Unconventional Method for Prediction of Petrophysical Properties from Basic Logs using Machine Learning in a Brown Field: A Case Study

Ashok[1], M V Ramanand, Sujay Thakur
[1]Email: soni_ashok@ongc.co.in, Oil and Natural Gas Corporation Limited

Abstract

Petrophysical properties (effective porosity, water saturation and volume of clay) are key inputs for reservoir characterization during exploration and development phase of a field. These properties generally computed from basic logs through conventional log data processing and interpretation using deterministic / multi-mineral probabilistic approach. This conventional method is relatively more time consuming and expensive as compared to machine learning approach. Recent advances in petroleum exploration technologies using machine learning have granted a new light in the industry for more economical, efficient and accurate reservoir characterization. This paper introduces a successful application of two supervised machine learning algorithms viz. k-Nearest Neighbour (k-NN) and Multi-layer feed forward Artificial Neural Network (ANN) to predict the petrophysical properties i.e. effective porosity (PHIE), effective water saturation (SWE) & volume of clay (VCL) in the fluvial channel sandstone reservoir in brown field of ONGC.

In the present paper, thirty numbers of wells of the brown field have been shortlisted for predicting the petrophysical properties using the ML models. Twenty numbers of wells have been used to train and build the model while ten wells were kept blind to assess the efficacy of the trained ML models. Feature selection (selecting inputs which contribute most to the predicted output) and feature extraction (composite new input from other inputs) play very vital role in model's performance.The high correlation of the neutron-density separation with volume of clay has been observed as featured log for clay computation. The ML models were trained with taking five input logs i.e. gamma ray, deep resistivity, neutron porosity, density and neutron-density separation. Once both the ML models were trained, they were used to predict the petrophysical properties in the blind wells. The predicted results using ML models have been compared with earlier processed results through petrophysical evaluation. A high correlation coefficient of more than 90% have been achieved between the predicted and conventional processed petrophysical results for all the blind wells and were analysed on the paralog as well. The predicted petrophysical properties were validated with the testing results in the blind wells.

The present study has brought out unconventional method through development of two machine learning models (k-NN & ANN) which successfully predicted petrophysical properties that further validated with testing results. The developed workflow may be applied for the prediction of petrophysical properties in another area which will save time and increase efficiency. Predicted petrophysical properties can be used in GCM modelling, G&G studies viz. Geomechanics, Seismic inversion etc.

1. Introduction

Reservoir characterization and formation evaluation is one of the most important stages in developing oil and gas field for an accurate and optimal field development strategy. In recent years, advancement in machine learning has raised its capabilities across various sectors and disciplines including the oil industry. In the present paper, a novel approach of predicting the petrophysical properties i.e. effective porosity ($ø_e$), effective water saturation (S_{we}) & volume of clay (V_{cl}) using machine learning (ML) algorithms were carried out. Machine learning algorithms i.e. k-Nearest neighbor (k-NN) and artificial neural network (ANN) have been successfully deployed to achieve the objective of the study.

The total 30 numbers of studied wells is shown in Fig-1. The studied formation is composed of sand/sandstone with minor claystone and deposited in middle shelf environment. The sands in the study area are multi-layered oil & gas. Out of 30 wells, 20 wells were chosen to train ML models and 10 wells were kept blind to assess the ML models. While building the ML models from 20 wells, it was ensured that representative samples covering the variations in the formation are present within the training wells in order to improve the accuracy of the log prediction.

2. Machine Learning (ML) Algorithms

Machine learning is a branch of artificial intelligence that allows computer systems to learn directly from examples, data and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules. Machine learning is a subfield of artificial intelligence (AI) is to understand the structure of data and fit that data into models. Machine learning methods are divided into three categories (1) supervised learning (2) unsupervised learning and (3) Reinforcement learning.

Supervised learning: a set of target output is used to train the network by adjusting the weights and biases of the network to get an output that is closer to the target output.

Reinforcement learning: No target output is provided, instead, the network is graded for the performance of its algorithm.

Unsupervised learning: The weights and biases are adjusted according to the response to the input only using clustering operations. The applied supervised learning models have been described below:

- **K-Nearest Neighbors (k-NN)**

K-nearest neighbors (k-NN) is a supervised learning algorithm that can be used to solve both classification and regression tasks. The main idea behind k-NN is that the value or class of a data point is determined by the data points around it. K-NN classifier determines the class of a data point by majority voting principle. Prediction is done according to the majority class. Similarly, k-NN regression takes the mean value of k closest points. It is very important to determine an optimal k value. If k is too low, the model is too specific and not generalized well. It also tends to be sensitive to noise. On the other hand, if k is too large, the model is too generalized and not a good predictor on both train and test datasets.

- **Artificial Neural Network (ANN)**

Artificial neural network is usually a computational network based on biological neural networks that construct the structure of the human brain. An artificial neural network is a system of several simple processing units known as nodes, neurons, or processing elements. These processing elements are associated with one another through simple connections known as synaptic connections. The strength of the synaptic connections changes with attaching a weight to them. Neurons in a network are organized in layers; each layer is responsible for a particular task.

Typically, there are three kinds of layers in an artificial neural network. Input layer is responsible for presenting the network with the necessary information from the outside world in a normalized manner. Hidden layers (there may be more than one hidden layer in a network, a problem-dependent factor) contain neurons that are responsible for the main part of the input to the output mapping. Output layer contains output neurons that communicate the outcome of the neural network computation with the user.

3. Workflow / Methodology

Fig-1 shows the base map of the area highlighting the training wells (black dots) and testing wells (red dots). The proposed workflow in the present study is consisting of data collection, data pre-processing, building datasets, feature selection, model training and refinement, evaluation and deployment to production. The workflow for predicting petrophysical properties (PHIE, SWE & VCL) is shown in Fig-2. The steps are briefly discussed below:-

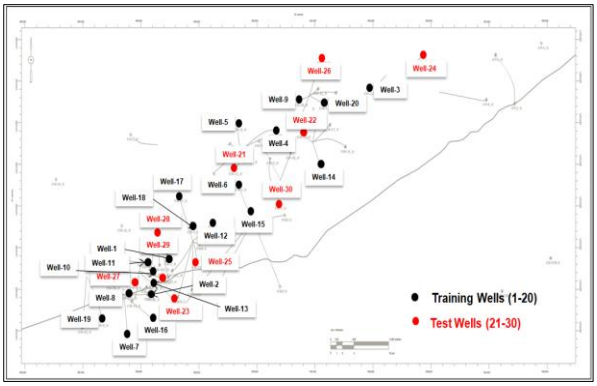

Fig-1: Base Map of study area

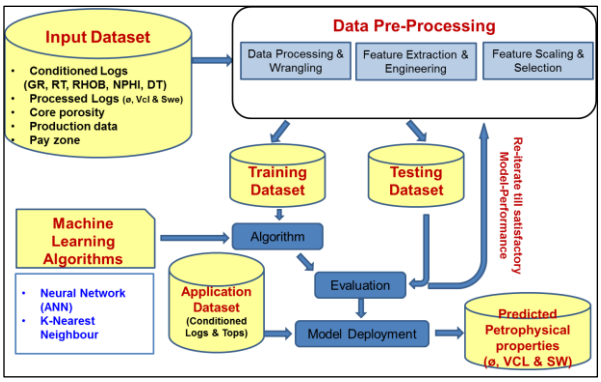

Fig-2: Machine Learning Workflow adopted for Predicting Petrophysical Properties (PHIE, SWE & VCL)

(I) Gathering ML data:
Gathering the well log data was one of the most important stages of machine learning workflows. Since the ML model mimics the training dataset, hence it became essential to include large number of reliable, good data and unbiased training samples of study area.

(II) Data Pre-Processing:
Data Pre-processing of the log data involved depth shifting, applying environmental correction, removing spikes or erroneous log data and synthetic log generation to condition the affected well-log data.

(III) Building Datasets
The conditioned well log data is divided into three datasets—training, validating, and testing.
Training wells: This dataset comprised of the well logs which were used to initially train the algorithm and teach it how to process information.
Validation wells: This dataset was used to estimate the accuracy of the model on few blind wells. This dataset was also used to fine-tune model parameters.
Testing wells: This dataset was used to assess the accuracy and performance of the models. This set was meant to expose any issues in the model.

(IV) Feature Selection
The features (inputs) for the ML algorithm were carefully chosen with the knowledge of correlation between input and output logs. The high correlation between neutron-density separations with volume of clay is a well-known fact. Hence five input logs viz. gamma ray, deep resistivity, neutron porosity, density and neutron-density separation were chosen to build the ML model for predicting petrophysical properties.

(V) Training and Refinement
The two ML model (k-NN and ANN) was deployed in the current study. Once training was complete, both models were refined using the error on validation dataset. This process suggested to discard compressional slowness log for the purpose of petrophysical property prediction. The optimum value of k was observed to be 10 in prediction of petrophysical. Euclidean distance was found to perform well for k-NN model. The data was normalized using the data range for k-NN model. For ANN, 3 hidden layers with 5 nodes each were used to predict the output log.

(VI) Machine Learning Evaluation
The ML models were used to predict the petrophysical parameter after refinement and training of both the models. The outputs from both the model were then compared and analyzed with the available petrophysical properties from conventional petrophysical evaluation.

4. Prediction of Petrophysical Properties (PHIE, SWE & VCL) using ML

In the present study, 30 numbers of wells was selected for predicting the petrophysical property using machine learning models. Out of 30 wells, 20 wells were chosen to train ML models and 10 wells were kept blind to assess the ML models. While building the ML models from 20 wells, it was ensured that representative samples covering the variations in the formation are present within the training wells in order to improve the accuracy of the log prediction.

The data samples used for training (from 20 wells) and its statistics is shown in Fig-3. The model logs (features/inputs) used was gamma ray, deep resistivity, neutron porosity, density and neutron-density separation. The associated logs (target/output) were ø, Swe & Vcl.

TRAINING DATA STATISTICS

Number of Samples	127589
Number of Model Logs	5
Number of Associated Logs	3

MODEL LOG STATISTICS:

LOG	UNITS	MINIMUM	MAXIMUM	MEAN	STD.DEV.
DIFFND	V/V	-0.3837	0.6556	0.1282	0.1419
GR	GAPI	17.9628	340.9458	69.1817	22.6233
LLD	OHMM	0.3301	618.5278	2.9295	13.8522
NPHI	V/V	0.0125	0.7851	0.3224	0.1059
RHOB	G/C3	1.9420	2.9363	2.3297	0.1271

ASSOCIATED LOG STATISTICS:

LOG	UNITS	MINIMUM	MAXIMUM	MEAN	STD.DEV.
PHIE	V/V	0.0011	0.3357	0.1322	0.0882
VOL_WCS	V/V	0.0000	1.0000	0.3992	0.2971
SWE	V/V	0.0686	1.0634	0.9766	0.1179

Fig. 3: Training Data Statistics

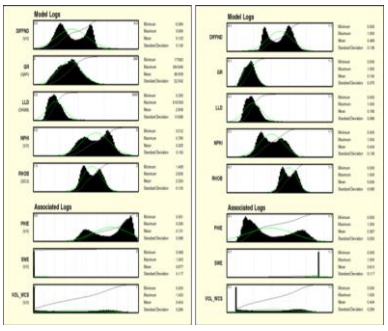

Fig. 4: Histogram of Features (Input logs), Target (Output logs) pre (left) and post (right) normalization

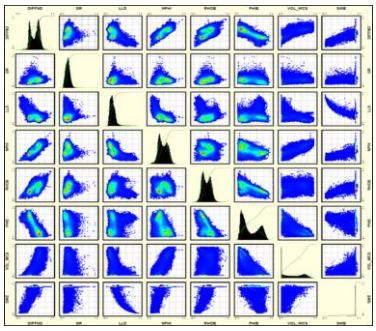

Fig. 5: Paiplot of the training samples indicating correlativity of output logs with input logs

In order to assign same order of weightage to each dataset it is necessary to normalize the input logs to avoid biasness. This may be achieved through normalization using data range (min-max), plot limits, standard deviation or histogram approach. The input for k-NN model was normalized using data range prior to being fed into the model. Fig-4 shows pre and post normalization data distribution of the input and output logs where the rescaled data ranges are from 0-1. A pair-plot of the input and output logs is shown in Fig-5. The pairplot shows the dependence of the output logs with each input logs. It clearly illustrates the high dependence of the effective porosity and volume of clay with neutron-density separation.
The k-NN method uses the membership value to weigh the associated log (output) values of the nearest points. The predicted log value is defined as the summation of the weight associated log values. Multiple predictions with varied k helped in optimizing its value which was 10.

While using the ANN model, a feed forward network was used. The network architecture was same for predicting effective porosity, effective water saturation and volume of clay. The three no of hidden layer were used with 5 neuron in each layer (Fig-6). The error convergence curve (Fig-7) shows the error magnitude with number of cycles for PHIE. The no of cycles was kept as 135, 305 and 188 for effective porosity, effective water saturation and volume of clay respectively during modelling.

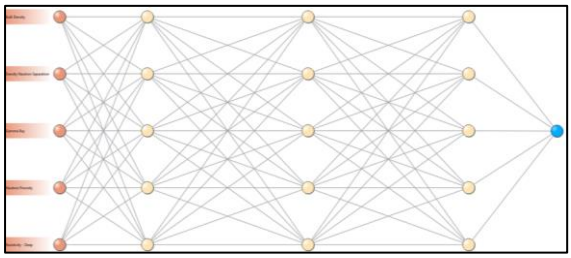

Fig-6: ANN architecture for predicting ø, Swe & Vcl (5 input log, 3 hidden layers with 5 node each and 1 output log)

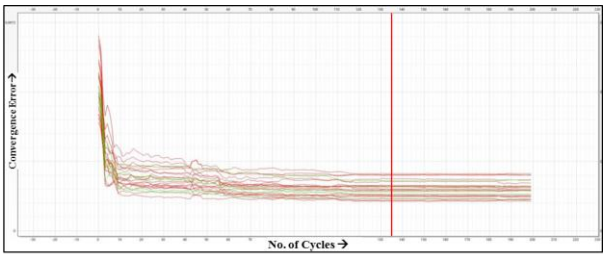

Fig-7: Convergence error to optimize no. of cycles for ANN network (for output log PHIE)

The error convergence graph serves as a guide to optimize no of cycle for ANN. The optimum no. of cycles bring the error curve low for both training and validation dataset. Once both k-NN and ANN model were trained, they were used to predict ø, Swe & Vcl. The cross plot between conventional petrophysical evaluation and output logs predicted from k-NN ML model shows high correlation coefficient for Swe, PHIE and Vcl (Fig-8). Correlation coefficient for SWE is ~94%, for PHIE is ~95% and for VCL is ~92% for all 10 blind wells together.

Similarly, the cross plot between conventional petrophysical evaluation and output logs predicted from ANN ML model shows high correlation coefficient for Swe, PHIE and Vcl (Fig-9). Correlation coefficient for SWE is ~95%, for PHIE is ~95% and for VCL is ~93% for all 10 blind wells together. Both k-NN and ANN models shows quite high correlation coefficient value on the cross plot and were indicative of a robust and well-trained ML model.

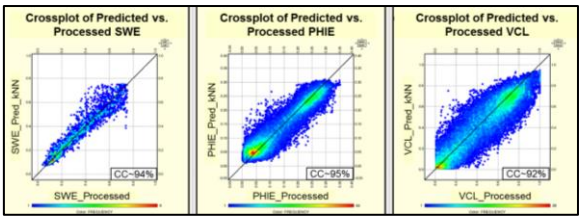

Fig-8: X-plot showing correlation coefficient of predicted Swe (left), ø (mid) & Vcl (right) from k-NN with conventional petrophysical evaluation

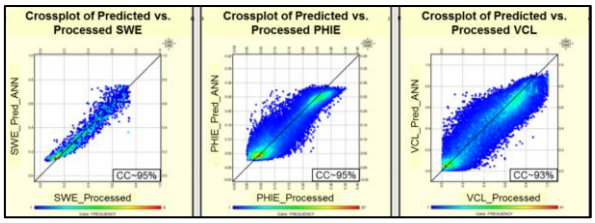

Fig-9: X-plot showing correlation coefficient of predicted Swe (left), ø (mid) & Vcl (right) from ANN with conventional petrophysical evaluation

For all the blind wells the predicted porosity and saturation results shows good to excellent match with the effective porosity and saturation from conventional petrophysical evaluation. However, in case of volume of clay the predicted results from both ML models are predicted relatively lesser than the volume of clay from conventional petrophysical evaluation for shale volumes more than 60%. This may be attributed to the variation in the end parameters of clay used in multi-mineral models of 20 training wells based on the log motif.

Two such typical paralog for wells: 26 & 28 are shown in Fig-10 & Fig-11 respectively. Interval 2250-2252m was tested as object-I in well: 28 (Fig. 10) which flowed gas@31,200m3/d. The predicted petrophysical parameters are satisfactory against the perforated object in Well-28. The machine learning model successfully predicted the petrophysical properties even in high gamma ray sands of studied formation (Fig-11, well-26).

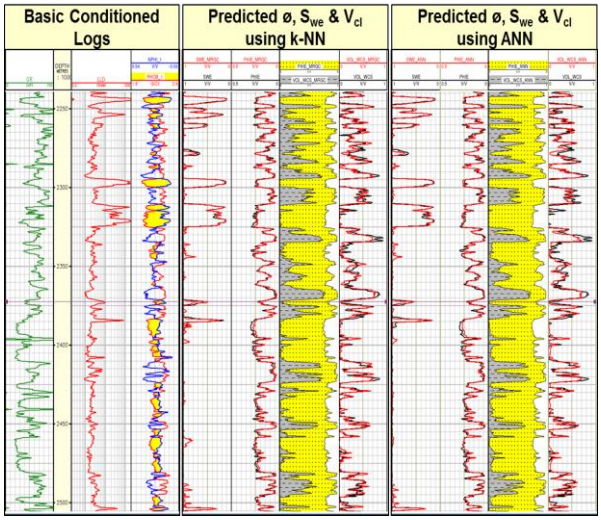

Fig-10: Well-28, Prediction of Petrophysical property in multi-layer sands using k-NN & ANN model

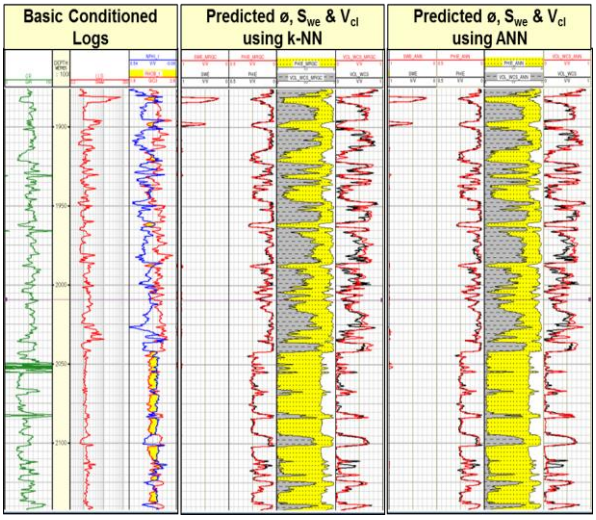

Fig-11: Well-26, ML models successfully predicted the petrophysical properties against high gamma ray in the studied formation

The excellent match of the predicted output with the conventional petrophysical evaluation establishes the fact that the training samples (from 20 wells) had enough representative samples to address a new well. The variation in shaliness, porosity and saturation is well captured in both k-NN & ANN models. In addition to this, both k-NN and ANN models predicted a reasonably satisfactory output against the high gamma ray sand of the studied formation (Fig-11) with high correlation coefficient for porosity, water saturation and volume of shale.

5. Discussions & Results

The effectiveness of machine learning models was assessed by predicting and comparing the petrophysical properties in the study area. Two machine learning models (k-NN and ANN) very well established the fact that ML algorithm has the potential to unfold the complex non-linear relationship in the data. While predicting the petrophysical properties (ø, Swe & Vcl) the models showed high level of correlativity with the outputs from conventional petrophysical evaluation. The study area has brine, hydrocarbon bearing sands and shales layers with it. Petrophysical properties against all such layers were very well addressed by both the models. Fig. 12 shows high correlation coefficient between the predicted petrophysical properties i.e. ø, Swe & Vcl and conventionally evaluated petrophysical properties for each individual blind well.

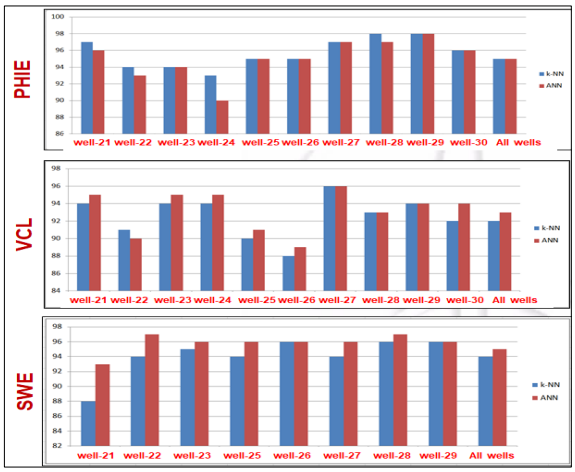

Fig-12: Correlation coefficient of ø, S_{we} & V_{cl} between predicted and processed logs for each blind well

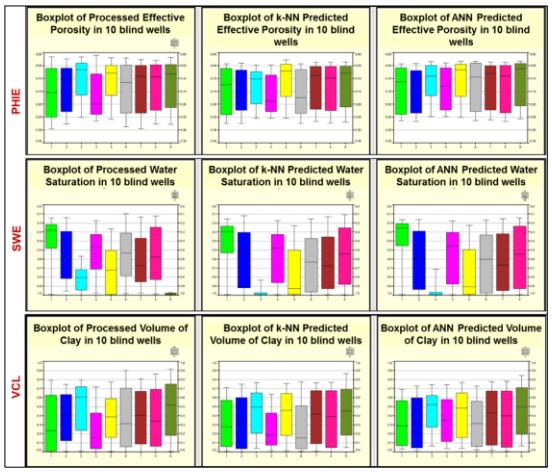

Fig-13: Boxplot of processed and predicted (k-NN and ANN model) ø, S_{we} & V_{cl}

Dataset is the key factor for any ML algorithm to perform well. The high number of training samples i.e. more than 1.25 lakhs samples in present study, established the presence of representative samples for all the facies in the training dataset. Fig 13 shows the boxplot indicating the range of recorded and predicted ø, S_{we} & V_{cl}.. The boxplot shows the data distribution of output predicted from both the ML models lies in the same range as the processed petrophysical properties.

6. Conclusions

- The present paper has brought out unconventional method for prediction of petrophysical properties (ø, Swe & Vcl) from basic logs using Machine Learning algorithms without conventional log data processing in the study area.
- The present study have developed two machine learning models (k-NN & ANN) helped to predict the petrophysical properties successfully with correlation coefficient more than 92% in volume of clay, 95% in porosity & 94% in water saturation in the blind wells.
- Feature selection (selecting inputs which contribute most to the predicted output) and feature extraction (composite new input from other inputs) play very vital role in ML model's performance.
- The developed models predicted the petrophysical properties which were successfully used in GCM modeling/reservoir characterization. The predicted petrophysical properties have been also validated with production testing results in all the wells.
- The present methodology may be applied for the prediction of petrophysical properties in another area which will save time and increase efficiency. Predicted petrophysical properties can be used in GCM modelling, G&G studies viz. Geomechanics, Seismic inversion etc.

Acknowledgement

The authors are thankful to Director(Exploration), ONGC for granting permission to publish this paper in 5th South Asian Geosciences conference and exhibition, GEOINDIA-2022, 14-16th, Oct, 2022 at Jaipur, Rajasthan, India. The authors express their sincere gratitude to Shri Rajesh Kumar Sharma, ED-Asset Manager, Assam Asset, ONGC, Nazira, Sivsagar for providing all facilities and continuous guidance during the study. The authors are also thankful to everyone who involved directly or indirectly to the complete the study.

NB. The views expressed in this paper are solely of the authors and do not necessarily reflect the view of ONGC.

References

1. Hamada Ghareb M, Elsakka Ahmed and Nyein Chaw Y, 2018, “Artificial Neural Network (ANN) Prediction of Porosity and Water Saturation of Shaly Sandstone Reservoirs”, Advances in Applied Science Research, 9(2), 26-31.

2. Chicheng Xu, Siddharth Misra and Shouxiang Ma, 2019, “When Petrophysics Meets Big Data: What can Machine Do?”, SPE-195068-MS.

3. D. Zhang, Y. Chen, and J. Meng, 2018, “Synthetic well logs generation via recurrent neural networks,” Petroleum Exploration and Development, vol. 45, no. 4, pp. 629–639.

4. Mohamed Bennis and Carlos Torres-Verdín, 2019, “Estimation of dynamic petrophysical properties from Multiple well logs using machine learning and Unsupervised rock classification”, SPWLA 60th Annual Logging Symposium, June 15-19.

5. M. Korjani, A. S. Popa, E. Grijalva, S. Cassidy, and I. Ershaghi, 2016, “Reservoir characterization using fuzzy kriging and deep learning neural networks,” in SPE Annual Technical Conference and Exhibition, Dubai, UAE.

6. M. Korjani, A. S. Popa, E. Grijalva, S. Cassidy, and I. Ershaghi, 2016, “A new approach to reservoir characterization using deep learning neural networks,” in SPE Western Regional Meeting, Anchorage, AK, USA,

7. O. Akinnikawe, S. Lyne, and J. Roberts, 2018, “Synthetic well log generation using machine learning techniques,” in Proceedings of the 6th Unconventional Resources Technology Conference, Houston, TX, USA.