / Accepted: 13 September 2021 / Published online: 24 September 2021  
© The Author(s) 2021

## Abstract

This study offered a detailed review of data sciences and machine learning (ML) roles in different petroleum engineering and geosciences segments such as petroleum exploration, reservoir characterization, oil well drilling, production, and well stimulation, emphasizing the newly emerging field of unconventional reservoirs. The future of data science and ML in the oil and gas industry, highlighting what is required from ML for better prediction, is also discussed. This study also provides a comprehensive comparison of different ML techniques used in the oil and gas industry. With the arrival of powerful computers, advanced ML algorithms, and extensive data generation from different industry tools, we see a bright future in developing solutions to the complex problems in the oil and gas industry that were previously beyond the grip of analytical solutions or numerical simulation. ML tools can incorporate every detail in the log data and every information connected to the target data. Despite their limitations, they are not constrained by limiting assumptions of analytical solutions or by particular data and/or power processing requirements of numerical simulators. This detailed and comprehensive study can serve as an exclusive reference for ML applications in the industry. Based on the review conducted, it was found that ML techniques offer a great potential in solving problems in almost all areas of the oil and gas industry involving prediction, classification, and clustering. With the generation of huge data in everyday oil and gas industry activities, machine learning and big data handling techniques are becoming a necessity toward a more efficient industry.

**Keywords** Oil and gas industry · Systematic review · Machine learning · Future of data science in oil and gas

## Introduction

Artificial Intelligence (AI) is the field that integrates computational power with human intelligence to produce smart and reliable solutions to extremely nonlinear and highly complicated problems. AI is the field of science that allows computers to think and decide on their own. Machine learning (ML) is a subset of AI that provides statistical tools to explore and analyze big data. ML is comprised of further subsets such as supervised, unsupervised, and reinforced learning. Supervised learning is the data learning technique applied when some past or labeled data is available for future forecasting by function approximation. The unsupervised learning technique is the machine learning technique when

the past labeled data is unavailable and is usually used for clustering purposes. Reinforced learning is the combination of supervised and unsupervised learning techniques in which some part of the data is labeled and some part is not.

In the last two decades, engineering journals have reported numerous articles utilizing ML for regression, function approximation, and classification problems. With the development of intelligent oilfields and big data technology, the adoption of the ML method has gained new vitality for the study of problems in the oilfield development process. With the advent of computing techniques, several correlations utilizing ML have come to the fore, especially in reservoir characterization (Anifowose 2012; Fatai A Anifowose et al. 2013a, b), reservoir engineering (Al-Marhoun and Osman 2002; Gharbi et al. 1999; Gharbi and Elsharkawy 1999); and reservoir geomechanics (Tariq et al. 2017a, b) and many other areas in petroleum engineering applications.

The most repeated question that ML petroleum researchers faced in their everyday life is that ML models are usually

✉ Sulaiman A. Alarifi  
salarifi@kfupm.edu.sa

<sup>1</sup> King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

limited to the data set tested, so how to globalize this and produce more general correlations? ML applications have common limitations and challenges that hinder the globalization of the created models, such as overfitting, coincidence, excessive training, lack of interpretability of results, and bias. Besides, these models require a large amount of data that is not available in many cases.

Overfitting is considered the most common problem in ML applications. This is due to the lack of an appropriate amount of data to be used for training. To overcome this issue, the ratio of data points to the total number of weights used by the connections ( $\rho$ ) was used to lessen the effect of insufficient data. The coincidence effect is another issue that accompanies the AI supervised learning models as they try to match a specific dataset, so there is a probability of getting a good match by coincidence. This also can happen in other regression analysis techniques, which require working on methods to minimize that occurrence (Livingstone et al. 1997). Overtraining can happen when there is no clear stopping stage for the training. The error may stay decreasing by updating the model structure, including the weights. The real risk, in that case, is that the model can be more complex to fit a specific dataset, becoming impossible to generalize after that. A training methodology named “early stopping” uses a control set that monitors the training process to overcome this. If error begins to rise, the early stopping will end the training process. Other techniques are being used to save time and effort, such as reinforcement learning with in-stream supervision, such as generative adversarial networks that monitor the learning of two competing networks to better understand the model concept (Hossain 2018).

The availability of large datasets is also a concern, which affects the training accuracy and goodness of the model. If the gathered data is limited, a methodology like single-shot learning is implemented, in which the AI model is pre-trained on a similar dataset and is enhanced with experience.

Interpretability is the key to data analysis. AI models are not that simple, and even in some cases, it is impossible to interpret the results even in modeling small linear problems. The single connections in the models do not alone affect results, but the whole combined connections do. One of the methods developed to help in that regard is the local interpretable model and its agnostic explanations, which try to detect which parts of the raw data the model depends on mostly for estimations. In the generalized additive models' method, the separation between model features enhances each feature's interpretation.

The lack of AI models' generalization ability is a major limitation that delays the widespread of AI in the oil and gas industry. It is hard for many models to be used in circumstances different from those used in building the original model (Virginia 2018). Additional resources are to be utilized each time for training new datasets, even if they were

similar to previous cases (Ramamoorthy and Yampolskiy 2018). The reusability of the ML models is also quite challenging. Usually, the trained models on one geological field are less reliable when applied to other geological fields. It is highly recommended to implement the model when the input parameters of the given dataset lay within the range of the input parameters on which the model is to be implemented (Mohaghegh 2017).

Lastly, the effect of bias cannot be ignored and sometimes is hard to be detected and mitigated. Many researchers are solving the issues related to AI bias by understanding the model's objective and its associated results. Using model-independent perturbations by substituting the inputs with random values obtained from a normal distribution will help avoid biases (Samek et al. 2018). Table 1 provides a summary of all limitations of AI and ML models.

Covering all AI and ML application content to the oil and gas industry in a single article is a challenge, so we focused this article on the application of AI and ML in petroleum exploration, drilling, production, stimulation, and reservoir characterization. The issues highlighted in this article include comparing commonly used AI techniques, how AI can be used as a standalone technique, and how to make the AI model generalized. Furthermore, this review highlights the present status of data-driven machine learning predictive models. It also addresses the commonly asked questions related to machine learning and future research.

## AI applications challenges

### AI as a standalone predictive tool

Should AI be a standalone predictive tool? Or it can be combined with analytical models, numerical models, statistical and probabilistic approaches, numerical simulation software's, imaging software's, etc. AI applications are recently getting more attraction as an enabler of state-of-the-art technologies for digitalization among industries, including digital twins (data and physics). Because physics-based models are based on simplifying assumptions to formulate the problem, the models lack the physics controlling the processes (Rasheed et al. 2018). However, artificial intelligence is like a black box, which does not explain the model outcomes. Since data contains the known and unknown part of physics, constructing data-driven models incorporate the full physics behind it. However, the black box nature inhibited these models from prevailing in critical systems, which have a culture of zero error tolerance, such as what oil and gas have in field operations. However, with the increased number of applications with proved concepts in oil and gas, the industry leaders are now emphasizing these applications' potential in optimizing operations like predictive maintenance.

**Table 1** A summary of the limitations of AI and ML models

Limitation	Reason	Solution	References
Overfitting	Lack of an appropriate amount of data to be used for training	Using the ratio of input data points to the total number of network weights used by the connections ( $\rho$ )	Andrea and Kalayeh (1991) Livingstone et al. (1997)
Coincidence	Getting a good match by coincidence for a specific dataset	Using discriminant analysis	Livingstone and Manalack (1993)
Overtraining	When the error keeps decreasing by updating the model structure and the model can be more complex to fit a specific dataset	A training methodology that is named “early stopping” can be used Reinforcement learning with in-stream supervision, for example, the generative adversarial networks	Hossain (2018)
Data availability	Sometimes the gathered data is limited	Single-shot learning in which the AI model is pre-trained on a similar dataset and then is enhanced with experience	Weyrauch and Herstatt (2016)
Interpretability	The single connections in the models do not affect alone but the whole model connections combined affect results	Local interpretable model and its agnostic explanations The generalized additive models method	Shabbir et al. (2018) Bas (2016)
Generalization	Model failure in the circumstances different from the set of circumstances, which were used in building the original model	Additional resources are to be utilized for training new datasets	Virginia (2018) Ramamoorthy and Yampolskiy (2018)
Bias	The nature of black-box models makes it to be prone to biases	Using model-independent perturbations	Samek et al. (2018)

AI algorithms and statistical models have advanced significantly, leading to computers overtaking creative tasks such as art drawings, summarizing texts, translation, and even interpreting languages. Deep learning or deep neural networks were used for image classification. Restricted Boltzmann machines are utilized as stochastic networks that understand the distribution of inputs in the supervised and unsupervised way; thus, they can be used as powerful tools in detecting anomalies (Evangelatosorn and Payne 2016). Table 2 provides a summary of the comparison between physics-based and data-driven models.

To benefit from the advantages of physics-based and AI models', hybrid models that engage the interpretability and reliable mathematical concepts of physics-based models are proposed. Rasheed et al. (2020) explained the “digital twin” models and gave an example of Kongsberg's dynamical digital twin for oil and gas. The model linked process schematics and virtual 3D graphics of an oil and gas production facility connected to real-time data from sensors where the data-driven models are used. It could also be provided with synthetic data generated from simulators, representing the physics part. Humans can interact with the model using an avatar to give expert opinions. Finally, the model can also do “what if?” scenarios using digital siblings, which are copies of the physical asset. Al-Hajri et al. (2020) suggested a coupled machine learning and probabilistic models to predict the scale prediction and plan for inhibition. The coupled model was able to predict the scale and quantify the data-driven models' goodness, then quantify the savings. Sun et al. (2019) used a coupled model of ML image processing techniques and reservoir simulation to achieve better reservoir characterization and overcome classical methods' drawbacks. Shahkarami et al. (2014) used a Surrogate Reservoir Model (SRM) to assist the reservoir history matching process. Table 3 shows a summary of the hybrid models applied to the oil and gas industry.

Based on the review of the coupled models, it is evident that coupling with other methods such as analytical models, numerical models, statistical and probabilistic approaches, and imaging software is beneficial for more robust, accurate, and unbiased models. Figure 1 shows how the different models should interact to achieve the desired goals.

### What is needed from AI in the oil and gas industry?

Many oil and gas industry giants are currently applying AI in oil and gas operations. AI advances made it suitable for several applications such as precision in drilling and automation, saving oil and gas producers' time and money. These advances are going to serve different aspects of the oil and gas industry, such as:

**Table 2** Physics-based against data-driven models

Models	Advantages	Disadvantages	References
Physics driven	Strong basics, based on existing solid knowledge Easy to interpret Can detect errors and uncertainties and avoid them Lower probability of bias Easy to be generalized to other problems Fundamental relationships give insight and help in understanding Valid prediction at a full range of model coverage	Hard to integrate historical or archived data with the models Prone to numerical instability as a result of having complex boundary conditions and inputs uncertainties Vast physics knowledge in the domain is required High computational power requirement, so it suffers if used for real time Assumptions need to be set in advance	Rasheed et al. (2020) Boschert and Rosen (2016)
Data-driven	Considers the historical data and experiences into the model Able to stably make predictions after training Does not require knowledge of the domain as it depends mainly on data Deals with heterogeneous data Able to enhance performance over time Can detect complicated relationships and patterns	Black box nature and interpretability issues Cannot detect errors or uncertainties Affected by bias in data Not easy to generalize Data availability is the main concern It is an approximation Lower performance outside the scope of the training data Hard to predict critical conditions or extremes	

**Table 3** A summary of selected hybrid models in the oil and gas industry

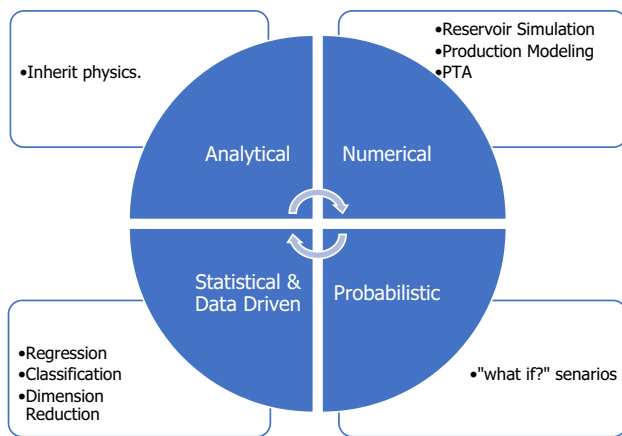
Model	Application	Components	Reference
Digital twin	Drilling engineering	Sensor data in near-real time (Data), Synthetic data generated from simulators (Physics) Humans to interact using avatar (Expert) Digital siblings for “what if?” scenarios	Rasheed et al. (2020)
ML & probabilistic approach	Oil and gas production	Calculated input parameters using existing principles (Physics) Classification using ML models (Data) A probabilistic model to quantify the uncertainty associated with each method Cost model to predict financial impact	Al-Hajri et al. (2020)
ML & digital rock analysis (DRA)	Reservoir characterization	Rock image acquisition Image processing using ML models (Data) Numerical simulation (Physics) Result Analysis	Sun et al. (2019)
Surrogate reservoir model (SRM)	Reservoir characterization	Multiple neuro-fuzzy systems Numerical simulation model Spatiotemporal database	Shahkarami et al. (2014)

### Precise drilling

Drilling activities are always accompanied by high risk and a high level of uncertainty. AI techniques coupled with the big data recorded by the smart sensors mounted on drilling strings such as pressure, temperature, and seismic surveys in real time can be used to overcome these challenges. Precise drilling using AI can enhance the control level of the rate of penetration and identify risks in advance.

### Production optimization

Every oil and gas company focuses on production optimization and efficiency, which eventually increases profits with the help of AI, automated pattern recognition, and classification to prepare production data for generating analytics. Estimation and prediction models can then be built based on the refined data. It can also isolate the effects of the reservoir from the production control



**Fig. 1** A figure shows how different models interact to achieve objectives

responses such as gas lift rates, choke openings, network routing, and artificial lift methods.

### Reservoir management

Multiple teams from several aspects such as seismic, geology, reservoir, and production engineering are required to collaborate to achieve better reservoir management. The AI models can be trained with historical data of seismic surveys, geological descriptions, and production methodologies and then can be applied in the characterization or modeling of reservoirs and field monitoring.

### Inspections

Frequent inspections are scheduled for detecting abnormal equipment performance to prevent failures of the equipment and potential accidents. That is why companies are looking for automated and smart detective approaches. Robots driven by AI models can help investigate abnormal equipment behavior by identifying anomalies using techniques such as pattern recognition. Besides, drones can inspect pipelines and offshore facilities that can detect, in real time, cracks or leaks in pipelines. They can also help in case of an emergency, such as gas leaks. In certain situations, these robots can intervene in emergency cases and use the procedure, which applies to that case, which will elevate the company's safety measures.

### Chatbots

AI-powered chatbots can help engineers and scientists by digging in a database or archive of historical data, suggesting possible solutions to problems, providing correct standards of job execution, or help in teaching junior staff using

natural language processing. Jacobs (2019) discussed three newly released chatbots in the oil and gas industry: Sandy, Nesh, and Ralphie. They are designed intentionally to provide answers to oil and gas professionals' complex questions. These are also named virtual assistants that use artificial intelligence (AI) natural language processing (NLP), which has quickly entered the market through the tech giants Amazon, Apple, and Google, which enabled many millions of people to engage in dialogue with laptops, smartphones, and speakers.

### Facilities monitoring

Intelligent cameras can reduce potential damage by detecting hazardous activities such as smoking in dangerous areas. They can be trained using photos and recordings of dangerous activities to alert the staff or take predefined actions. Moreover, they can detect if the employees are watering their protective PPE or not. Using this approach will help enhance safety management.

### Commonly used machine learning techniques in oil and gas industry

Several ML techniques such as ANN, FL, SVM, DT, RF, KNN, RNN, CNN, and fuzzy C-means clusters are widely used in different applications of oil and gas. Table 4 summarizes some of the algorithms with their advantages and disadvantages.

### Exploration and geosciences

The applications discussed here include fault and salt-body delineation, Petrofacies classification, and well correlation. We also discuss potential further development in emerging applications.

### Fault and salt-body delineation

Accurate fault detection and delineation of the salt-body boundary from 3D seismic data are essential for building a realistic 3D reservoir model (Bahorich and Farmer 1995, b; Melville and Guruswamy 2002). Seismic attributes analysis has been traditionally used to map faults and salt bodies. Some examples of such attributes include the semblance (Marfurt et al. 1998), coherence (Bahorich and Farmer 1995a, b; Qi et al. 2017), edge detection (Di and Gao 2014), and seismic curvature (Di and Gao 2016; Somasundaram et al. 2017). Due to the complex geology and the noise level frequently encountered in 3D seismic data, the use of multiple seismic attributes is frequently needed to detect faults or salt-body geometry (Berthelot

**Table 4** Commonly used machine learning tools with their advantages and disadvantages

Algorithms	Application	Advantages	Disadvantages	References
ANN Artificial neural networks MLP multi-layer perceptron FF feed forward RBF radial basis function CN convolutional FN functional PN probabilistic	Regression/ classification/ clustering	Learning algorithms are simple With available data it can superior any other model Does not depend on linearity of any function Can be used for problems which are hard or not practical to get a formula for  ANN can be used for tasks that linear programs cannot handle Due to the parallel nature of the networks, they can proceed without problems even if an element fails They can learn from experience and avoid reprogramming Applicable in most problems It is tolerant to faults Can learn from experience Effect of small changes is minor Handle nonlinear data Excellent in fitting applications	They are “black box” in nature so it is not easy to be understood or interpreted Lack the ability of generalization as they are exposed to overtraining and might memorize specific data For small datasets, the predictions are not acceptable  Neural networks need training to be used The architecture is different from problem to another For big networks, the training and processing time is high	Mohaghegh (2000)          Sibai et al. (2011)
FL Fuzzy logic	Classification/clustering	Quick, easy, strong and effect of environment changes is minor Gives a combination of numeric and symbolic picture of systems Can handle problems with strict conditions or even without exact solution Can be described with few data points or approximated datasets Simple reasoning, application and can deal with uncertainties and nonlinearity It is able to detect hyperplane of optimal separation Deals with higher degrees of dimensionality Its kernels can learn precise concepts as they have infinite Vapnik–Chervonenkis dimension Works well usually	Needs parallel processing abilities  Exposed to overfitting Can be trapped in local optimum solution Consumes large time in training If mathematical model is existing, FL is used only in case of low computational capabilities Not easy to prove the system characteristics as it lacks mathematics  Lacks robustness	Raj (2019)  Liu and Lang (2019)  Kar et al. (2014)  Raj (2019)  Omar et al. (2013)

Table 4 (continued)

Algorithms	Application	Advantages	Disadvantages	References
SVM Support vector machine	Regression/ classification/ clustering	Provide high accuracy classifiers. Overfitting occurrence is little, excellent in dealing with noise Preferred for text classification applications that are normally high dimensional problems Intensive memory consumption Get useful information from little datasets Has generalization capabilities	It is a binary classification technique, so it needs pairwise classification to perform multi-class classification that means one class against all others, for all classes Runs slowly and require high computational power Low performance with big data or multi-classification tasks Kernel function parameters affect the performance	Anifowose and Abdulraheem (2011)      Liu and Lang (2019)
DT Decision tree	Regression/ classification/ clustering	Easy to understand and can be interpreted Data preparation is fast Can deal with numerical and categorized data It has white box interpretable model Statistical tests can be used to validate the model accuracy Robust Efficiently handle huge data in a little time Nonlinearity among parameters do not affect the performance of DT Interpretable and explainable In case of few predictor variables, it is easy to understand Can be used in building models that contain special data types, such as text	Even for most simple concepts, the learning of an optimal DT is known as NP-complete Complex DT models cannot generalize the data well Fails to learn some concepts as it is not easy for DT to express them	Omar et al. (2013)      Devi et al. (2015)  Omar et al. (2013)
KNN $K$ nearest neighbors	Classification/ clustering	Classes do not need to be linearly divisible Modest and powerful Understandable and easy to implement technique Can be trained quickly Robust in case of associated noise It is mainly well suited for multimodal classification High performance Arise variable measures Quick in implementation Less complex Robust with noisy data Can learn in increments	Tends to disregard the attributes importance Sluggish and expensive Sensitive to local Data structure Memory restriction Supervised type of learning Sluggish algorithm  Computationally expensive Overfit issues Does not depend on variables Disregards original geometry of data Low performance with attribute-related training data	Somvanshi et al. (2016)  Devi et al. (2015)      Taunk et al. (2019)  Taunk et al. (2019)  Liu and Lang (2019)
RF Random forest	Classification/ clustering			

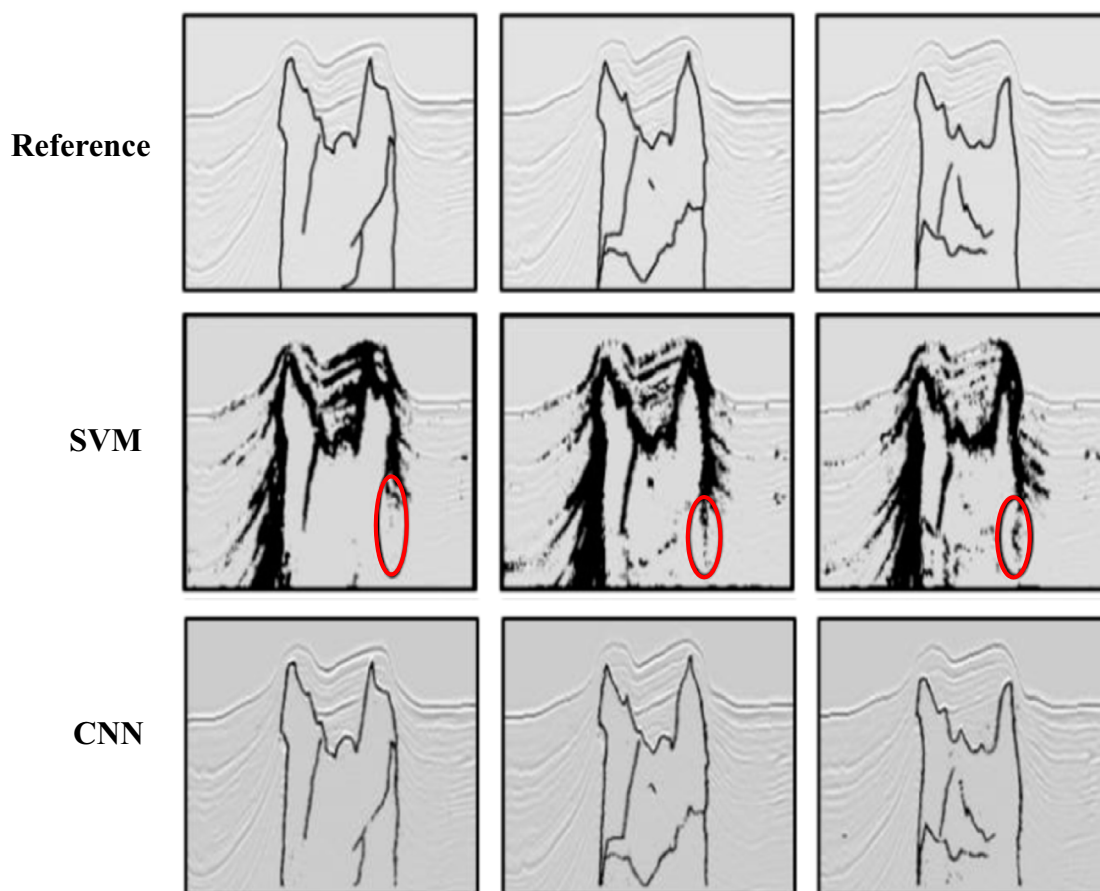
Table 4 (continued)

Algorithms	Application	Advantages	Disadvantages	References
K-means	Classification/ clustering	Data point is allowed to exist in different clusters Normal representation of the behavior of genes  Changeable model that can adapt different dataset distribution If training data increase, the parameters number does not change	Need to define number of clusters $c$ Membership cutoff value has to be set Initial assignment of centroids affects the clusters  In some cases, the convergence is slow	Omar et al. (2013)  Omar et al. (2013)
Fuzzy C-means	Classification/ clustering	Modest and easy to-understand the working algorithm As a topological clustering unsupervised technique, it can deal with dataset nonlinearity Unique in directionality reduction being able to convert high dimensions problem to 1–2 dimensions	Time consuming technique	Omar et al. (2013)
RNN Recurrent neural network	Regression/ classification/	Can record the information as activations with time Manipulate consecutive information that are random in length	It is affected by the gradient vanishing type Not able to be stacked within extra deep modeling	Pandey and Janghel (2019)
CNN Convolutional neural network	Regression/ classification/ image processing	Able to detect relevant features only from given dataset Same parameters can be utilized in different problems  Quick training No approximation techniques needed Does not require several entries in the samples Layer by layer strategy of learning makes it capable of learning the features Deals with non-labelled data and can be safe from the overfitting and underfitting issues Not affected by the fragmentation of training data thus it reduces over-smooth problem	Tuning of parameters is difficult Requires large amount of data	Pandey and Janghel (2019)
GAN Generative adversarial network	Regression/ classification/		Quality might be low Unstable training Generating discrete data is difficult	Yishuang et al. (2019) Pandey and Janghel (2019)
DBN Deep belief network	Regression/ classification/		Some pre training algorithms decrease the performance as the input data is clamped Run time is long	Yuming Hua et al. (2015)
			Lower output quality	Hu et al. (2019)



et al. 2013; Hale 2013). Several studies have then used machine learning techniques to integrate multiple seismic attributes for better detection accuracy and faster operation (Di and Gao 2014; Huang et al. 2017; Zhao et al. 2015). Support vector machine (SVM) is one common algorithm that has been used by several studies, particularly for fault detection (Guitton et al. 2017). In this case, correlation and cluster analysis are used to select the suitable seismic attributes that can best identify fault from seismic. The use of SVM could successfully improve the accuracy and efficiency of fault detection, especially in large-scale faults (Zou et al. 2019). Nevertheless, some researchers pointed out two main shortcomings of the SVM attributes-based approach (Xiong et al. 2018). Firstly, it requires precomputed attributes by experienced interpreters to map the faults, which can be labor-intensive as this step has to be repeated for each data set. Secondly, the SVM attributes-based approach can fail in zones of weak reflections,

as highlighted in Fig. 2. This can be critical for heavily faulted zones and salt-body delineation due to the frequently weak signal associated. Recent studies (Di and Gao 2016, 2014; Tschannen et al. 2020; Xiong et al. 2018) have shown that deep learning technologies such as convolutional neural networks (CNN) can help overcome the previous two shortcomings and to map complex geological structures/features. An example of improved performance of CNN versus SVM is demonstrated in Fig. 2. In the CNN approach, the network is trained based on annotated seismic images where faults or salt-body boundaries are labeled, relying more on the reflection patterns and reducing the effect of seismic noises or processing artifacts (Di and Gao 2016; Xiong et al. 2018). Additionally, the relationship between seismic reflection patterns and the target fault or salt bodies is constructed based on the original seismic amplitude, eliminating the need for precomputed attributes (Di and Gao 2014).



**Fig. 2** Comparison of salt-body boundaries delineation using the traditional multi-attribute-based support vector machine (SVM; second row) and convolutional neural network (CNN; third row) in three different inline sections (modified after Di et al. 2018). The reference

manually labeled sections are shown in the first row. The seismic sections were extracted from the synthetic SEG-SEAM dataset. Poor detection of the boundary from the SVM results is highlighted by red circles

## Petrofacies classification and fractures identification

Reservoir rocks can be classified and grouped based on their reservoir quality. Such classification can be done based on petrophysical rock properties (e.g., porosity, permeability, and pore size) and geological features (e.g., textures, diagenetic overprints, and pore types). Petrofacies are usually defined based on combining both petrophysical and geological attributes, which can be an essential tool for reservoir characterization (Avseth and Mukerji 2002). Petrofacies classification is frequently done using both core samples and wireline log data. Cores are not frequently available from all wells due to the time and cost associated, and thus several studies (Bhattacharya and Mishra 2018; Qi and Carr 2006; Sebtosheikh and Salehi 2015) have examined how machine learning algorithms can be trained on data obtained from certain cored well and then used to perform petrofacies classification in other un-cored wells. Petrofacies labels, defined as a function of depth based on the integration of well-log and core data, are used to train the models (Sebtosheikh and Salehi 2015; Silva et al. 2015). The utilized logs for facies identification are usually Gamma Ray (GR), resistivity (Rt), neutron (NPHI), density (RHOB), and lithology (PEF). In addition, other features could be extracted from these logs to improve the prediction, such as total organic matter (TOC), matrix grain density (RHOMAA), and apparent volumetric cross-section (UMA).

Earlier studies have used ANN, SVM, and RF to classify petrofacies from well logs in both sandstone and carbonate reservoirs (Silva et al. 2015; Al-Anazi and Gates 2010; Martinelli et al. 2013; Salehi and Honarvar 2014). Nevertheless, more recent studies have suggested that Gradient Boosting (GB) algorithm outperforms ANN and SVM, especially when a limited number of features are available (Silva et al. 2015). Another algorithm that has shown success is the Random Forest (RF), which reduces the computational time for the training phase compared to GB (Bhattacharya and Mishra 2018). Based on the existing literature, it seems that there is no consensus regarding the most suitable machine learning technique for petrofacies classification. This could be due to several factors, including the wide variations in the features selected or available data, as well as differences in terms of complex geology and reservoir heterogeneity. Indeed, as pointed out by Silva et al. (Silva et al. 2015), the applicability of various algorithms has to be tested for each training/testing data set to be used. One major challenge that remains for the success of machine learning in this application is to have/select the right petrophysical and geological attributes/features to distinguish between facies. Such tasks remain mainly subjective and far from being automated or objective.

Fractures and facies identification are usually made through personal judgments based on field log and laboratory core analysis data. Recently, AI has been used to identify fractures and facies in unconventional formations. Tian and Daigle (2019) could identify micro-fractures and organic matter in siliceous and carbonate-rich shale samples and find the association between them using AI. That was to automate the process of understanding micro-fractures in shale samples to make it fast and avoid personal evaluations. SEM and EDS images were used to find fractures and organic matters in intact and deformed samples. The single-shot detector (SSD) deep learning approach was used to train the data obtained from the images. Around 97% of fractures in intact samples and 92% in deformed ones were identified using SSD. Also, detected organic matter images were overlapped over detected fractures to find the associations. It was found the clear majority of micro-fractures penetrated the OM and clay minerals. It seems that the combination of the soft OM and clay and brittle materials (quartz and calcite) enhances the fracability according to the study.

## Well correlation

Correlating different reservoir units and formation tops across different wells is essential in reservoir characterization and modeling. Such a task may require significant time from experienced geologists, especially in large fields with hundreds of wells. The use of machine learning to handle this issue has been recognized many years back (Luthi and Bryant 1997). An interpreter has first to pick formation tops and perform well correlations in several wells, which will be used as a training dataset to perform interpretation in tens to hundreds of other wells. An increasing body of studies (Maniar et al. 2018; Zheng et al. 2019) has demonstrated that a deep convolutional neural network (CNN) can provide an accurate and efficient approach for well-log correlations. The most common log data used for the correlation includes gamma ray and resistivity, although any other geophysical well-log data with sufficient log character can be used. One crucial observation documented by Zheng et al. (2019) was the drastic reduction in prediction accuracy as the number and percentage of the training dataset decreases. This might be explained by the complexity of geology that would require wells covering different depositional environments and stratigraphic sequences throughout a field.

To produce a “universal” model for well correlation, Brazell et al. (2019) developed a deep CNN architecture trained based on five million data points derived from thousands of well-log and experienced interpreter correlations. The data was obtained from various depositional environments and basins within the USA. The authors have implemented a 3D search logic to determine the marker propagation pathway and the optimum correlation. The model does require some

interpreted-top examples to be provided from the specific dataset to account for particular complexity within the geology of a given area. Nevertheless, no need for extensive training data set from the specific field due to the rich dataset used to build the model. The model could provide an accuracy of around 96% on the testing dataset. It is important to note that more interpreted examples might be needed for the training if the model is to be applied to a dataset outside the US with very different regional complex geology. Another potential consideration is incorporating seismic sequence stratigraphy into the workflow, which currently relies only on well-log data. This can be important, especially in benching out strata and faulted reservoirs where the spatial continuation of a given unit might be heterogonous.

## Reservoir characterization

Machine learning has an increasing number of applications in the field of geosciences. Still, we focus here on applications directly related to reservoir characterization in the oil/gas industry. The areas discussed are petrophysical properties prediction from the seismic, core, and well-log data. Other properties such as water saturation, petroleum geochemical parameters, and reservoir geomechanics will be predicted.

### Petrophysical properties prediction

Reservoir characterization plays a critical role in the oil and gas industry, such as developing optimal production and reservoir management strategies. Permeability, which determines the ability and direction of oil flow, is central in reservoir characterization. An accurate permeability determination is essential for material balance calculations, reservoir flow simulation, estimating oil production rate, stimulation strategies, and enhancing oil recovery. However, permeability is very difficult to determine due to its complexity and highly nonlinear nature. Therefore, machine learning techniques are widely used to predict petrophysical parameters such as porosity, permeability, capillary pressure, relative permeability, and bulk density. Table 5 shows a summary of the studies used to predict porosity and permeability.

### Water saturation prediction

Water saturation defines a fraction of pore space occupied by water. A good estimation of water saturation is considered a difficult task in petroleum engineering. In fact, there are very few empirical models that existed to predict water saturation directly from the petrophysical well logs. Nevertheless, water saturation is an essential parameter in petrophysics and reservoir engineering calculations such as material balance

calculations, simulation model optimizations, history matching, and oil and gas reserves estimation. In 1942, Archie was the first to present an equation to determine water saturation in a clean, non-clay reservoir. Several researchers have tried to deconvolute the water distribution in composite formations by formulating empirical correlations that depend on log-derived data, which is not a very precise representation. Hence, no consensus exists among log analysts about which model can be universally used. The most commonly utilized models/correlations are Simandoux (Simandoux 1963), Fertl and Hammack (Fertl and Hammack 1971), and Waxman and Smith models; however, the variables involved in each contain inherent uncertainties and eventually lead to misconstrued results. Determining water saturation in the laboratory is a time-consuming and challenging task. Therefore, AI and ML techniques have widely been used to predict water saturation. Table 6 provides some of the insight on how to predict water saturation using machine learning algorithms. Most of the presented research integrates well-log and core data to predict water saturation.

### Geomechanics

A better estimation of the reservoir rock elastic and failure properties is instrumental to minimizing wellbore instability problems, avoiding differential sticking, improving hole cleaning, improving casing placement, improving hydraulic fracturing operations, minimizing subsidence, and many more (Khamidy et al. 2019). Carrying out mechanical rock tests such as triaxial compression, uniaxial compression, scratch, and impulse hammer is an accurate way to determine these properties (Elkatatny et al. 2019). These tests are usually carried out on the downhole samples retrieved from some depth of interest. In the absence of core samples and well-log data, analytical and empirical models determine rock mechanical properties. In the last two decades, predicting the mechanical rock properties using AI tools was thoroughly investigated. Table 7 lists the summary of the selected work done to relate core mechanical properties with well-log data using ML tools. Most of the work has utilized ANN, ANFIS, SVR, DT, and RF.

### Drilling and completions

Drilling operations for oil and gas reservoirs are usually expensive. Hence, several approaches are utilized to reduce the operational cost, mainly by improving the drilling efficacy and reduce the drilling time. Usually, the drilling performance is improved by selecting proper drilling fluids, improving cementing jobs, maximizing the drilling rate of penetrations, and minimizing the required drilling energy (Bilgesu et al. 1997; Dupriest

**Table 5** Summary of the researches related to the application of AI in predicting permeability and porosity

References	Target parameter	Method techniques	Input parameters	No. of data points	Lithology	Geological formation studied	Accuracy	
							Error	$R^2$
Fatai A Anifowose et al. (2013a, b)	Permeability	Type 2 fuzzy logic, SVM, and ANN	NPHI*, RHOB*, CALI*, GR*, SWT*, and seismic parameters	–	Carbonate	Middle eastern carbonate reservoir	–	0.95
Fatai A Anifowose et al. (2013a, b)	Permeability and Porosity	FN, ANFIS-SC, ANFIS-FCM*, SVM, and ANN	NPHI*, RHOB*, CALI*, GR*, SWT*, and seismic parameters	–	Carbonate	Middle eastern carbonate reservoir	–	0.98
Nooruddin et al. (2014)	Permeability	ANN, SVR, FN, and GRNN	Winland parameter, displacement pressure, porosity, pittman parameter, swanson parameter, purcell integration, and Dastidar parameter	206	Carbonate	Middle eastern carbonate reservoir	AAPE134	0.962
Anifowose et al. (2017)	Permeability and porosity	FN and type 2 fuzzy logic	GR*, NPHI*, RHOB*, SWT*, RT*, MSFL*, NPHI*, and CALI*	500+	Carbonate, dolomite, and sandstone	–	–	0.96
Elkatatny et al. (2018a)	Permeability	ANN, SVM, and ANFIS	RHOB*, NPHI*, LLS*, and LLD*	488	Carbonate	Middle eastern carbonate reservoir	5	0.98
Elkatatny et al. (2018b)	Porosity	ANN, SVM, and ANFIS	RHOB*, NPHI*, and DTC*	700	Carbonate	Middle eastern carbonate reservoir	1.5	0.99
You et al. (2018)	Capillary pressure and relative permeability	PSO-ANN	Porosity, Permeability, pore-throat radius, and sorting coefficient	–	–	Marine-sand reservoir	0.1–8%	–
Bondi et al. (2018)	Bulk density of soil	Linear regression and decision Trees	Soil parameters	500+	–	–	–	–
Ahmadi and Chen (2019)	Permeability and porosity	LSSVM* and ANN	RHOB*, NPHI*, PHIT*, and DTC	1000	Carbonate	South of Iran, Khuzestan state	1%	Above 0.99
Ba alawi et al. (2020)	Capillary Pressure and relative permeability	ANN	Resistivity data	–	–	Middle east	–	Above 0.9
Gowida et al. (2020)	Bulk density of the rock	ANN, ANFIS, SVM, and FN	Drilling parameters	2400	Carbonate	Middle east	1%	0.95

GR\* gamma ray, RHOB\* bulk density, LLD\* deep lateral log, MSFL\* micro spherical focused log, RILD deep induction resistivity log, DT\* compressional wave travel time, TH\* thorium, U\* uranium, K\* potassium, RT\* resistivity log, NPHI\* neutron porosity, SP\* spontaneous potential, CNL\* compensated neutron log, PE\* photoelectric index, RT\* deep resistivity, LSSVM\* least square support vector machine, PHIT\* true porosity, SWT\* water saturation, ANFIS\* adaptive-network-based fuzzy inference system

**Table 6** Summary of the researches related to the application of AI in predicting water saturation

References	Target parameter	Method techniques	Input parameters	No. of data points	Lithology	Geological formation studied	Accuracy	
							Error	$R^2$
Helle and Bhatt (2002)	Sw	Committee neural networks	RT, NPHI, DT, and RHOB	150	–	–	–	0.98
Shokir (2004)	Sw	ANN	GR, SP, RT, NPHI, and RHOB	–	Shale	Egyptian field	–	0.996
Kamalyar et al. (2012)	Sw	ANN	Porosity and K from core, and $h^*$	263	Carbonate	Iran	6.4%	0.985
Mardi et al. (2012)	Sw, $m^*$ , and $n^*$	ANN	RT, NPHI, RHOB, DT, and core porosity	–	Carbonate	Azadegan oilfield, sarvak formation, Iran	–	0.969
Al-Bulushi et al. (2012)	Sw	ANN	RHOB, NPHI, resistivity, and PE	14	Sandstone	Gharif and hardh formations of Oman	2.5	0.91
Kenari and Mashohor (2013)	Sw	ANN, FL, and ANFIS	SP, RT, NPHI, RHOB, PE, and effective PHI	–	Carbonate	Khark oilfields, Iran	0.89–0.95	0.95
Mollajan et al. (2013)	Sw	ANN and SVR	RT, NPHI, RHOB, and DT	–	Carbonate	Sarvak formation, Iran	0.21 MAE*	0.81
Amiri et al. (2014)	Sw	ICA-ANN	GR, RT, NPHI, RHOB, and effective NPHI DP	–	Shaly gas sandstone	–	–	–
Amiri et al. (2015)	Sw	Progressive quasi—Static (PQS) algorithm	Well-log data	–	Tight shaly gas sandstones	–	–	–
Bageri et al. (2015)	Sw	ANN and FL	Resistivity logs and core data	378	Carbonate	Middle eastern carbonate reservoir	–	0.95
Gholanlo et al. (2016)	Sw	RBF ANN	NPHI, DT, RHOB, and core data	–	Carbonate	Sarvak formation, Iran	0.027 MSE*	0.870
Khan et al. (2018)	Sw	ANN and ANFIS	GR, RT, Rxo, NPHI, RHOB, and caliper log	150	Carbonate	Middle eastern carbonate reservoir	5%	0.94
Tariq et al. (2020a)	Sw	FN	GR, RHOB, NPHI, 15FR, LLS, and LLD	150	Carbonate	Middle eastern carbonate reservoir	AAPE=4.3	0.92

$K$  permeability,  $h^*$  cementation factor,  $n^*$  saturation exponent,  $h^*$  height above free water level,  $DT^*$  decision tree forest,  $RBF$ -ANN radial based function artificial neural network,  $MAE$  mean absolute error,  $MSE$  meansquare error

and Koederitz 2005; Hegde and Gray 2018). Therefore, accurate prediction of the drilling fluid parameters, cement strength, and rate of penetrations (ROP) is an essential element for evaluating or improving the drilling performance.

### Drilling performance prediction

Several analytical models were developed to evaluate and optimize the drilling performance; however, most of these models were developed based on weak assumptions,

**Table 7** Summary of the researches related to the application of AI in predicting geomechanical parameters

References	Target parameter	Method techniques	Input parameters	No. of data points	Lithology	Geological formation studied	Accuracy	
							Error	$R^2$
Yilmaz and Yuksek (2008)	Estatic and UCS	ANN and NLR	Effective porosity, point load index, schmidt rebound number, and slake durability index	39	Gypsum	Hafik formation, sivas basin, Turkey	Less than 5%	Above 0.9
Abdullaheem et al. (2009)	Young's modulus and Poisson's ratio	ANN, Fuzzy logic, and FN	DTC and RHOB	77	Carbonate	Saudi Arabian field	–	0.72
Ameen et al. (2009)	Elastic moduli	NLR	DTC, DTS, and $E_{dyn}$	–	Carbonate	Arab D, Ghawar, Saudi Arabian field	–	Above 0.9
Sharma et al. (2010)	UCS	ANN	Bulk density, compression wave velocity, volume of shale and porosity	–	Carbonate	Caspian Sea and South Atlantic Ocean	MAE=1	Above 0.9
Rabbani et al. (2012)	UCS	ANN	RHOB, NPHI, and Sw	5000	Carbonate	South Iran	–	0.9
Najibi et al. (2015)	Estatic and UCS	NLR	DTC, DTS, and RHOB	45	Carbonate	Asmari and sarak limestones, Iran	–	0.96
Tariq et al. (2016)	DTC and DTS	ANN, SVM, and ANFIS	GR, RHOB, and NPHI	5000+	Carbonate	Saudi Arabian field	Less than 5%	Above 0.9
Negara et al. (2017)	UCS	SVR	Elemental spectroscopy, RHOB, and NPHI	–	Carbonate	Middle east	–	Above 0.9
Tariq et al. (2017a)	UCS	ANN, SVM, and ANFIS	GR, RHOB, and NPHI	100	Carbonate	Saudi Arabian field	Less than 5%	Above 0.9
Tariq et al. (2017b)	Friction angle and Cohesion	ANN, ANFIS, and SVM	RHOB, NPHI, DTC, DTS, and $E_{dyn}$	100	Carbonate	Saudi Arabian field	Less than 5%	Above 0.9
Tariq (2018, , Tariq et al.2018)	Poisson's Ratio	FN, ANFIS, and NLR	DTC, DTS, and RHOB	600	Carbonate	Saudi Arabian field	Less than 5%	Above 0.9
Matin et al. (2018)	UCS	RF	Porosity, $I_s^*$ , $V_p$ , and $R_n^*$	–	Carbonate	Iran field	–	0.93
Anemangely et al. (2019)	DTC and DTS	LSSVM	Depth, NPHI, RHOB, PEF, GR, CAL, RT, and $V_p$	–	–	–	0.23	0.954
Abdi et al. (2018)	UCS	ANN	Porosity, density, $V_p$ , and water absorption	196	Sandstone, conglomerate, limestone, and marl	Iran	–	Above 0.9
Tariq et al. (2019)	Poisson's ratio	ANN	DTC, DTS, and RHOB	600	Carbonate	Saudi Arabian field	Less than 5%	0.96

$I_s$  point load index,  $R_n$  Schmidt hammer rebound number

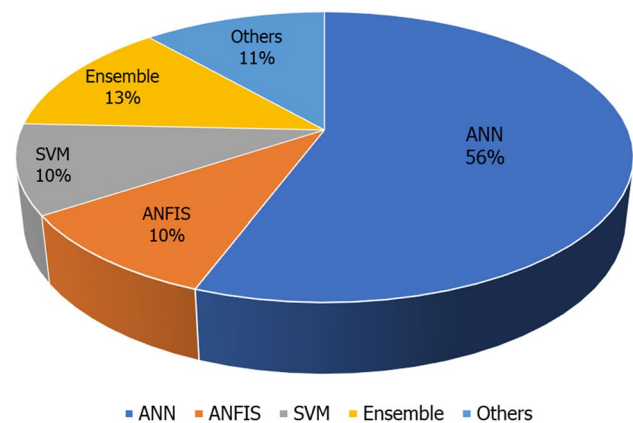
reducing their reliability (Aadnoy et al. 2010; Reiber et al. 1999). Besides, numerical approaches such as finite element methods were also utilized for evaluating the drilling performance by estimating the ROP (Yang et al. 2008). Usually, the depth is divided into individual sections, mainly based on rock lithology. Thereafter, the drilling performance is estimated based on the affected forces in each section (Bourgoyne and Young 1999; Murray and Cunningham 1955). Thus, the numerical approaches improved the prediction performance, but the main issue is the computational speed, limiting the applications of these numerical approaches, especially for real-time operations.

In the last decade, artificial intelligence (AI) techniques present practical tools for prediction purposes, and therefore, have been widely applied in the oil and gas industry (Barbosa et al. 2019; Rolon et al. 2009; Sun and Ertekin 2020; Van Si and Chon 2018). ANN is the most used AI technique among all AI methods because an empirical correlation can be extracted from the optimized ANN model. Hence, numerous ANN models were developed for real-time applications, such as estimating the RPO and the drilling performance (Ahmed et al. 2019; Arabjamaloei and Shadizadeh 2011).

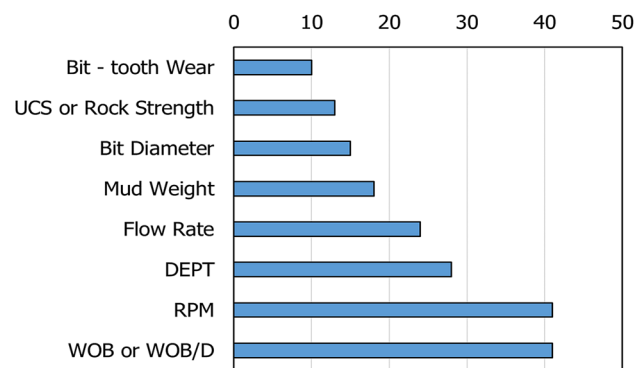
Gidh et al. (2012) improved the drilling performance by predicting and managing the bit wear utilizing an artificial neural network technique. A new ANN-based system was developed to predict the bit performance at different ROP values. They mentioned that the developed approach could be used to determine the optimum range for the surface drilling parameters (such as revolutions per minute (RPM), torque, and WOB) to extend the lifetime of the drilling bit. Their developed approach was tested in several field operations, and successful results were reported.

Evangelatos and Payne (2016) presented an advanced model to describe the motion and dynamics of the bottom-hole assembly (BHA). The developed BHA model was coupled with neural network analysis to estimate the drilling performance for different conditions of WOB and RPM. They reported that the coupled model could consider the acting forces on the BHA and thereby provide very accurate predictions for the ROP profile at the wide range of BHA conditions.

Barbosa et al. (2019) presented extensive reviews on ROP modeling and prediction using machine learning techniques. They classified the ROP models into conventional (physics-based) models, statistical (regression) models, and machine learning (data-driven) models. Based on their reviews, machine learning techniques can outperform all ROP models and provide very reasonable ROP predictions. However, the reliability of ROP prediction depends mostly on two factors; the type of AI method and the inputs used for ROP predictions. However, they concluded that there is a lack of field implementations of AI techniques in the oil and gas industry. They attributed these limited field applications



**Fig. 3** The common machine learning techniques used for ROP predictions, considering 53 ROP studies (Barbosa et al. 2019)



**Fig. 4** The common inputs used for predicting the ROP based on machine learning techniques, considering 53 ROP works (Barbosa et al. 2019)

to the difficulties of selecting the input parameters and the suitable AI models. Figures 3 and 4 show the most common AI techniques and input parameters used to predict the ROP, respectively. ANN is the most AI technique for developing ROP models among the reviewed works, while the most common inputs are the weight on bit (WOB) and RPM. ANN tends to be the most common machine learning method when it comes to dealing with large data sets. The main reason behind that is the availability and the relatively simple structure and layout of ANN models when compared with other machine learning models. Besides, they mentioned that humans are usually resistant to change, which leads to limited field applications. Finally, they recommended that downhole parameters such as nozzle size, drill bit wear, and rock strength should be considered inputs for predicting the ROP profiles.

Ahmed et al. (2019) presented a comparative study of predicting ROP using several intelligence techniques. ROP was predicted for two wells using an extreme learning machine, ANN, and SVR techniques. They selected the input

parameters for the ROP models based on the specific energy concept. The ROP was predicted for more than 8800 data points based on the RPM, WOB, torque, depth, mud weight, flow rate, nozzle sizes, and standpipe pressure (SPP). They reported that all ROP models showed acceptable prediction performance with a correlation coefficient higher than 0.70 for the testing data. However, among all tested techniques, support vector regression showed the best ROP estimation with a correlation coefficient of 0.94.

Mehrad et al. (2020) used a machine learning approach to develop a rigorous ROP model for vertical wells. They used different parameters to determine the ROP, including logging, drilling, and geomechanical parameters. They found that the best ROP prediction can be obtained by using the uniaxial compressive strength (UCS), mudflow rate, weight on bit (WOB), Depth, mud density (MD), and revolutions per minute (RPM) as input parameters. After that, they combined the least-squares support vector machines (LSSVM) with different optimization algorithms to estimate the ROP profile. The examined optimization algorithms are genetic algorithms (GA), particle swarm optimization (PSO), and cuckoo optimization algorithm (COA). LSSVM-GA, LSSVM-PSO, and LSSVM-COA hybrid algorithms were used to predict the ROP for two vertical wells, and more than 2000 data points were used to train and tests the hybrid models. LSSVM-COA showed the best prediction performance for training and testing wells among all tested algorithms, and an R-square of around 0.802 was achieved.

Artificial intelligence showed an effective approach for estimating the drilling performance, and accurate profiles of ROP can be predicted. However, it is noticeable that there is a lack of implementation of those techniques for real-time operations, especially for gas wells. Also, most of the available ANN-based models were developed to predict the ROP for a certain section, usually for the reservoir section. No attempt was reported for predicting the full profile of ROP using the ANN technique. Predicting the complete profile of ROP in real time can significantly improve the drilling performance and reduce the operational time and cost.

Furthermore, the coupling of different drilling efficiency indicators can help in improving the drilling operations by considering more than one parameter. For example, the ROP models can be coupled with the MSE concept to determine the best drilling conditions in drilling time (ROP) and required drilling energy (MSE). Hassan et al. (2018) coupled the torque modeling with the mechanical specific energy (MSE) to optimize the drilling performance. First, artificial intelligent techniques were used to predict the torque and ROP profiles for around 18000 ft. Then, the MSE was calculated for the whole drilling section using the surface drilling parameters. After that, the MSE was coupled with the torque and ROP profiles to identify the optimum drilling conditions that will result in maximizing the ROP and minimizing the

required drilling energy (MSE). They mentioned that the developed approach would enable the drilling engineers to evaluate and optimize the drilling performance in real-time applications; hence, the surface drilling parameters can be controlled to maintain the drilling operations within the optimum conditions.

Besides, AI techniques were used to estimate several drilling problems, such as loss of circulation, one of the most common drilling problems that can increase the overall drilling cost by around 25–40%. Solomon et al. (2017) developed a new ANN model to estimate the loss circulation zones. The developed model can also recommend the suitable sizes of loss circulation materials based on the characteristics of the depleted zones. They used 30 case studies to train and validate the developed ANN model. They mentioned that the ANN model showed a very acceptable prediction performance, and a coefficient of determination of 0.8 was obtained. Besides, they compared the reliability of the developed model with different fracture predictive models, and they concluded that the developed ANN model could reduce the estimation error from around 26% to less than 16%.

Manshad et al. (2017) used an SVM and radial basis function to assess the loss of circulation problems for 30 oil wells. They reported that SVM showed high performance in predicting the amount of loss circulation material required to overcome the thief zones. A coefficient of determination of 0.8 was obtained between the predicted results and actual field data. In comparison, the radial basis function was able to estimate the mitigation of loss of circulation problems with an accuracy of 78.3%.

Al-Hameedi et al. (2018) estimated the volume of lost circulation materials for 500 wells using the machine learning technique. They predicted the volume of fluid losses based on the profiles of mud weight, bit nozzle sizes, ROP, equivalent circulation density (ECD), plastic viscosity (PV), and WOB. They reported that the machine learning models were able to predict the volume of fluid losses with very acceptable error for different types of mud loss, including partial, seepage, severe, and total mud losses.

Alkinani et al. (2020) used an ANN technique to predict the volume of drilling fluids losses during drilling fractured zones. They developed and validated the ANN model using 1500 wells. Also, the lost circulation volume was determined based on the profiles of mudflow rate, yield point (YP), PV, ECD, bit nozzle sizes, RPM, and WOB. They reported that the ANN model was able to predict the loss of circulation with a coefficient of determination higher than 0.92.

Abbas et al. (2019) applied SVM and ANN techniques to estimate the severity of loss of circulation while drilling. They used 1120 case studies from 385 wells to train and validate the new AI models for different types of mud losses such as seepage, partial, severe, and total fluids losses. They used the rock lithology, mud properties, and drilling surface



parameters to predict the severity of loss of circulation. They reported that the developed ANN model was able to estimate the fluids loss with a correlation coefficient higher than 0.82. While the SVM model showed better prediction performance compared to the ANN model, a correlation coefficient higher than 0.91 was obtained.

Overall, different AI techniques were utilized to estimate the loss of circulation problems. ANN and SVM methods are the common AI tools that are used for this purpose. The very practical performance was reported for predicting the loss circulation based on the mud properties, rock lithology, and drilling parameters. However, the application of these models in real-time operation might be restricted due to the huge drilling data, leading to misleading results or delaying the model prediction. Therefore, proper data cleaning could be required to improve the data quality and reduce the data size for problems in real-time applications (Elkatatny et al. 2016).

## Drilling fluids

Drilling is one of the most critical tasks, with challenges including lost circulation, clogged pipes, wellbore instability, and kicks occurring regularly. Drilling fluid, sometimes known as the "blood of the drill," is a direct or indirect remedy to the challenges stated above during the drilling process. It helps to keep the wellbore clean and retain the wellbore's integrity. For instance, high mud weight controls the high wellbore pressures and prevents kicks. On the other hand, high mud weight has a tendency to frack the formation. Similarly, low mud weight prevents fractures but can cause kick or blowout. Further drilling fluids prevent the pipe from sticking during drilling by building thin filter cake on the wellbore wall as well as by removing drilling cuttings out the wellbore. The drilling fluid works as an architect for the wellbore. The operation's success or failure is largely determined by the drilling fluid's performance and compatibility (Agwu et al. 2018). Many drilling issues can be avoided by using the proper drilling fluids. Drilling fluids are always chosen based on data analysis and expertise gained from previously drilled wells in the area. Each well design includes a drilling fluid program that specifies drilling fluid, additives, rheology, density, filtration, and other drilling fluid parameters. Combating wellbore difficulties involves comprehensive analysis and decision-making to build the drilling fluid to satisfy specific needs that suit distinct formation features.

The majority of drilling fluid design is done in the laboratory through trial and error. Hence, a system that can use existing data and provide a deeper knowledge of drilling fluid is required. Machine learning models are created using the parameters of drilling fluids and the downhole circumstances. These models aid in forecasting changes in drilling

fluid parameters and recommend the optimum course of action. Rheological models express a mathematical relationship between the shear rate and the shear stress to describe the fluid flow behavior. This relationship is complicated in the case of drilling fluids. However, no single rheological model can accurately fit all drilling fluids' shear stress-shear rate data across all shear rate ranges. Instead, a plethora of mathematical models with varying degrees of relevance has been utilized. These mathematical models do not precisely capture the behavior of non-Newtonian fluids. For instance, the Bingham plastic model does not describe the drilling fluid flow behavior at a low shear rate. Further, it overestimates the yield point of the drilling fluid. The power-law model does not account for the yield point of drilling fluids. There are challenges in performing hydraulic calculations due to many rheological parameters involved in the case of the Herschel-Bulkley model (Huang et al. 2020).

Regression approaches are utilized to predict rheological proficiencies such as an ANN. For greater accuracy, the ANN model can be trained continuously with more data sets. It gives a more comprehensive view of how to comprehend the drilling performance. For example, if there is a reduction in pump pressure during the drilling operation, which happens for several reasons, including thinning effect on the drilling fluid, quick transport of the cuttings to the surface, reservoir fluid influx in the wellbore, and lost circulation, etc. Here AI interlinks different parameters, improves the decision-making process, and brings back the engineers on the right track within a short time.

Tables 8 and 9 outlines several studies of artificial intelligence in drilling fluids. The tables summarize the drilling fluids properties investigated and the AI technique used. They also show the input and output parameters and accuracy of a performance evaluation using correlation coefficient ( $R^2$ ), mean square error (MSE), average absolute percent relative error (AAPE), etc.

## Oil well cementing

The main objective of the oil well cement is to prevent the movement of fluid between the geological formation and behind the casing string (Murtaza et al. 2020; Tariq et al. 2020b). A slurry of cement is pumped down into the annulus between the casing and the geological formation. A cement slurry is a mixture of various additives such as strength enhancers, friction reducers, fluid loss agents, and expanding agents, etc. In the field of oil well cementing, AI is mostly used for the prediction of cement strength development and rheological properties. Table 10 provides some of the applications of AI in the field of oil well cement. Compressive strength development is one of the most critical parameters which significantly affects the drilling operation. Accurate prediction of compressive strength development can save

**Table 8** Summary of the application of AI in predicting properties of drilling fluids

References	Study conducted	Type of AI method	Input parameters	Output parameters	Accuracy	
					AAPE	R <sup>2</sup>
Jeirani and Mohebbi (2006)	Filtrate and mud cake permeability of WBDF	Backpropagation network (ANN)	Filtration Time; pressure drop; water and NaCl percentages	Fluid loss volume; mud cake permeability	–	Filtrate: 0.9815; mud cake permeability: 0.9433
Miri et al. (2007)	Prediction of differential pipe sticking during drilling	Multilayer perceptron (MLP); Radial basis functions (RBF)	Mud filtrate viscosity; plastic viscosity (PV); yield point (YP); 10 s and 10 min gel strength; differential pressure; hole depth; API filtration; solid percent;	Stuck pipe index	–	MSE for MLP = 0.00547 MSE for RBF = 0.00419
Deosarkar and Sathe (2012)	Effective viscosity of magnetic slurries	Back-propagation (BP) ANN	Solids concentrations; viscosity	Effective viscosity	–	0.9768
Rooki et al. (2012)	Particle settling velocity in newtonian and non-Newtonian fluids	ANN	Density and diameter of particle; fluid density; fluid consistency; flow behavior index; gravity	Terminal velocity	–	0.986
Razi et al. (2013)	Rheological properties of WBDF	ANN	Temperature; concentration of additives	PV; YP	–	0.965
Maghrabi et al. (2014)	Behavior of shale erosion in aqueous drilling fluids	ANN	Shale mineralogy and fluid composition	% recovery	–	0.85–0.9
Rooki et al. (2014)	Wellbore cleaning efficiency of foam fluid	ANN	Pressure, temperature, pipe rotation, foam quality, foam velocity, and annular eccentricity are all factors to consider	Cuttings concentration	5.93%	0.914
Shadravan et al. (2017)	Drilling fluid design	Backpropagation (ANN)	Drilling fluid density; ingredient A content; temperature	Shear Rates: 600, 300, 200, 100, 6 & 3 RPM	–	–
Hoang (2016)	Drilling fluid viscosity estimation	Fuzzy logic, feedforward artificial neural network; Feedback artificial neural network	Density; shear stress	Mud viscosity	MAPE 8.12%	0.98
Bispo et al. (2017)	Estimating apparent viscosity of WBDFs	Feedforward multilayer perceptron (MLP)	Temperature; concentrations of barite, bentonite and xanthan gum	Apparent viscosity (AV)	–	0.9486
Chhantyal et al. (2017)	Drilling fluid flow rate prediction	ANN	Ultrasonic level measurements	Mud flow rate	–	–
Hmood Mnati and Abdul Hadi (2018)	Prediction of penetration rate and cost	ANN	Weight on bit; rotation per minute; hydraulic (HIS); travel transit time(DT)	ROP and Cost	–	0.9689

Table 8 (continued)

References	Study conducted	Type of AI method	Input parameters	Output parameters	Accuracy	
					AAPE	R <sup>2</sup>
Abdelgawad et al. (2018)	Real-time predication of rheological properties of spud drilling fluids	ANN	Mud density, Marsh funnel viscosity and solid percent	PV, YP, AV, and flow behavior index	<5%	> 95%
Tomiwa et al. (2019)	Application of artificial neural network in predicting mud Rheological properties in water based mud using solanum tuberosum formulated biopolymer	ANN	Quantities of Bentonite and solanum tuberosum biopolymer	PV, YP & AV		R <sup>2</sup> :0.99373 & MSE: 1.25
Oguntade et al. (2020)	Prediction of rheology and filtration properties of water based drilling fluids	ANN	Quantities of water, bentonite and biopolymer	YP, PV & AV		R <sup>2</sup> : 0.92234 and MSE: 9.2494
Samnejad et al. (2020)		ANN	Drilling fluid composition, shear stresses at ambient conditions	Drilling fluid rheology		MAPE: 8.74%, R <sup>2</sup> : 0.75
Kumar et al. (2020)	Non-Newtonian fluid pressure drop prediction	Bayesian Neural Network (BNN); random forest (RF); ANN; support vector machines (SVM)	Ratio of Diameter; Flow Behavior Index; Eccentricity of Annulus; Consistency Index; Yield Stress;	Annulus pressure drop	–	ANN: 0.904; RF: 0.986; BNN: 0.989; SVM: 0.83
Osarogiabon et al. (2020)	Kick detection during drilling	Long short-term memory recurrent neural network	Flow rate	Influx size	–	0.98
Osarogiabon et al. (2020)	Kick detection and influx size estimation	Long Short-Term Memory (LSTM), a Recurrent Deep Neural Network (RNN)	d-exponent and SPP	Kick size and flow rate	–	0.97

WBDF\*: water based drilling fluids, OBDF\*: oil-based drilling fluids,

**Table 9** Summary of the application of AI in predicting density of drilling fluids

References	Study conducted	Type of AI method	Input parameters	Output parameters	Accuracy	
					AAPE	R <sup>2</sup>
Osman and Aggour (2003)	The density of WBDF and OBDF	Backpropagation network (ANN)	Drilling fluid density; temperature; pressure; water and NaCl wt.%; pressure drop	Density	0.367	0.9998
Adesina et al. (2015)	Prediction of the density of oil-based drilling fluids in HPHT wells	ANN	Downhole pressure; downhole temperature; initial drilling fluid density	Downhole drilling fluid density	12.70%	0.999
Zhou et al. (2016)	HTHP drilling fluid density of WBDF & OBDF	BP neural network-Particle Swarm Optimization (PSO)	Initial mud density; pressure; difference temperature difference; water volume fraction; oil volume fraction	Density	4.5	0.9998
Ahmadi et al. (2018)	Determination of drilling fluid density at wellbores conditions	particle swarm optimization (PSO); artificial neural network (ANN)	Initial mud density, Pressure (Psi) and Temperature °C	Drilling fluid density	–	0.9964
Rahmati and Tatar (2019)	Prediction of mud density	Radial Basis Function (RBF) network	Initial density; pressure and temperature of OBM and WBM; Colloidal Gas Apron (CGA); synthetic drilling fluids	Density	0.0752	0.999
Alkinani et al. (2020)	Estimation of equivalent circulation density (ECD)	Bayesian regularization (BR) algorithm	Mud density; YP; PV; RPM; Flow rate, Weight on bit; Nozzles total flow area	Equivalent circulation density (ECD)	–	0.982

**Table 10** Summary of the application of AI in Cementing

Reference	Study conducted	Type of AI method	Input parameters	Output parameters	Accuracy	
					Error	R <sup>2</sup>
Kawalramani and Gupta (2006)	Compressive strength prediction using ultrasonic pulse velocity	Multiple regression (MLR) and ANN	weight and UPV	Compressive strength	–	–
Shahriar and Nehdi (2011)	Rheological Properties of Oil Well Cements Slurries	Back-propagation neural network	Shear rate; dosage of admixtures; test temperature	Shear stress	3.43%	–
Jui-Sheng et al. (2011)	Prediction accuracy of compressive strength	ANN	Cement; FA; BFS; water; superplasticizer; coarse and fine aggregate; curing age	compressive strength	9.64	0.9091
Yadollahi et al. (2015), Trtnik et al. (2009)	Prediction of compressive strength of geopolymer composites	ANN	Silica modulus; Na <sub>2</sub> O content; water binder ratios; curing time	compressive strength of geopolymer	8.75	0.958
Safuiddin et al. (2016)	Compressive Strength for Self-Consolidating High-Strength Concrete	ANN	Aggregates; cement; palm oil as; water; water reducer	Compressive strength	4.39	0.948
Khashman and Akpinar (2017)	Prediction of nondestructive concrete compressive strength	ANN	Water; cement; fly ash; slag	Compressive strength	–	0.86
Naderpour et al. (2018)	Compressive strength prediction of environmentally friendly concrete	ANN	Water-cement coefficient; water absorption; percentage of fine aggregate; recycled coarse aggregate; natural coarse aggregate; water-total material coefficient	Compressive strength	–	0.8926
Bagheri et al. (2019)	Compressive strength of boron based geopolymers	ANN; genetic programming (GP)	% fly ash; % GGBS, ratio of boron; silicon; sodium ions	Compressive strength	RMSE=0.07	0.9
Dao et al. (2019)	Prediction of Compressive Strength of Geopolymer Concrete	ANFIS and ANN	fly ash; sodium hydroxide; sodium silicate solution; water	Compressive strength	ANFIS RMSE=2.265 and ANN RMSE=2.423	ANFIS: 0.879 ANN: 0.851
Tariq et al. (2020c)	Prediction of rheological parameters of Class G with nanoclay as a strength enhancer	ANN	Nanoclay Concentration, curing temperature, and Shear rate	PV*, YP*, AV*, Shear Stress	Less than 5%	More than 0.9

PV\* plastic viscosity, YP\* yield point, AV\* apparent viscosity

millions of dollars by reducing the wait on the cement after cementing operation. In cement design, various additives are mixed with the cement, and each additive impacts the performance of cement slurry. The AI models can predict these impacts on cement performance without conducting detailed laboratory investigations.

## Reservoir and production

### PVT properties curves prediction

Pressure–Volume–Temperature (PVT) crude oil properties are considered essential in petroleum engineering for reservoir and production calculations. Determination of these properties in the laboratory is the most accurate and expensive way to obtain representative values (Tariq et al. 2021). In the absence of such facilities, other approaches such as analytical solutions and empirical correlations are used. Some of these correlations can be seen in the works of Al-Shammasi (2001) and Osman et al. (2001), who used ANN to predict the bubble point ( $P_b$ ). Al-Marhoun and Osman (2002) used ANN to predict  $P_b$  for crude oils in Saudi Arabia. Their model was developed using 283 data sets with an AAPE of 5.8% and  $R^2$  of 0.996. SVM and ANFIS were also used to predict the  $P_b$  by several authors (El-Sebakhy 2009; Baarimah et al. 2014; Ganji-Azad et al. 2014; Shojaei et al. 2014). Numbere et al. (2013) used ANN to predict the  $P_b$  for the Niger Delta crude oil. Their model utilized 1248 data sets with an AAPE of 17.17%, and  $R^2$  of 0.96. From the literature survey, it was revealed that most of the previous work utilizing AI and Machine Learning (ML) to predict PVT properties used black-box approach (Ahmadi et al. 2015; Cheshmeh Sefidi and Ajorkaran 2019; Hemmati-Sarapardeh et al. 2014; Ramirez et al. 2017; Seyyedattar et al. 2020; Shokrollahi et al. 2015).

### Production in the reservoir

Predicting Well performance prediction is one of the key parameters in developing and managing oil and gas fields. Several approaches are used to determine productivity, such as conducting deliverability tests or using mathematical models. Deliverability tests are usually time-consuming and costly operations, while the available correlations showed considerable deviations between the actual and predicted values. Data-driven models present a promising approach for estimating production based on reservoir properties and well configurations. The common AI techniques are an artificial neural network (ANN), SVM, and fuzzy logic system. ANN is the most applicable technique among all AI methods that showed a very effective performance in several applications. Several ANN models were

developed to evaluate the hydrocarbon productivity for several well types and operations (Alarifi et al. 2015; Hassan et al. 2017; Sun and Ertekin 2020). The performance of several enhanced oil recovery (EOR) treatments was also evaluated using ANN models such as CO<sub>2</sub> injection and miscible gas flooding (Le Van and Chon 2017a; Van and Chon 2017b).

Alarifi et al. (2015) applied three AI tools to estimate the productivity index of horizontal wells producing from oil reservoirs. FL, ANN, and FN were used to determine the well production rate for more than 100 wells. They reported that the developed AI models provided very good predictions and outperformed the industry's well-known correlations. Chen et al. (2015) and Feifei et al. (2015) determined the productivity index for horizontal wells using AFL, ANN, and FN. They mentioned that the developed models investigate the influences of reservoir parameters (such as reservoir size, thickness, and reservoir permeability) on well performance. Also, they reported that the AI models showed very acceptable predictions and reduced the estimation error compared to the available correlations. Buhulaigah et al. (2017) estimated the productivity of multilateral wells utilizing artificial neural networks (ANN). They presented an ANN model to determine the oil production rate for multilateral wells based on the reservoir and well parameters. They compared the developed model with analytical models and correlations. The ANN model showed good performance; strong matching between the actual and predicted flow rates was achieved, with an overall error of 7.9%.

Hassan et al. (2017, 2020) applied different artificial intelligence techniques to predict the well performance for fishbone well types. FL, radial basis network, and ANN were used to determine the well production rate from more than 250 cases of different reservoir properties and wellbore conditions. The developed models were able to estimate the well productivity of fishbone with an absolute error of 7.23%. Furthermore, a new correlation was presented utilizing the optimized ANN model. The developed correlation was validated using actual field data with an estimation error of 6.92%. They mentioned that the suggested correlation could be inserted into the commercial production software, which will reduce the deviations between the simulated results and the actual field measurements.

Ariturk (2019) used artificial intelligence to optimize the flow rates for injection and production wells operated in geothermal Fields. The flow rates were predicted based on the wellhead pressure, wellhead temperature, and valve positions. The injection and production rates were forecasted for around 500 days. It was mentioned that AI models could provide a very acceptable prediction for the geothermal reservoirs since the models were developed based on the field data/measurements. Also, it was concluded that AI presents

a reliable approach that can minimize the complexity and uncertainties associated with geothermal reservoirs.

Sagheer and Kotb (2019) have used the LSTM to predict the productivity of unconventional shale reservoirs. They have found that LSTM models have resulted in comparable results with the physics-driven reservoir simulation. Aulia et al. (2014) have used RF to predict several wells' bottom-hole pressure (BHP). They have built the reservoir simulation model by tuning several input parameters. Through simulation, they have generated the BHP values and used them for the model training.

### Reservoir simulation and field development optimization

Reservoir simulation plays an essential role in modern oil and gas exploration and production. ML methods have been used to accelerate oil reservoir simulations and achieve higher accuracy as well. Navrátil et al. (2019) developed a model using deep learning methods to accelerate the simulations of oil reservoirs by three orders of magnitude compared to industry-strength physics-based partial differential equations (PDE) solvers.

Mohaghegh (2011) discussed the AI-based reservoir model which can be developed using pattern recognition capabilities of AI to build relationships between fluid production, reservoir characteristics and operational constraints. Masoudi et al. (2020) showed how AI and ML are used to build a purely data-driven reservoir simulation model that successfully history match all the dynamic variables for all the wells in a field and used to forecast production. They tested the model with a highly complex mature field with large number of wells and years of production. They found out that time, efforts and resources required for the development of the dynamic reservoir simulation models using AI and ML is considerably less than time and resources required using commercial numerical simulators.

ML techniques are being implemented in many areas related to field development optimization. ML can be used to predict production and the potential field productivity which mainly done by conducting history matching models and using them to forecast. Alarifi and Miskimins (2021) developed and new approach using ANN to predict the ultimate recovery of several unconventional oil and gas wells using historical production data along with completion data. They developed and tested using actual production and completion data from 989 multistage hydraulically fractured horizontal wells from four different formations. The models developed can be used to optimize future field development plan by optimizing the well completion and stimulation procedures. Using ANN to forecast the production of several wells using limited production history can potentially help identify the

expected productivity of new wells and therefore optimize the field development.

Khazaeni and Mohaghegh (2011) developed production data analysis method with AI techniques using production history data to build a field-wide performance prediction model. In their work, production history is paired with field geological information to build datasets containing the spatiotemporal dependencies among different wells. They formed intelligent time-successive production-modeling (ITSPM) system using data from 165 wells. Input data includes data from the well itself and offset wells' static data. Dynamic data includes ultimate drainage area and initial production rate for offset wells.

He et al. (2021) developed a methodology to optimize the field development plans (FDPs). This includes optimizing well counts, well locations and the drilling sequence. They used deep reinforcement learning method (DRL) in which the AI model would provide an optimized FDPs. They showed that starting from no reservoir engineering knowledge, the AI model can learn basic reservoir engineering principles, such as placing optimized well locations with high porosity and permeability, choosing a reasonable number of wells and maintaining good well spacing. They also showed an example of how the resulted AI model has been used to obtain FDPs for a real field that is better than the one initially designed by human engineers.

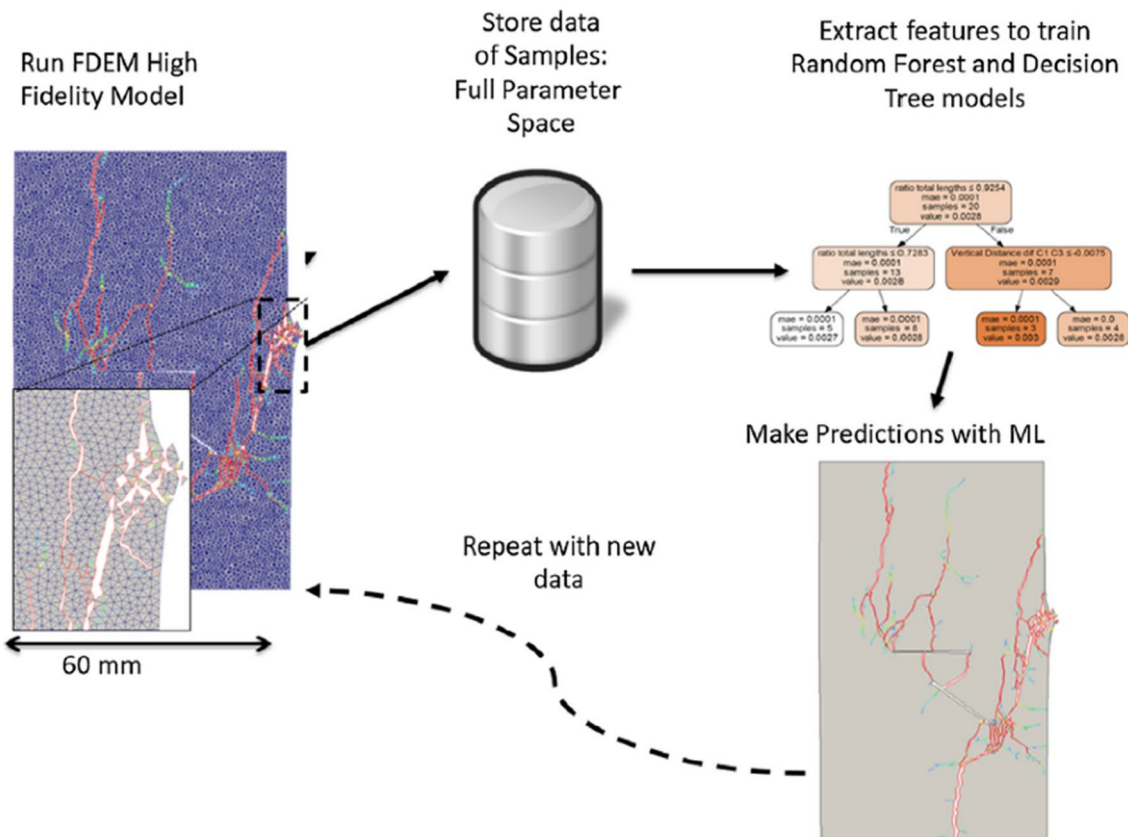
### Stimulation

In the past two decades, the development of unconventional formations was the focus of oil and gas industry operations and scientific research. The invention of horizontal drilling combined with multistage hydraulic fracturing (MSF) made it possible to produce economically from these oil and gas reservoirs. Nevertheless, the understanding of fracture propagation and production thereafter from such complex systems is still in its infancy. Millions of stages were performed in unconventional formations generating a tremendous amount of data that cannot be handled conventionally. Advancement of AI methods made it possible to utilize these data to understand the formation response to stimulation and optimize the MSF completion design. Moreover, AI methods were used to improve the computational efficiency of such complex models that are frequently used in the simulation of fracture propagation or reservoir production. Running these intensive models could take long time which makes it a challenge to optimize the design based on simulations. This section intends to review the utilization of AI algorithms to tackle unconventional formations from stimulation and production perspectives. Also, fracture propagation and conductivity estimation, in general, were reviewed.

Fracture propagation models were developed to understand the growth mechanism of a hydraulic fracture in a complex system containing different minerals and natural fractures. Early studies of fracture propagating were based on simple two-dimensional (2D) (i.e., PKN, KGD) or pseudo-three-dimensional planar models. These models usually oversimplify the problem and do not provide good accuracy if natural fractures dominate the behavior of hydraulic fracture propagation which usually occurs in unconventional formations. Advanced computational methods such as displacement discontinuity method (DDM), discrete element method (DEM), finite-discrete element method (FDEM), and extended finite element method (XFEM) are usually applied (Lecampion et al. 2018). These methods are computationally expensive especially if high accuracy is required in a heterogeneous system.

Accurate fracture propagation modeling even at a small scale can be computationally intensive. Zapico et al. (2008) ANN to fine-tune a finite element model (FEM) to predict experimental fracture propagation outcomes. Nevertheless, the methodology is computationally expensive as FEM should be utilized for the estimations. Moore et al. (2018) built an efficient machine learning model from a

physics-based finite-discrete element model (FDEM) to predict the fracture growth in brittle materials containing pre-existing fractures. Modeling fracture propagation at the microscale is computationally expensive while running that model at field-scale is prohibitively expensive. Hence, a machine learning model was trained to reduce the computational cost between 2 and 5 orders of magnitudes. Around 200 data set of 2D simulations were obtained which were performed on  $2\text{ m} \times 3\text{ m}$  domain containing 20 random fractures each has 30 cm length. The features of two neighbor fractures were studied and used as input to the model. These features were the length of the fractures, orientation, the distance between two fractures, and the minimum distance to the domain boundary. Labels were given to the features indicating if they coalesce and the time for that to take place. Figure 5 shows the methodology implemented where an FDEM model was first performed to produce the outcomes. Features are then extracted to train the AI algorithms which were used to provide similar outcomes to the FDEM. Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) were implemented by the study. From the 20 fractures, 190 pairs were generated containing a total of 5200 data points. The first outcome of their work was to



**Fig. 5** Flow chart showing the methodology implemented to train AI-based algorithms to predict fracture propagation in a domain contains random natural fractures (Moore et al. 2018)



classify if two fracture pairs will coalesce and then a regression model was implemented to estimate the time for that to happen (see the lower right part of the figure). The success of the model was based on the ability to track the fracture path until the domain splits. Also, the time of rupture (domain split) was used as an indication of the model accuracy. The correlation coefficient for ANN as compared to the FDEM simulations was 0.68 while DT and RF gave lower accuracy. One of the challenges (in terms of classification) was the class imbalance (coalesce) as only 5–9 pairs out of possible 190 pairs coalesce. The issue was partially resolved by giving more weights to the imbalanced class. Five orders of magnitudes in simulation time were reduced by using AI methods. Such a methodology is feasible to utilize at field-scale to predict fracture propagation in a formation containing micro and macro-scale natural features.

Fluid flow through a fracture network is challenging to simulate because of the structural complexity. Srinivasan et al. (2018) built a machine learning tool to predict the solute flow through a fractured network. A discrete fracture network (DFN) methodology was used to simulate fluid flow in fracture networks. Solute flow through the fracture network is usually taking the shortest path. Graph theory was used to reduce the number of fractures to those that only contribute to flow. Then, SVM and RF were used to identify the backbone of the fracture network that contributes to flow. This significantly reduced the computational power when simulating flow using the DNF model. The trained model could capture the early solute breakthrough precisely; nevertheless, it was not as useful in predicting late time flow.

Proppant distribution in a hydraulic fracture is crucial information as it could be used to optimize MSF design. Maity et al. (2019) identified proppant particles from cored samples based on imaging processes supported by machine learning classification tools. The goal was to understand the proppant distribution after an MSF job. This helps identify the location of the new infill wells to be drilled and the completion spacing as proppant distribution can tell the length of popped fracture and which clusters were propped. Images were taken for the particles obtained from a 600 ft cored interval using a dedicated slanted well to obtain these cores. Training ANN classification, the particles were divided into proppant, calcite, and others. The following attributes of particles were used as input: hue, roundness, size, darkness, roughness, translucence, and entropy. *K*-fold cross-validation was used for hidden layer size optimization for ANN. It was benchmarked against other classifiers such as SVM. It was concluded that the proppant is limited within 30 ft vertical distance in the studied formation. It was validated against field data using other classification techniques.

AI is also an active area in hydraulic fracture design optimization such as the number of horizontal wells, number of stages, volume of proppant and fluids, type of chemical

additives, and sweet spot identification (Awoleke and Lane 2011; Lolon et al. 2016). Most of the AI developed models ignore important geological and reservoir properties such as porosity, permeability, saturation, and pressure. These data are challenging to obtain especially along the horizontal sections of the wellbore. Some researchers replaced these data with the location of the well (i.e., coordinates) as the mentioned properties are spatially changing (Mishra et al. 2015; Wang and Chen 2019). Wang and Chen (2019) trained machine learning algorithms (RF, SVM, ANN, and AdaBoost) on 3160 horizontal well data of Montney unconventional formation to predict the first-year production and optimize the fracture design. Features such as proppant mass, well location, lateral length, fluids treatment size and type, completion type, number of stages were used for training. Recursive feature elimination with cross-validation (RFECV) was used to find the most significant features where RF was used for prediction. Then, algorithms were trained based on the most important features to predict the production rate from a fractured well. Using RFECV showed that the most important parameter in enhancing production is the mass of proppant pumped for the case of Montney formation and the location of the well. It was found that using more than the four features (proppant mass, latitude, longitude, and TVD) will not improve the correlation coefficient. It was also observed that the RF results in the best performance in terms of prediction accuracy. One drawback of the trained model is its lack of reservoir properties such as permeability, porosity, and pressure.

Optimization of hydraulic fracture stages using gradient-free (i.e., AI) methods has been applied by many researchers (Iino et al. 2020; Yu and Sepehrnoori 2013). The objective function that is usually optimized is the net present value (NPV) or cumulative production. Features such as fracture half-length, spacing, porosity, permeability, the distance between laterals, and fracture conductivity were used for the optimization. Different AI algorithms were tried such as covariance matrix adaptation evolution strategy (CMA-ES), simultaneous perturbation stochastic approximation (SPSA), genetic algorithm (GA), and non-dominated sorting genetic algorithm (NSGA-II). Rahmanifard and Plaksina (2018) aimed to optimize hydraulic fracture stages in unconventional gas formation based on cumulative production or NPV using AI-based optimization tools such as GA, Differential Evolution (DE), and Particle Swarm Optimization (PSO). Gradient-based methods are usually used for optimization purposes. However, they suffer from being trapped in local optima which means that the absolute optima could not be found. Also, many functions could not be differentiated at a certain value or range. Hence, this study was utilizing AI-based optimization tools that are gradient-free. The authors used Wattenbarger et al. (1998) analytical slap model to estimate gas cumulative flow within a certain production

period. The optimization function is the NPV which is a function of the cumulative gas production, water cumulative production, and cost of hydraulic fracturing and waste disposal. The objective is to find the optimum number of hydraulic fractures (NHF) that will maximize NPV. The PSO outperformed the other AI methods such as DE and GA as it required much fewer iteration for convergence.

Du et al. (2017) utilized embedded discrete fracture modeling (EDFM) to train an AI-based algorithm to estimate productivity in the Permian Basin. Authors used EDFM for fracture representation in a reservoir simulator; a method that reduces the need for using fine grids. The EDFM composes of two elements: matrix and fracture that can be represented separately. Mangrove which is a commercial software was used for hydraulic fracture network generation. AI was implemented to remove unnecessary fracture complexity that would not contribute to productivity. Using AI methods to reduce the complexity of the fracture and then implement it in EDFM resulted in significant simulation time reduction as compared only to using Mangrove. It enabled doing sensitivity analysis as it was feasible. However, the simplified structure resulted because the AI should be history matched to tune parameters such as reservoir permeability otherwise an error up 40% could be the outcome.

Bhattacharya et al. (2019) used machine learning algorithms to predict production in fractured Marcellus shale. The authors used the data of one well with 28 stages of hydraulic fractures in Marcellus shale to predict the production rate. The data used were petrophysical and geomechanical data (GR, sonic), pressure data (surface, casing, tubing), and fiber optics data such as distributed acoustic sensing (DAS) and distributed temperature sensing (DTS) while missing are hydraulic fracturing data and design. Ghahfarokhi et al. (2018) also implemented DAS and two years of DTS data for estimating production from Marcellus shale well. Bhattacharya et al. (2019) implemented the following machine learning tools: RF, ANN, and SVM. Feature engineering was implemented to find secondary attributes from the row data such as the brittleness index (BI). Collinearity analysis was implemented to find the most suitable features which reduced them from 34 originally to 18. All models could predict the production rate to good accuracy. However, SVM provided less accuracy with more computation time. Including hydraulic fracturing, reservoir, and PVT properties should improve accuracy. The model's lack of these data is a major limitation of their approach. Figure 6 shows that the Poisson ratio (PR) and brittleness index (BI) were the most important while DAS and DTS were not as significant.

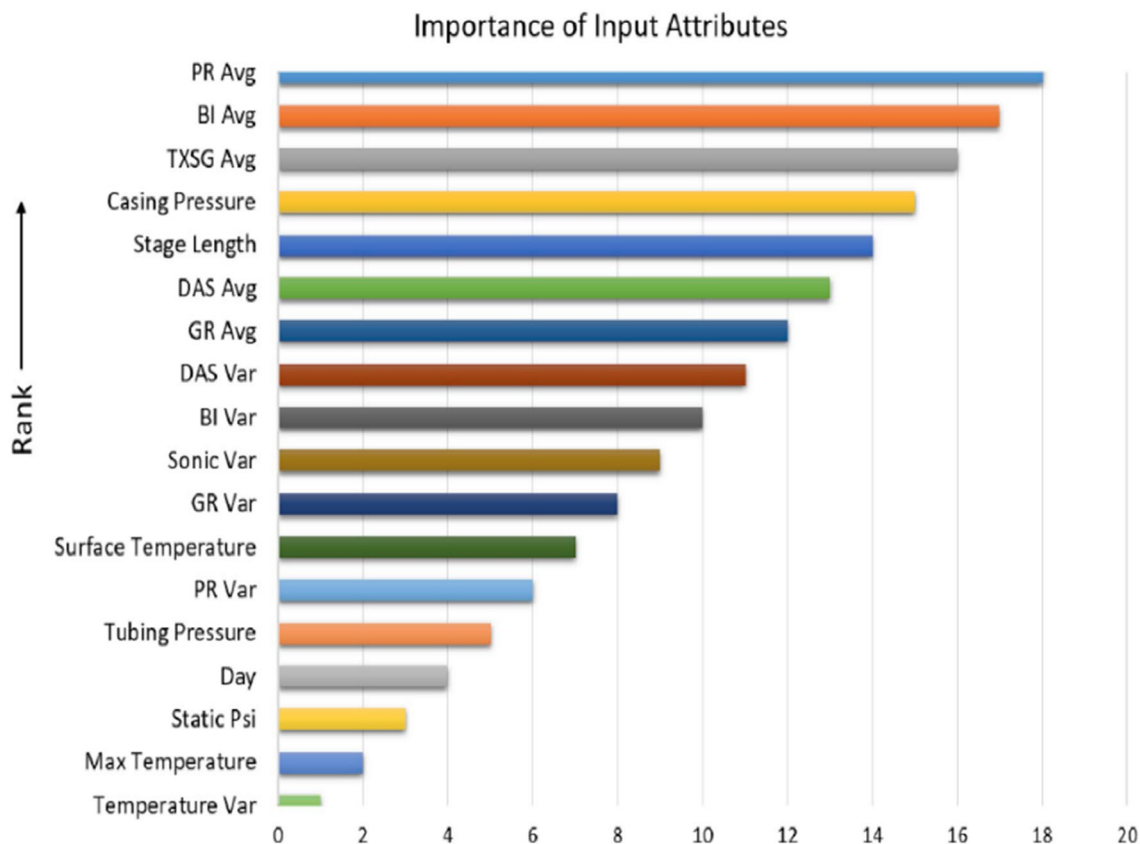


Fig. 6 Importance of attributes to production estimation of Marcellus shale (Bhattacharya et al. 2019)

Similar concepts were applied to other shale formations such as Bakken shale. Luo et al. (2019) investigated the possibility of predicting the productivity of horizontally drilled wells in Bakken shale based on completion and geological parameters. Geology and completion data of 2061 horizontal wells in the Bakken were used. These include vertical depth, amount of proppant, water saturation, porosity, permeability thickness... etc. Spearman correlation, RF, and joint mutual information (JMI) were used for feature selection. Deep learning (ANN) was used as a predictive model based on one-year production data. Based on feature selection, it was found that the formation thickness, depth, and amount of proppant are the most important parameters to predict the production in the first year. It was also observed that less porous spots require more proppant to increase productivity which agrees with the physics of unconventional. Wang et al. (2011) applied AI on 2780 MSF and 139 vertical wells in the Bakken to predict productivity. A deep neural network was used in the study with *k*-fold cross-validation to check the predictiveness of the model. The number of hidden layers and neurons was optimized to give the best prediction for 6 and 18 months. The model showed that the amount of proppant placed in each stage is the most important parameter in predicting productivity. The trained model resulted in a small root mean square error (RMSE) when predicting the 6 and 18 months of production.

Sweet spots identification in unconventional formation is an important process as horizontal drilling combined with MSF is an expensive process that should be justified by good productivity. Also, unconventional formations cover large areas, and hence, finding the right location to complete the well is critical. Tahmasebi et al. (2017) defined the sweet spots as the ones having high TOC and fracability index (FI). They used multiple linear regression (MLR) to train log data from shale formation to predict the TOC and FI. Nonlinear models such as FL, hybrid neural networks (NN)/FL, GA were also implemented. For variable selection, step-wise selection was implemented. Mineralogy composition was used to assess the fracability where quartz is the brittle mineral. MLR failed to predict FI where the correlation coefficient was 0.44 which is an unsatisfactory value. The prediction of TOC was better where the correlation coefficient was around 0.88. The hybrid (NN + FL) machine learning, among nonlinear models, (HML) could provide better accuracy and remove the weakly correlated variables.

Rastogi and Sharma (2019) used machine learning tools to find the impact of fracturing chemicals on production using one-year production data. Different algorithms were used for feature selection such as F-Regression, Decision tree-based regressions, recursive feature elimination ... etc. The data were obtained from different fracture jobs in the Powder River Basin. Chemicals additives were found to be

in the top 5 parameters that impact productivity out of 11 selected features.

AI has also been applied to the area of acid fracturing in terms of conductivity prediction. Acid self-prop the fracture by generating peaks and valleys that act as a conduit for the fluids to follow. Akbari et al. (2017) used a 106 data point generated experimentally to develop a conductivity correlation based on GA. The developed correlation resulted in better accuracy as compared to the popular correlations for acid fracture conductivity. Eleibide et al. (2018) applied ANN and adaptive network-based fuzzy inference systems on the same data set. The authors showed that the model accuracy was improved as compared to Akbari et al.'s model. Desouky et al. (2020a, 2020b) utilized more than 500 data points to generate a more accurate acid fracture conductivity correlation that considers rock type and etching pattern.

## Future and challenges

The utilization of ML techniques to handle a large data set and to predict several parameters in many aspects of the oil and gas industry is rapidly growing. The main reason behind that is the generation of large data in everyday activities of the oil and gas industry. To be able to process the large data and make it useful, a careful data processing and handling has to take place and ML techniques are a great tool to do that. Furthermore, due to the complexity of the different relationships between the many factors controlling the productivity of an oil or gas well, ML techniques are widely used to figure out these complex relationships and build a multilayered correlation to relate the different factors. Without ML, the classical linear/nonlinear regression methods do not have the capability to handle high complexity as ML models do. Also, the high uncertainty of the many oil and gas industry activities is a major concern given the capital-intensive nature of these activities, building a reliable forecast and prediction models are necessary to navigate through these challenges while optimizing the outcomes.

ML techniques have provided many solutions to the oil and gas industry to thrive. At the same time, there are many disadvantages of these models that are sometimes ignored or rarely mentioned. One of the main disadvantages of using ML in building a relationship between several parameters is whenever there is a high correlation, it does not necessarily imply causation (“correlation does not imply causation”). Building a high correlation model linking several parameters together based on the data used should not be taken as indication that these parameters are truly having a cause/effect relationship unless there is a proven physical or scientific relationship between them. Many developed models in the literature fail to address this fact and tend to associate correlation with causation. Another common challenge facing the

applicability of ML techniques is the availability and accuracy of data used to build and test these models. The data has to be accurate in order to produce a useful model. Otherwise, the model developed will never be useful no matter its high accuracy. Conducting data collection quality assurance is highly recommended to avoid this issue.

A common criticism of ML models is that they require a large and diverse data set to train the model. Any model needs a sufficient representative data in order to capture the underlying structure that allows it to generalize to new similar cases. For instance, a ML model built to predict production of a certain formation would only be applicable for that formation and under the same conditions when the training data it collected. Generalizing the predictive models has to be done with careful consideration of the constrain of these models and the diversity and inclusivity of the data used to build them. This is a major disadvantage of ML techniques as they tend to be generalized without careful consideration of this limitation.

Surly, ML cannot be used to predict anything related to oil and gas industry or build a correlation between any two or more parameters. Before undergoing building the relationships between the different factors, a scientific and factual explanation of the actual “physical” relationship between these parameters has to be addressed first. Also, using ML to predict and forecast based on historical data has to be done carefully by addressing and assuring that the future conditions are similar to the historical events. ML tend to be a very useful tool to deal with big data and to build the complex relationships between the different parameters that linear/nonlinear regression models cannot handle. Many of the correlation that has been established based on regression analysis of laboratory data are being replaced with correlation developed using ML methods that are more case specific rather than general correlations.

Deep learning which is a subset of ML based on ANN is very efficient for many tasks but it is not the solution to every problem as it faces many challenges. Deep learning algorithms need to be trained with large sets of data and the access and availability of accurate data is not always possible in many aspects of the oil and gas industry. Therefore, overfitting is considered the most common problem in ML applications which is mainly due to the lack of an appropriate amount of data to be used for training. Also, overtraining can happen when there is no clear stopping stage for the training and the error keeps decreasing by updating the model structure and the model become more complex to fit a specific dataset. Even when dealing with large data sets, a major challenge is the training cost. In many situations, supercomputers are needed to handle large oil and gas data sets to build and run ML models.

The future trend of ML applications in the oil and gas industry looks promising. With the arrival of the internet of

things and the automation of many of the oil and gas activities and the high reliance on data, it is possible to minimize risks and enhance productivity by integrating ML algorithms that are continuously trained and enhanced using the continuous flow of data. With the generation of large data in oil and gas industry, petroleum engineers and geoscientists must be exposed to big data handling techniques that are being developed in the AI domain. Making the most of the availability of data is something being addressed nowadays and will continue to be the trend for the future. Optimization cannot be reached without the utilization of the powerful capabilities of AI.

## Concluding remarks

Based on the review of the literature and the authors’ work on the applications of AI in petroleum engineering, the following remarks can be made:

- AI offers a huge potential in solving problems in almost all areas of the oil and gas industry involving prediction, classification, and clustering. Compared to other areas of engineering, petroleum engineering and geosciences have special relevance because of two important factors. Firstly, we deal with nature rather than man-made materials and processes. Variation in rocks, oil, and brine cannot be easily handled by any closed-form solution. The employment of numerical methods in such a situation is too cumbersome as well as unrealistic. Secondly, the amount of data produced every day from cores, logs, and seismic exploration in conventional reservoirs to multi-stage hydraulic fracturing in unconventional reservoirs is too huge to be interpreted properly using classical approaches.
- The biggest challenge for researchers is to have access to laboratory and/or field data. Oil companies should come forward to share data in whatever secured form acceptable to benefit the literature from the enormous potential that AI is ready to offer.
- With the arrival of the internet of things and the live relay of data from drilling and production facilities, it is possible to minimize risks and enhance production by integrating AI algorithms trained based on past data.
- There is a degree of uncertainty in the data coming out from the laboratories, logs, or seismic data. The depth record of the samples on which the measurements are taken may not be exact due to depth shifts, and the log data may not exactly represent the property corresponding to that depth point as it involves the averaging of the properties of layers that are intersected within the logging sensors such as transmitters and receivers. Seismic data also averages vast volumes of rocks in a given block.

As a result, a significant amount of data collected during the drilling operation is unreliable. Consequently, data cleaning and uncertainty quantification is another huge area that needs to be integrated with AI to develop realistic solutions to the problems.

- The solution to a problem using AI is rarely guaranteed if one attempts to relate the input with the target data without a thorough understanding of the physics of the problem. The major challenge in this approach is setting up the problem so that the algorithm can easily connect input and the target data. Wherever possible, information from analytical models should be suitably used in the input data to help the AI model arrive at the solution efficiently and more quickly.
- Problems involving the prediction of curves such as in viscosity-pressure curve in PVT data require extensive exploration of algebraic equations in addition to the understanding of the physics of the problem.
- Challenges remain in AI tools, too, in overfitting, coincidence effect, overtraining, etc. It is hoped that researchers in soft computing will develop modified and/or new AI tools.
- With the generation of huge data in petroleum engineering from cores to logs to seismic exploration, the petroleum engineers and geoscientists must be exposed to normal as well as big data handling techniques that are being developed in the AI domain. Exposure to AI tools should start right from their undergraduate education.

## Declarations

**Conflict of interest** Authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aadnoy BS, Fazaelizadeh M, Hareland G (2010) A 3D analytical model for wellbore friction. *J Can Pet Technol*. <https://doi.org/10.2118/141515-PA>
- Abbas AK, Al-haideri NA, Bashikh AA (2019) Implementing artificial neural networks and support vector machines to predict lost circulation. *Egypt J Pet* 28(4):339–347. <https://doi.org/10.1016/j.ejpe.2019.06.006>
- Abdelgawad K, Elkatatny S, Mousa T, Mahmoud M, Patil S (2018) Real time determination of rheological properties of spud drilling fluids using a hybrid artificial intelligence technique. *J Energy Resour Technol*. <https://doi.org/10.2118/192257-MS>
- Abdi Y, Garavand AT, Sahamieh RZ (2018) Prediction of strength parameters of sedimentary rocks using artificial neural networks and regression analysis. *Arab J Geosci* 11:587. <https://doi.org/10.1007/s12517-018-3929-0>
- Abdulraheem A, Ahmed M, Vantala A, Parvez T, (2009) Prediction of rock mechanical parameters for hydrocarbon reservoirs using different artificial intelligence techniques. In: Society of petroleum engineers—SPE Saudi Arabia section technical symposium 2009. Society of Petroleum Engineers. <https://doi.org/10.2118/126094-ms>
- Adesina FAS, Abiodun A, Anthony A, Olugbenga F (2015) Modeling the effect of temperature on environmentally safe oil based drilling mud using artificial neural network algorithm. *Pet Coal* 57(1):60–70
- Agwu OE, Akpabio JU, Alabi SB, Dosunmu A (2018) Artificial intelligence techniques and their applications in drilling fluid engineering: a review. *J Pet Sci Eng* 167:300–315. <https://doi.org/10.1016/j.petrol.2018.04.019>
- Ahmadi MA, Chen Z (2019) Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum* 5:271–284. <https://doi.org/10.1016/j.petlm.2018.06.002>
- Ahmadi MA, Pournik M, Shadizadeh SR (2015) Toward connectionist model for predicting bubble point pressure of crude oils: application of artificial intelligence. *Petroleum* 1:307–317. <https://doi.org/10.1016/j.petlm.2015.08.003>
- Ahmadi MA, Shadizadeh SR, Shah K, Bahadori A (2018) An accurate model to predict drilling fluid density at wellbore conditions. *Egypt J Pet* 27:1–10. <https://doi.org/10.1016/j.ejpe.2016.12.002>
- Ahmed A, Ali A, Elkatatny S, Abdulraheem A (2019) New artificial neural networks model for predicting rate of penetration in deep shale formation. *Sustain*. <https://doi.org/10.3390/su11226527>
- Akbari M, Ameri MJ, Kharazmi S, Motamedi Y, Pournik M (2017) New correlations to predict fracture conductivity based on the rock strength. *J Petr Sci Eng* 152:416–426. <https://doi.org/10.1016/j.petrol.2017.03.003>
- Al-Anazi A, Gates ID (2010) On the capability of support vector machines to classify lithology from well logs. *Nat Resour Res*. <https://doi.org/10.1007/s11053-010-9118-9>
- Al-Bulushi NI, King PR, Blunt MJ, Kraaijeveld M (2012) Artificial neural networks workflow and its application in the petroleum industry. *Neural Comput Appl* 21:409–421. <https://doi.org/10.1007/s00521-010-0501-6>
- Al-Hajri NM, Al-Ghamdi A, Tariq Z, Mahmoud M (2020) Scale-prediction/inhibition design using machine-learning techniques and probabilistic approach. *SPE Prod Oper* 35:0987–1009. <https://doi.org/10.2118/198646-PA>
- Al-Hameedi AT, Alkinani HH, Dunn-Norman S, Flori RE, Hilgedick SA, Amer AS, Alsaba MT (2018) using machine learning to predict lost circulation in the rumaila field, Iraq. In: Paper presented at the SPE Asia pacific oil and gas conference and exhibition, Brisbane, Australia. <https://doi.org/10.2118/191933-MS>

- Al-Marhoun MA, Osman EA (2002) Using artificial neural networks to develop New PVT correlations for Saudi crude oils. In: Society of petroleum engineers—Abu Dhabi international petroleum exhibition and conference 2002, ADIPEC 2002. Society of Petroleum Engineers, Abu Dhabi. <https://doi.org/10.2523/78592-ms>
- Al-Shammasi AA (2001) A review of bubblepoint pressure and oil formation volume factor correlations. *SPE Res Eval Eng* 4:146–160. <https://doi.org/10.2118/71302-PA>
- Alarifi SA, Miskimins J (2021) A new approach to estimating ultimate recovery for multistage hydraulically fractured horizontal wells by utilizing completion parameters using machine learning. *SPE Prod Oper* 36:468–483. <https://doi.org/10.2118/204470-PA>
- Alarifi S, AlNuaim S, Abdurraheem A (2015) Productivity index prediction for oil horizontal wells using different artificial intelligence techniques. In: All days. SPE. <https://doi.org/10.2118/172729-MS>
- Alkinani HH, Al-Hameedi ATT, Dunn-Norman S, Lian D (2020) Application of artificial neural networks in the drilling processes: can equivalent circulation density be estimated prior to drilling? *Egypt J Pet*. <https://doi.org/10.1016/j.ejpe.2019.12.003>
- Ameen MS, Smart BGD, Somerville JM, Hamilton S, Naji NA (2009) Predicting rock mechanical properties of carbonates from wireline logs (A case study: Arab-D reservoir, Ghawar field, Saudi Arabia). *Mar Pet Geol* 26:430–444. <https://doi.org/10.1016/j.marpetgeo.2009.01.017>
- Amiri M, Zahedi G, Yunan MH (2015) Water saturation estimation in tight shaly gas sandstones by application of progressive quasistatic (PQS) algorithm—a case study. *J Nat Gas Sci Eng* 22:468–477. <https://doi.org/10.1016/j.jngse.2014.12.027>
- Amiri M, Salamati A, Hatampour A, Rostami A, Heidari M (2014) Application of progressive quasistatic (PQS) algorithm in prediction of water saturation in tight gas sandstones—A case study. In: 20th Formation evaluation symposium of Japan 2014
- Andrea TA, Kalayeh H (1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J Med Chem* 34(9):2824–2836. <https://doi.org/10.1021/jm00113a022>
- Anemangely M, Ramezanzadeh A, Amiri H, Hoseinpour S-A (2019) Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *J Pet Sci Eng* 174:306–327. <https://doi.org/10.1016/j.petrol.2018.11.032>
- Anifowose F, Abdurraheem A (2011) Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization. *J Nat Gas Sci Eng*. <https://doi.org/10.1016/j.jngse.2011.05.002>
- Anifowose FA, Abdurraheem A, Al-Shuhail A, Schmitt DP (2013a) Improved permeability prediction from seismic and log data using artificial intelligence techniques. *SPE Middle East Oil Gas Show Conf MEOS Proc* 3:2190–2196. <https://doi.org/10.2118/164465-ms>
- Anifowose FA, Labadin J, Abdurraheem A (2013b) Prediction of petroleum reservoir properties using different versions of adaptive neuro-fuzzy inference system hybrid models. *Int J Comput Inf Syst Ind Manag Appl* 5:413–426
- Anifowose FA, Labadin J, Abdurraheem A (2017) Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. *J Pet Sci Eng* 151:480–487. <https://doi.org/10.1016/j.petrol.2017.01.024>
- Anifowose FA (2012) Advances in hybrid computational intelligence application in oil and gas reservoir characterization. In: Society of petroleum engineers—SPE Saudi Arabia section young professionals technical symposium 2012, YPTS 2012. pp. 1–8. <https://doi.org/10.2118/160921-ms>
- Arabjamaloei R, Shadzadeh S (2011) Modeling and optimizing rate of penetration using intelligent systems in an Iranian southern oil field (ahwaz oil field). *Pet Sci Technol*. <https://doi.org/10.1080/10916460902882818>
- Ariturk MS (2019) Optimizing the production and injection wells flow rates in geothermal field using artificial intelligence. In: Environmental science
- Aulia A, Rahman A, Quijano Velasco JJ (2014) Strategic well test planning using random forest. In: All days. SPE. <https://doi.org/10.2118/167827-MS>
- Avseth P, Mukerji T (2002) Seismic lithofacies classification from well logs using statistical rock physics. *Petrophysics* 43:70–81
- Awoleke OO, Lane RH (2011) Analysis of data from the barnett shale using conventional statistical and virtual intelligence techniques. *SPE Reserv Eval Eng*. <https://doi.org/10.2118/127919-PA>
- Ba alawi M, Gharbi S, Mahmoud M (2020) Modeling and prediction of resistivity, capillary pressure and relative permeability using artificial neural Network. In: Day 3 Wed, January 15, 2020. IPTC. <https://doi.org/10.2523/IPTC-19854-MS>
- Baarimah SO, Abdurraheem A, Anifowose FA (2014) Artificial intelligence (AI) techniques for predicting the reservoir fluid properties of crude-oil systems. In: Society of petroleum engineers—international petroleum technology conference 2014, IPTC 2014—innovation and collaboration: keys to affordable energy. pp. 3953–3968
- Bageri BS, Anifowose FA, Abdurraheem A (2015) Artificial intelligence based estimation of water saturation using electrical measurements data in a carbonate reservoir. In: SPE middle east oil and gas show and conference, MEOS, Proceedings. Society of Petroleum Engineers, pp. 499–515. <https://doi.org/10.2118/172564-ms>
- Bagheri A, Nazari A, Sanjayan J (2019) The use of machine learning in boron-based geopolymers: function approximation of compressive strength by ANN and GP. *Meas J Int Meas Confed*. <https://doi.org/10.1016/j.measurement.2019.03.001>
- Bahorich M, Farmer S (1995) The coherence cube. *Lead Edge* 14(10):1053–1058
- Bahorich MS, Farmer SL (1995) 3-D seismic discontinuity for faults and stratigraphic features: the coherence cube. In: 1995 SEG annual meeting. pp. 93–96. <https://doi.org/10.1190/1.1887523>
- Barbosa LFFM, Nascimento A, Mathias MH, de Carvalho JA (2019) Machine learning methods applied to drilling rate of penetration prediction and optimization: a review. *J Pet Sci Eng* 183:106332. <https://doi.org/10.1016/j.petrol.2019.106332>
- Bas CL (2016) Frugal innovation, sustainable innovation, reverse innovation; why do they look alike? Why are they different? *J Innov Econ Manag* 21:9–26
- Berthelot A, Solberg AHS, Gelius LJ (2013) Texture attributes for detection of salt. *J Appl Geophys*. <https://doi.org/10.1016/j.jappgeo.2012.09.006>
- Bhattacharya S, Ghahfarokhi PK, Carr TR, Pantaleone S (2019) Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiberoptic, completions, and surface data: a case study from the Marcellus Shale, North America. *J Petrol Sci Eng* 176:702–715
- Bhattacharya S, Mishra S (2018) Applications of machine learning for facies and fracture prediction using bayesian network theory and random forest: case studies from the Appalachian basin. *USA J Pet Sci Eng* 170:1005–1017. <https://doi.org/10.1016/j.petrol.2018.06.075>
- Bilgesu HI, Tetrack LT, Altmis U, Mohaghegh S, Ameri S (1997) New approach for the prediction of rate of penetration (ROP) values. In: Proceedings—SPE eastern regional conference and exhibition. <https://doi.org/10.2523/39231-ms>
- da Bispo VDS, Scheid CM, Calçada LA, da Meleiro LAC (2017) Development of an ANN-based soft-sensor to estimate the apparent viscosity of water-based drilling fluids. *J Pet Sci Eng*. <https://doi.org/10.1016/j.petrol.2016.11.030>

- Bondi G, Creamer R, Ferrari A, Fenton O, Wall D (2018) Using machine learning to predict soil bulk density on the basis of visual parameters: tools for in-field and post-field evaluation. *Geoderma* 318:137–147. <https://doi.org/10.1016/j.geoderma.2017.11.035>
- Boschert S, Rosen R (2016). Digital twin-the simulation aspect. In: *Mechatronic futures: challenges and solutions for mechatronic systems and their designers*. [https://doi.org/10.1007/978-3-319-32156-1\\_5](https://doi.org/10.1007/978-3-319-32156-1_5)
- Bourgoyne AT, Young FS (1999) A multiple regression approach to optimal drilling and abnormal pressure detection. *SPE Repr Ser*. <https://doi.org/10.2118/4238-pa>
- Brazell S, Bayeh A, Ashby M, Burton D (2019) A Machine-learning-based approach to assistive well-log correlation. *Petrophysics* 60:469–479. <https://doi.org/10.30632/PJV60N4-2019a1>
- Buhulaigah A, Al-Mashhad AS, Al-Arifi SA, Al-Kadem MS, Al-Dabbous MS (2017) Multilateral wells evaluation utilizing artificial intelligence. In: *SPE middle east oil & gas show and conference*. Society of petroleum engineers. <https://doi.org/10.2118/183688-MS>
- Chen F, Duan Y, Zhang J, Wang K, Wang W (2015) Application of neural network and fuzzy mathematic theory in evaluating the adaptability of inflow control device in horizontal well. *J Pet Sci Eng* 134:131–142
- Cheshmeh Sefidi A, Ajorkaran F (2019) A novel MLP-ANN approach to predict solution gas-oil ratio. *Pet Sci Technol* 37:2302–2308. <https://doi.org/10.1080/10916466.2018.1490759>
- Chhantyal K, Viumdal H, Mylvaganam S (2017) Ultrasonic level scanning for monitoring mass flow of complex fluids in open channels—a novel sensor fusion approach using AI techniques. In: *Proceedings of IEEE sensors*. <https://doi.org/10.1109/ICSENS.2017.8234010>
- Van Dao D, Ly HB, Trinh SH, Le TT, Pham BT (2019) Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete. *Materials (basel)*. <https://doi.org/10.3390/ma12060983>
- Deosarkar MP, Sathe VS (2012) Predicting effective viscosity of magnetite ore slurries by using artificial neural network. *Powder Technol*. <https://doi.org/10.1016/j.powtec.2011.12.058>
- Desouky M, Tariq Z, Aljawad MS, Alhoori H, Mahmoud M, AlShehri D (2020a) Data-driven acid fracture conductivity correlations honoring different mineralogy and etching patterns. *ACS Omega* 5(27):16919–16931. <https://doi.org/10.1021/acsomega.0c02123>
- Desouky M, Aljawad MS, Alhoori H, Al-Shehri D (2020b) Evaluating the effects of acid fracture etching patterns on conductivity estimation using machine learning techniques. In: *Paper presented at the SPE Europec, Virtual*. <https://doi.org/10.2118/200527-MS>
- Devi PRS, Baskaran R, Abirami S (2015) Multi-label learning with class-based features using extended centroid-based classification technique (CCBF). *Procedia Comput Sci* 54:405–411. <https://doi.org/10.1016/j.procs.2015.06.047>
- Di H, Gao D (2014) Gray-level transformation and Canny edge detection for 3D seismic discontinuity enhancement. *Comput Geosci* 72:192–200. <https://doi.org/10.1016/j.cageo.2014.07.011>
- Di H, Gao D (2016) Improved estimates of seismic curvature and flexure based on 3D surface rotation in the presence of structure dip. *Geophysics*. <https://doi.org/10.1190/GEO2015-0258.1>
- Di H, Wang Z, AlRegib G (2018) search and discovery.com; Deep convolutional neural networks for seismic salt-body delineation. (AAPG annual convention & exhibition)
- Du S, Liang B, Yuanbo L (2017) Field study: embedded discrete fracture modeling with artificial intelligence in permian basin for shale formation. In: *SPE annual technical conference and exhibition*. Society of petroleum engineers. <https://doi.org/10.2118/187202-MS>
- Dupriest FE, Koederitz WL (2005) Maximizing drill rates with real-time surveillance of mechanical specific energy. In: *All days*. SPE. <https://doi.org/10.2118/92194-MS>
- Eleibide M, Hassan AM, Mahmoud M, Abdurraheem A, Elkhatny S (2018) Intelligent prediction of acid-fracturing performance in carbonates reservoirs. *Society of Petroleum Engineers—SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition 2018, SATS 2018*,
- Elkhatny S, Mahmoud M, Tariq Z, Abdurraheem A (2018a) New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. *Neural Comput Appl* 30:2673–2683. <https://doi.org/10.1007/s00521-017-2850-x>
- Elkhatny S, Tariq Z, Mahmoud M (2016) Real time prediction of drilling fluid rheological properties using artificial neural networks visible mathematical model (white box). *J Pet Sci Eng* 146:1202–1210. <https://doi.org/10.1016/j.petrol.2016.08.021>
- Elkhatny S, Tariq Z, Mahmoud M, Abdurraheem A (2018b) New insights into porosity determination using artificial intelligence techniques for carbonate reservoirs. *Petroleum* 4:408–418. <https://doi.org/10.1016/j.petlm.2018.04.002>
- Elkhatny S, Tariq Z, Mahmoud M, Abdurraheem A, Mohamed I (2019) An integrated approach for estimating static Young's modulus using artificial intelligence tools. *Neural Comput Appl* 31:4123–4135. <https://doi.org/10.1007/s00521-018-3344-1>
- El-Sebakhy EA (2009) Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme. *J Petrol Sci Eng* 64(1–4):25–34. <https://doi.org/10.1016/j.petrol.2008.12.006>
- Evangelatosorn GI, Michael L. Payne, BP plc (2016) “Advanced BHA-ROP modeling including neural network analysis of drilling performance data”, IADC/SPE178852-MS, Fort Worth, Texas
- Feifei C, Yonggang D, Junbin ZW, Weifeng W (2015) Application of neural network and fuzzy mathematic theory in evaluating the adaptability of inflow control device in horizontal well. *J Pet Sci Eng*. <https://doi.org/10.1016/j.petrol.2015.07.020>
- Fertl WH, Hammack GW (1971) Comparative look at water saturation computations in shaly pay sands. *Soc. Petrophysicists Well-Log Anal*
- Ganji-Azad E, Rafiee-Taghanaki S, Rezaei H, Arabloo M (2014) Hossein Ali Zamani, Reservoir fluid PVT properties modeling using adaptive neuro-fuzzy inference systems. *J Nat Gas Sci Eng* 21:951–961. <https://doi.org/10.1016/j.jngse.2014.10.009>
- Gharbi RBC, Elsharkawy AM (1999) Neural network model for estimating the PVT properties of middle east crude oils. *SPE Reserv Eval Eng* 2:255–265. <https://doi.org/10.2118/56850-PA>
- Gharbi RB, Elsharkawy AM, Karkoub M (1999) Universal neural-network-based model for estimating the PVT properties of crude oil systems. *Energy Fuels* 13:454–458. <https://doi.org/10.1021/ef980143v>
- Gholanlo HH, Amirpour M, Ahmadi S (2016) Estimation of water saturation by using radial based function artificial neural network in carbonate reservoir: a case study in Sarvak formation. *Petroleum* 2:166–170. <https://doi.org/10.1016/j.petlm.2016.04.002>
- Gidh Y, Purwanto A, Bits S (2012) Artificial neural network drilling parameter optimization system improves ROP by predicting/managing bit wear. *Paper presented at the SPE Intelligent Energy International*, Utrecht, The Netherlands. <https://doi.org/10.2118/149801-MS>
- Gowida A, Elkhatny S, Al-Afnan S, Abdurraheem A (2020) New computational artificial intelligence models for generating synthetic formation bulk density logs while drilling. *Sustainability* 12:686. <https://doi.org/10.3390/su12020686>
- Guitton A, Wang H, Trainor-Guitton W (2017) Statistical imaging of faults in 3D seismic volumes using a machine learning approach. <https://doi.org/10.1190/segam2017-17589633.1>

- Hale D (2013) Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. *Geophysics*. <https://doi.org/10.1190/GEO2012-0331.1>
- Hassan A, Abdurraheem A, Elkhatatny S, Ahmed M (2017) New approach to quantify productivity of fishbone multilateral well. In: Proceedings—SPE annual technical conference and exhibition. <https://doi.org/10.2118/187458-ms>
- Hassan A, Al-Majed A, Elkhatatny S, Mahmoud M, Abdurraheem A, Nader M, Abughaban M, Khamis M (2018) Developing an efficient drilling system by coupling torque modelling with mechanical specific energy. In: Paper presented at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia. <https://doi.org/10.2118/192251-MS>
- Hassan A, Mahmoud M, Al-Majed A, Abdurraheem A (2020) A new technique to quantify the productivity of complex wells using artificial intelligence tools. In: Paper presented at the international petroleum technology conference, Dhahran, Kingdom of Saudi Arabia. <https://doi.org/10.2523/IPTC-19706-Abstract>
- He J, Tang M, Hu C, Tanaka S, Wang K, Wen X, Yusuf N (2021) Deep reinforcement learning for generalizable field development optimization. *SPE J*. <https://doi.org/10.2118/203951-PA>
- Hegde C, Gray K (2018) Evaluation of coupled machine learning models for drilling optimization. *J Nat Gas Sci Eng*. <https://doi.org/10.1016/j.jngse.2018.06.006>
- Helle HB, Bhatt A (2002) Fluid saturation from well logs using committee neural networks. *Pet Geosci* 8:109–118. <https://doi.org/10.1144/petgeo.8.2.109>
- Hemmati-Sarapardeh A, Shokrollahi A, Tatar A, Gharagheizi F, Mohammadi AH, Naseri A (2014) Reservoir oil viscosity determination using a rigorous approach. *Fuel* 116:39–48. <https://doi.org/10.1016/j.fuel.2013.07.072>
- Hoang M (2016) Tuning of viscosity and density of non-Newtonian fluids through mixing process using multimodal sensors, sensor fusion and models. 156
- Hossain M (2018) Frugal innovation: a review and research agenda. *J Clean Prod* 182:926–936
- Hu Y, Ding Y, Wen F, Liu L (2019) Reliability assessment in distributed multi-state series-parallel systems. *Energy Procedia* 159:104–110. <https://doi.org/10.1016/j.egypro.2018.12.026>
- Hua Y, Guo J, Zhao H (2015) Deep belief networks and deep learning. In: Proceedings of 2015 international conference on intelligent computing and internet of things. IEEE, pp. 1–4. <https://doi.org/10.1109/ICAOT.2015.7111524>
- Huang L, Dong X, Clee TE (2017) A scalable deep learning platform for identifying geologic features from seismic attributes. *Lead Edge*. <https://doi.org/10.1190/tle36030249.1>
- Huang Y, Zheng W, Zhang D, Xi Y (2020) A modified Herschel-Bulkley model for rheological properties with temperature response characteristics of poly-sulfonated drilling fluid. *Energy Sour Part A Recover Util Environ Eff* 42:1464–1475. <https://doi.org/10.1080/15567036.2019.1604861>
- Iino A, Jung HY, Onishi T, Datta-Gupta A (2020) Rapid simulation accounting for well interference in unconventional reservoirs using fast marching method. <https://doi.org/10.15530/urtec-2020-2468>
- Jacobs T (2019) The oil and gas chat bots are coming. *J Pet Technol* 71:34–36. <https://doi.org/10.2118/0219-0034-JPT>
- Jafari Kenari SA, Mashohor S (2013) Robust committee machine for water saturation prediction. *J Pet Sci Eng* 104:1–10. <https://doi.org/10.1016/j.petrol.2013.03.009>
- Jeirani Z, Mohebbi A (2006) Artificial neural networks approach for estimating filtration properties of drilling fluids. *J Japan Pet Inst* 49:65–70. <https://doi.org/10.1627/jpi.49.65>
- Jui-Sheng C, Chien-Kuo C, Mahmoud F, Ismail A-T (2011) Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. *J Comput Civ Eng* 25:242–253. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000088](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000088)
- Kamalyar K, Sheikhi Y, Jamialahmadi M (2012) Using an artificial neural network for predicting water saturation in an Iranian oil reservoir. *Pet Sci Technol* 30:35–45. <https://doi.org/10.1080/10916461003752561>
- Kar S, Das S, Ghosh PK (2014) Applications of neuro fuzzy systems: a brief review and future outline. *Appl Soft Comput* 15:243–259. <https://doi.org/10.1016/j.asoc.2013.10.014>
- Kewalramani MA, Gupta R (2006) Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. *Autom Constr*. <https://doi.org/10.1016/j.autcon.2005.07.003>
- Khamidy NI, Tariq Z, Syihab Z (2019) Development of ANN-based predictive model for miscible CO<sub>2</sub> flooding in sandstone reservoir. In: SPE middle east oil and gas show and conference, MEOS, Proceedings. Society of petroleum engineers. <https://doi.org/10.2118/194726-MS>
- Khashman A, Akpınar P (2017) Non-destructive prediction of concrete compressive strength using neural networks. *Procedia Comput Sci*. <https://doi.org/10.1016/j.procs.2017.05.039>
- Khazaeni Y, Mohaghegh SD (2011) Intelligent production modeling using full field pattern recognition. *SPE Reserv Eval Eng* 14(6):735–749. <https://doi.org/10.2118/132643-PA>
- Kumar A, Ridha S, Ganet T, Vasant P, Ilyas SU (2020) Machine learning methods for herschel-bulkley fluids in annulus: pressure drop predictions and algorithm performance evaluation. *Appl Sci*. <https://doi.org/10.3390/app10072588>
- Lecampion B, Bungler A, Zhang X (2018) Numerical methods for hydraulic fracture propagation: a review of recent trends. *J Nat Gas Sci Eng*. <https://doi.org/10.1016/j.jngse.2017.10.012>
- Liu H, Lang B (2019) Machine learning and deep learning methods for intrusion detection systems: a survey. *Appl Sci* 9:4396. <https://doi.org/10.3390/app9204396>
- Livingstone DJ, Manalack DT (1993) Statistics using neural networks: chance effects. *J Med Chem* 36(9):1295–1297. <https://doi.org/10.1021/jm00061a023>
- Livingstone D, Manalack D, Tetko I (1997) Data modelling with neural networks: advantages and limitations. *J Comput Aided Mol Des* 11:135–142. <https://doi.org/10.1023/A:1008074223811>
- Lolon E, Hamidieh K, Weijers L, Mayerhofer M, Melcher H, Oduba O (2016) Evaluating the relationship between well parameters and production using multivariate statistical models: a middle bakken and three forks case history. In: Day 3 Thu, February 11, 2016. SPE. <https://doi.org/10.2118/179171-MS>
- Luo G, Tian Y, Bychina M, Ehlig-Economides C (2019) Production-strategy insights using machine learning: application for bakken shale. *SPE Reservoir Eval Eng* 22(3):800–816
- Luthi SM, Bryant ID (1997) Well-log correlation using a back-propagation neural network. *Math Geol* 29:413–425. <https://doi.org/10.1007/BF02769643>
- Maghrabi S, Kulkarni D, Teke K, Kulkarni SD, Jamison D (2014) Modeling of shale-erosion behavior in aqueous drilling fluids. In: Society of petroleum engineers—European unconventional resources conference and exhibition 2014: unlocking European potential. <https://doi.org/10.2118/167691-ms>
- Maity D, Ciezobka J (2019) Designing a robust proppant detection and classification workflow using machine learning for subsurface fractured rock samples post hydraulic fracturing operations. *J Petrol Sci Eng* 172:588–606. <https://doi.org/10.1016/j.petrol.2018.09.062>
- Maniar H, Ryali S, Kulkarni MS, Abubakar A (2018) Machine-learning methods in geoscience. In: SEG technical program expanded abstracts 2018. Society of exploration geophysicists, pp. 4638–4642. <https://doi.org/10.1190/segam2018-2997218.1>



- Manshad A, Rostami H, Niknafs H, Mohammadi A (2017) Integrated lost circulation prediction in oil field drilling operation. *Research Gate* 978–1–53610–852–1
- Mardi M, Nurozi H, Edalatkhah S (2012) A water saturation prediction using artificial neural networks and an investigation on cementation factors and saturation exponent variations in an Iranian oil well. *Pet Sci Technol* 30:425–434. <https://doi.org/10.1080/10916460903452033>
- Marfurt KJ, Kirilin RL, Farmer SL, Bahorich MS (1998) 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics* DOI 10(1190/1):1444415
- Martinelli G, Eidsvik J, Sinding-Larsen R, Rekstad S, Mukerji T (2013) Building bayesian networks from basin-modelling scenarios for improved geological decision making. *Pet Geosci*. <https://doi.org/10.1144/petgeo2012-057>
- Masoudi R, Mohaghegh SD, Yingling D, Ansari A, Amat H, Mohamad N, Sabzabadi A, Dipak M (2020) Subsurface analytics case study; reservoir simulation and modeling of highly complex offshore field in malaysia, using artificial intelligent and machine learning. In: SPE annual technical conference and exhibition, October 2020. <https://doi.org/10.2118/201693-MS>
- Matin SS, Farahzadi L, Makaremi S, Chelgani SC, Sattari G (2018) Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. *Appl Soft Comput* 70:980–987. <https://doi.org/10.1016/j.asoc.2017.06.030>
- Mehrad M, Bajolvand M, Ramezanzadeh A (2020) Jalil Ghavidel Neycharan, developing a new rigorous drilling rate prediction model using a machine learning technique. *J Petrol Sci Eng* 192:107338. <https://doi.org/10.1016/j.petrol.2020.107338>
- Melville PD, Guruswamy S (2002) Integration of 3D seismic with a reservoir model for reservoir characterisation. In: Abu Dhabi international petroleum exhibition and conference. <https://doi.org/10.2118/78511-MS>
- Miri R, Sampaio JHB, Afshar M, Lourenco A (2007) Development of artificial neural networks to predict differential pipe sticking in iranian offshore oil fields. In: International oil conference and exhibition in Mexico. Society of petroleum engineers. <https://doi.org/10.2118/108500-MS>
- Mishra S, Schuetter J, Zhong M, Lafollette R (2015) Data analytics for production optimization in unconventional reservoirs. <https://doi.org/10.15530/urtec-2015-2167005>
- Mnati KH, Hadi HA (2018) Prediction of penetration rate and cost with artificial neural network for alhafaya oil field. *Iraqi J Chem Pet Eng* 19:21–27. <https://doi.org/10.31699/IJCPE.2018.4.3>
- Mohaghegh S (2000) Virtual-intelligence applications in petroleum engineering: part 3—fuzzy logic. *J Pet Technol* 52:82–87. <https://doi.org/10.2118/62415-JPT>
- Mohaghegh S (2011) Reservoir simulation and modeling based on artificial intelligence and data mining (AI & DM). *J Nat Gas Sci Eng* 3(6):697–705. <https://doi.org/10.1016/j.jngse.2011.08.003>
- Mohaghegh SD (2017) Shale analytics: data-driven analytics in unconventional resources. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-48753-3>
- Mollajan A, Memarian H, Jalali MR (2013) Prediction of reservoir water saturation using support vector regression in an Iranian carbonate reservoir. In: 47th US rock mechanics/geomechanics symposium 2013. pp. 1872–1877
- Moore BA, Rougier E, O'Malley D, Srinivasan G, Hunter A, Viswanathan H (2018) Predictive modeling of dynamic fracture growth in brittle materials with machine learning. *Comput Mater Sci*. <https://doi.org/10.1016/j.commatsci.2018.01.056>
- Murray AS, Cunningham RA (1955) Effect of mud column pressure on drilling rates. *Trans AIME*. <https://doi.org/10.2118/505-g>
- Murtaza M, Mahmoud M, Tariq Z (2020) Experimental investigation of a novel, efficient, and sustainable hybrid silicate system in oil and gas well cementing. *Energy Fuels* 34:7388–7396. <https://doi.org/10.1021/acs.energyfuels.0c01001>
- Naderpour H, Rafiean AH, Fakharian P (2018) Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *J Build Eng*. <https://doi.org/10.1016/j.jobbe.2018.01.007>
- Najibi AR, Ghafoori M, Lashkaripour GR, Asef MR (2015) Empirical relations between strength and static and dynamic elastic properties of Asmari and Sarvak limestones, two main oil reservoirs in Iran. *J Pet Sci Eng* 126:78–82. <https://doi.org/10.1016/j.petrol.2014.12.010>
- Navrátil J, King A, Rios J, Kollias G, Torrado R, Coda A (2019) Accelerating physics-based simulations using end-to-end neural network proxies: an application in oil reservoir modeling. *Front Big Data* 2:33. <https://doi.org/10.3389/fdata.2019.00033>
- Negara A, Ali S, AlDhamen A, Kesserwan H, Jin G (2017) Unconfined compressive strength prediction from petrophysical properties and elemental spectroscopy using support-vector regression. In: Day 4 Thu, April 27, 2017. SPE. <https://doi.org/10.2118/188077-MS>
- Nooruddin HA, Anifowose F, Abdulraheem A (2014) Using soft computing techniques to predict corrected air permeability using Thomeer parameters, air porosity and grain density. *Comput Geosci* 64:72–80. <https://doi.org/10.1016/j.cageo.2013.11.007>
- Numbere OG, Azuibuikwe II, Ikiensikimama SS (2013) bubble point pressure prediction model for niger delta crude using artificial neural Network approach. In: Paper presented at the SPE Nigeria Annual international conference and exhibition, Lagos, Nigeria. <https://doi.org/10.2118/167586-MS>
- Ogutade T, Ojo T, Efajemue E, Oni B, Idaka J (2020) Application of ANN in predicting water based mud rheology and filtration properties. <https://doi.org/10.2118/203720-MS>
- Omar S, Ngadi A, Jebur HH (2013) machine learning techniques for anomaly detection: an overview. *Int J Comput Appl* 79:33–41. <https://doi.org/10.5120/13715-1478>
- Osarogiagbon A, Muojeke S, Venkatesan R, Khan F, Gillard P (2020) A new methodology for kick detection during petroleum drilling using long short-term memory recurrent neural network. *Process Saf Environ Prot* 142:126–137. <https://doi.org/10.1016/j.psep.2020.05.046>
- Osman EA, Aggour MA (2003) Determination of drilling mud density change with pressure and temperature made simple and accurate by ANN. In: Proceedings of the middle east oil show. <https://doi.org/10.2118/81422-ms>
- Osman EA, Abdel-Wahhab OA, Al-Marhoun MA (2001) Prediction of Oil PVT Properties Using Neural Networks. Paper presented at the SPE Middle East Oil Show, Manama, Bahrain. <https://doi.org/10.2118/68233-MS>
- Pandey SK, Janghel RR (2019) Recent deep learning techniques, challenges and its applications for medical healthcare system: a review. *Neural Process Lett* 50:1907–1935. <https://doi.org/10.1007/s11063-018-09976-2>
- Qi L, Carr TR (2006) Neural network prediction of carbonate lithofacies from well logs, big bow and sand arroyo creek fields, Southwest Kansas. *Comput Geosci* 32:947–964. <https://doi.org/10.1016/j.cageo.2005.10.020>
- Qi J, Li F, Marfurt K (2017) Multiazimuth coherence. *Geophysics*. <https://doi.org/10.1190/GEO2017-0196.1>
- Rabbani E, Sharif F, Koolivand Salooki M, Moradzadeh A (2012) Application of neural network technique for prediction of uniaxial compressive strength using reservoir formation properties. *Int J Rock Mech Min Sci* 56:100–111. <https://doi.org/10.1016/j.ijrmms.2012.07.033>
- Rahmanifard H, Plaksina T (2018) Application of fast analytical approach and AI optimization techniques to hydraulic fracture

- stage placement in shale gas reservoirs. *J Nat Gas Sci Eng* 52:367–378. <https://doi.org/10.1016/j.jngse.2018.01.047>
- Rahmati AS, Tatar A (2019) Application of radial basis function (RBF) neural networks to estimate oil field drilling fluid density at elevated pressures and temperatures. *Oil Gas Sci Technol—Rev d'IFP Energies Nouv.* <https://doi.org/10.2516/ogst/2019021>
- Raj J (2019) A comprehensive survey on the computational intelligence techniques and its applications. *J ISMAC* 1:147–159. <https://doi.org/10.36548/jismac.2019.3.002>
- Ramamoorthy A, Yampolskiy R (2018) Beyond map?: the race for artificial general intelligence. *ITU J* 1(1):77–84
- Ramirez AM, Valle GA, Romero F, Jaimes M (2017) Prediction of PVT properties in crude oil using machine learning techniques MLT. In: SPE Latin American and Caribbean petroleum engineering conference proceedings. Society of petroleum engineers. <https://doi.org/10.2118/185536-ms>
- Rasheed A, San O, Kvamsdal T (2020) Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access* 8:21980–22012. <https://doi.org/10.1109/ACCESS.2020.2970143>
- Rasheed Khan M, Tariq Z, Abdurraheem A (2018) Machine learning derived correlation to determine water saturation in complex lithologies. In: Society of petroleum engineers—SPE Kingdom of Saudi Arabia annual technical symposium and exhibition 2018, SATS 2018. Society of petroleum engineers. <https://doi.org/10.2118/192307-ms>
- Rastogi A, Sharma A (2019) Quantifying the impact of fracturing chemicals on production performance using machine learning. In: Paper presented at the SPE Liquids-Rich Basins conference—North America, Odessa, Texas, USA. <https://doi.org/10.2118/197095-MS>
- Razi MM, Mazidi M, Razi FM, Aligolzadeh H, Niazi S (2013) Artificial neural network modeling of plastic viscosity, yield point, and apparent viscosity for water-based drilling fluids. *J Dispers Sci Technol.* <https://doi.org/10.1080/01932691.2012.704746>
- Reiber F, Vos BE, Eide SE (1999) On-line torque & drag: a real-time drilling performance optimization tool. In: Proceedings of the IADC/SPE Asia Pacific drilling technology conference, APDT. <https://doi.org/10.2523/52836-ms>
- Rolon L, Mohaghegh SD, Ameri S, Gaskari R, McDaniel B (2009) Using artificial neural networks to generate synthetic well logs. *J Nat Gas Sci Eng.* <https://doi.org/10.1016/j.jngse.2009.08.003>
- Rooki R, Doulati Ardejani F, Moradzadeh A (2014) Hole cleaning prediction in foam drilling using artificial neural network and multiple linear regression. *Geomaterials.* <https://doi.org/10.4236/gm.2014.41005>
- Rooki R, Doulati Ardejani F, Moradzadeh A, Kelessidis VC, Nourozi M (2012) Prediction of terminal velocity of solid spheres falling through newtonian and non-newtonian pseudoplastic power law fluid using artificial neural network. *Int J Miner Process.* <https://doi.org/10.1016/j.minpro.2012.03.012>
- Safiuddin M, Raman SN, Salam MA, Jumaat MZ (2016) Modeling of compressive strength for self-consolidating high-strength concrete incorporating palm oil fuel ash. *Materials (basel).* <https://doi.org/10.3390/ma9050396>
- Sagheer A, Kotb M (2019) Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* 323:203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>
- Salehi SM, Honarvar B (2014) Automatic identification of formation lithology from well log data: a machine learning approach. *J Pet Sci Res.* <https://doi.org/10.14355/jpsr.2014.0302.04>
- Samek W, Wiegand T, Muller KR (2018) Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU J* 1(1):39
- Samnejad M, Gharib Shirangi M, Etehad R (2020) A digital twin of drilling fluids rheology for real-time rig operations. <https://doi.org/10.4043/30738-MS>
- Sebtosheikh MA, Salehi A (2015) Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir. *J Pet Sci Eng* 134:143–149. <https://doi.org/10.1016/j.petrol.2015.08.001>
- Seyyedattar M, Ghiasi MM, Zendejboudi S, Butt S (2020) Determination of bubble point pressure and oil formation volume factor: extra trees compared with LSSVM-CSA hybrid and ANFIS models. *Fuel* 269:116834. <https://doi.org/10.1016/j.fuel.2019.116834>
- Shabbir J, Anwer T (2018) Artificial intelligence and its role in near future. *J Latex Class Files* 1(8):1–11
- Shadravan A, Tarrahi M, Amani M (2017) Intelligent tool to design drilling, spacer, cement slurry, and fracturing fluids by use of machine-learning algorithms. *SPE Drill Complet* 32:131–140. <https://doi.org/10.2118/175238-PA>
- Shahkarami A, Mohaghegh S, Gholami V, Haghghat A, Moreno D (2014) Modeling pressure and saturation distribution in a CO<sub>2</sub> storage project using a surrogate reservoir model (SRM). *Greenh Gases Sci Technol* 4:289–315. <https://doi.org/10.1002/ghg.1414>
- Shahriar A, Nehdi ML (2011) Modeling rheological properties of oil well cement slurries using artificial neural networks. *J Mater Civ Eng.* [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0000340](https://doi.org/10.1061/(ASCE)MT.1943-5533.0000340)
- Sharma MSR, O'Regan M, Baxter CDP, Moran K, Vaziri H, Narayanasamy R (2010) Empirical relationship between strength and geophysical properties for weakly cemented formations. *J Pet Sci Eng* 72:134–142. <https://doi.org/10.1016/j.petrol.2010.03.011>
- Shojaei M-J, Bahrami E, Barati P, Riahi S (2014) Adaptive neuro-fuzzy approach for reservoir oil bubble point pressure estimation. *J Nat Gas Sci Eng* 20:214–220. <https://doi.org/10.1016/j.jngse.2014.06.012>
- Shokir EMEM (2004) Prediction of the hydrocarbon saturation in low resistivity formation via artificial neural network. In: Proceedings of the SPE Asia Pacific conference on integrated modelling for asset management. Society of petroleum engineers, pp. 35–40. <https://doi.org/10.2118/87001-ms>
- Shokrollahi A, Tatar A, Safari H (2015) On accurate determination of PVT properties in crude oil systems: committee machine intelligent system modeling approach. *J Taiwan Inst Chem Eng* 55:17–26. <https://doi.org/10.1016/j.jtice.2015.04.009>
- Van Si L, Chon BH (2018) Effective prediction and management of a CO<sub>2</sub> flooding process for enhancing oil recovery using artificial neural networks. *J Energy Resour Technol Trans ASME.* <https://doi.org/10.1115/1.4038054>
- Sibai FN, Hosani HI, Naqbi RM, Dhanhani S, Shehhi S (2011) Iris recognition using artificial neural networks. *Expert Syst Appl* 38:5940–5946. <https://doi.org/10.1016/j.eswa.2010.11.029>
- Silva AA, Lima Neto IA, Misságia RM, Ceia MA, Carrasquilla AG, Archilha NL (2015) Artificial neural networks to support petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. *J Appl Geophys* 117:118–125. <https://doi.org/10.1016/j.jappgeo.2015.03.027>
- Simandoux P (1963) Mesures dielectriques en milieu poreux, application a mesure des saturation eau. *Inst Fr du Pet Suppl Issue*
- Solomon O, Adewale D, Anyanwu C (2017) Fracture width prediction and loss prevention material sizing in depleted formations using artificial intelligence. In: Paper presented at the SPE Nigeria annual international conference and exhibition, Lagos, Nigeria. <https://doi.org/10.2118/189068-MS>
- Somasundaram S, Mund B, Soni R, Sharda R (2017) Seismic attribute analysis for fracture detection and porosity prediction: a case study from tight volcanic reservoirs, Barmer Basin, India. *Lead Edge.* <https://doi.org/10.1190/le36110947b1.1>
- Somvanshi M, Chavan P, Tambade S, Shinde SV (2016) A review of machine learning techniques using decision tree and support vector machine. In: 2016 International conference on computing

- communication control and automation (ICCUBEA). IEEE, pp. 1–7. <https://doi.org/10.1109/ICCUBEA.2016.7860040>
- Srinivasan G, Hyman JD, Osthus DA, Moore BA, O'Malley D, Karra S, Rougier E, Hagberg AA, Hunter A, Viswanathan HS (2018) Quantifying topological uncertainty in fractured systems using graph theory and machine learning. *Sci Rep*. <https://doi.org/10.1038/s41598-018-30117-1>
- Sun H, Belhaj H, Tao G, Vega S, Liu L (2019) Rock properties evaluation for carbonate reservoir characterization with multi-scale digital rock images. *J Pet Sci Eng* 175:654–664. <https://doi.org/10.1016/j.petrol.2018.12.075>
- Sun Q, Ertekin T (2020) Screening and optimization of polymer flooding projects using artificial-neural-network (ANN) based proxies. *J Pet Sci Eng*. <https://doi.org/10.1016/j.petrol.2019.106617>
- Tahmasebi P, Javadpour F, Sahimi M (2017) Data mining and machine learning for identifying sweet spots in shale reservoirs. *Expert Syst Appl* 88:435–447. <https://doi.org/10.1016/j.eswa.2017.07.015>
- Tariq Z, Elkhatny S, Mahmoud M, Abdurraheem A (2016) A New artificial intelligence based empirical correlation to predict sonic travel time, In: International petroleum technology conference. International petroleum technology conference. <https://doi.org/10.2523/iptc-19005-ms>
- Tariq Z, Elkhatny S, Mahmoud M, Ali AZ, Abdurraheem A, (2017a) A new technique to develop rock strength correlation using artificial intelligence tools. In: Society of petroleum engineers—SPE reservoir characterisation and simulation conference and exhibition, RCSC 2017. Society of petroleum engineers, pp. 1340–1353. <https://doi.org/10.2118/186062-MS>
- Tariq Z, Elkhatny S, Mahmoud M, Ali AZ, Abdurraheem A (2017b) A new approach to predict failure parameters of carbonate rocks using artificial intelligence tools. In: Society of petroleum engineers—SPE Kingdom of Saudi Arabia annual technical symposium and exhibition 2017. Society of petroleum engineers, pp. 1428–1440. <https://doi.org/10.2118/187974-MS>
- Tariq Z (2018) An automated flowing bottom-hole pressure prediction for a vertical well having multiphase flow using computational intelligence techniques. In: Society of petroleum engineers—SPE Kingdom of Saudi Arabia annual technical symposium and exhibition 2018, SATS 2018. Society of petroleum engineers. <https://doi.org/10.2118/192184-MS>
- Tariq Z, Abdurraheem A, Mahmoud M, Ahmed A (2018) A rigorous data-driven approach to predict Poisson's ratio of carbonate rocks using a functional network. *Petrophysics* 59:761–777. <https://doi.org/10.30632/PJV59N6-2018a2>
- Tariq Z, Mahmoud M, Abdurraheem A (2019) Core log integration: a hybrid intelligent data-driven solution to improve elastic parameter prediction. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04101-3>
- Tariq Z, Mahmoud M, Abdurraheem A (2020a) An intelligent data-driven model for dean-stark water saturation prediction in carbonate rocks. *Neural Comput Appl* 32:11919–11935. <https://doi.org/10.1007/s00521-019-04674-z>
- Tariq Z, Murtaza M, Mahmoud M (2020b) Effects of nanoclay and silica flour on the mechanical properties of class G cement. *ACS Omega* 5:11643–11654. <https://doi.org/10.1021/acsomega.0c00943>
- Tariq Z, Murtaza M, Mahmoud M (2020c) Development of new rheological models for class G cement with nanoclay as an additive using machine learning techniques. *ACS Omega* 5:17646–17657. <https://doi.org/10.1021/acsomega.0c02122>
- Tariq Z, Mahmoud M, Abdurraheem A (2021) Machine learning-based improved pressure–volume–temperature correlations for black oil reservoirs. *J Energy Resour Technol*. <https://doi.org/10.1115/1.4050579>
- Taunk K, De S, Verma S, Swetapadma A (2019) A Brief review of nearest neighbor algorithm for learning and classification. In: 2019 International conference on intelligent computing and control systems (ICCS). IEEE, pp. 1255–1260. <https://doi.org/10.1109/ICCS45141.2019.9065747>
- Tian X, Daigle H (2019) Preferential mineral-microfracture association in intact and deformed shales detected by machine learning object detection. *J Nat Gas Sci Eng* 63:27–37. <https://doi.org/10.1016/j.jngse.2019.01.003>
- Tomiwa O, Oluwatosin R, Temiloluwa O, Oluwasanmi O, Joy I (2019) Improved water based mud using solanum tuberosum formulated biopolymer and application of artificial neural network in predicting mud rheological properties. <https://doi.org/10.2118/198861-MS>
- Trtnik G, Kavčič F, Turk G (2009) Prediction of concrete strength using ultrasonic pulse velocity and artificial neural networks. *Ultrasonics*. <https://doi.org/10.1016/j.ultras.2008.05.001>
- Tschannen V, Delescluse M, Etrich N, Keuper J (2020) Extracting horizon surfaces from 3D seismic data using deep learning. *Geophysics* 85:N17–N26. <https://doi.org/10.1190/geo2019-0569.1>
- Van SL, Chon BH (2017b) Applicability of an artificial neural network for predicting water-alternating-CO<sub>2</sub> performance. *Energies*. <https://doi.org/10.3390/en10070842>
- Le Van S, Chon BH (2017a) Evaluating the critical performances of a CO<sub>2</sub>-Enhanced oil recovery process using artificial neural network models. *J Pet Sci Eng*. <https://doi.org/10.1016/j.petrol.2017.07.034>
- Virginia D (2018) Responsible artificial intelligence: designing AI for human values. *ITU J* 1(1):1–8
- Wang S, Chen Z, Chen S (2019) Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. *J Pet Sci Eng* 179:112–125. <https://doi.org/10.1016/j.petrol.2019.04.016>
- Wang Y, Duan M, Wang D, Liu J, Dong Y (2011) A model for deep-water floating platforms selection based on BP artificial neural networks
- Wattenbarger R, El-Banbi A, Villegas M, Maggard J (1998) Production analysis of linear flow into fractured tight gas wells. In: Paper SPE 39931 presented at the SPE rocky mountain regional/low-permeability reservoirs symposium, Denver, Colorado, pp. 5–8 April. Paper Presented at the Low-permeability Reservoirs Symposium, Denver, CO
- Weyrauch T, Herstatt C (2016) What is frugal innovation? Three defining criteria. *J Frugal Innov* 2(1):1–17
- Xiong W, Ji X, Ma Y, Wang Y, Benhassan NM, Ali MN, Luo Y (2018) Seismic fault detection with convolutional neural network. *Geophysics*. <https://doi.org/10.1190/geo2017-0666.1>
- Yilmaz I, Yuksek AG (2008) An example of artificial neural network (ANN) application for indirect estimation of rock parameters. *Rock Mech Rock Eng* 41:781–795. <https://doi.org/10.1007/s00603-007-0138-7>
- Yadolahi MM, Benli A, Demirboğa R (2015) Prediction of compressive strength of geopolymer composites using an artificial neural network. *Mater Res Innov* 19(6):453–458. <https://doi.org/10.1179/1433075X15Y.0000000020>
- Yang D, Rahman MK, Chen Y (2008) Bottomhole assembly analysis by finite difference differential method. *Int J Numer Methods Eng*. <https://doi.org/10.1002/nme.2221>
- You L, Tan Q, Kang Y, Xu C, Lin C (2018) Reconstruction and prediction of capillary pressure curve based on particle swarm optimization-back propagation neural network method. *Petroleum* 4:268–280. <https://doi.org/10.1016/j.petlm.2018.03.004>
- Yu W, Sepehrnoori K (2013) Optimization of multiple hydraulically fractured horizontal wells in unconventional gas reservoirs. In: SPE production and operations symposium, proceedings. <https://doi.org/10.2118/164509-ms>

- Zapico JL, Gonzalez-Buelga A, Gonzalez MP, Alonso R (2008) Finite element model updating of a small steel frame using neural networks. *Smart Mater Struct*. <https://doi.org/10.1088/0964-1726/17/4/045016>
- Zhao T, Jayaram V, Roy A, Marfurt KJ (2015) A comparison of classification techniques for seismic facies recognition. *Interpretation*. <https://doi.org/10.1190/INT-2015-0044.1>
- Zheng Y, Zhang Q, Yusifov A, Shi Y (2019) Applications of supervised deep learning for seismic interpretation and inversion. *Lead Edge* 38:526–533. <https://doi.org/10.1190/tle38070526.1>
- Zhou H, Niu X, Fan H, Wang G (2016) Effective calculation model of drilling fluids density and ESD for HTHP well while drilling. In: Society of petroleum engineers—IADC/SPE Asia Pacific drilling technology conference. <https://doi.org/10.2118/180573-MS>
- Zou G, Ren K, Sun Z, Peng S, Tang Y (2019) Fault interpretation using a support vector machine: A study based on 3D seismic mapping of the Zhaozhuang coal mine in the Qinshui Basin, China. *J Appl Geophys*. <https://doi.org/10.1016/j.jappgeo.2019.103870>

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