Data-Driven Approaches to Predict Thermal Maturity Indices of Organic Matter Using Artificial Neural Networks
Zeeshan Tariq, Mohamed Mahmoud,* Mohamed Abouelresh, and Abdulazeez Abdulraheem

ABSTRACT: Prediction of thermal maturity index parameters in organic shales plays a critical role in defining the hydrocarbon prospect and proper economic evaluation of the field. Hydrocarbon potential in shales is evaluated using the percentage of organic indices such as total organic carbon (TOC), thermal maturity temperature, source potentials, and hydrogen and oxygen indices. Direct measurement of these parameters in the laboratory is the most accurate way to obtain a representative value, but, at the same time, it is very expensive. In the absence of such facilities, other approaches such as analytical solutions and empirical correlations are used to estimate the organic indices in shale. The objective of this study is to develop data-driven machine learning-based models to predict continuous profiles of geochemical logs of organic shale formation. The machine learning models are trained using the petrophysical wireline logs as input and the corresponding laboratory-measured core data as a target for Barnett shale formations. More than 400 log data and the corresponding core data were collected for this purpose. The petrophysical wireline logs are γ-ray, bulk density, neutron porosity, sonic transient time, spontaneous potential, and shallow resistivity logs. The corresponding core data includes the experimental results from the Rock-Eval pyrolysis and Leco TOC measurements. A backpropagation artificial neural network coupled with a particle swarm optimization algorithm was used in this work. In addition to the development of optimized PSO-ANN models, explicit empirical correlations are also extracted from the fine-tuned weights and biases of the optimized models. The proposed models work with a higher accuracy within the range of the data set on which the models are trained. The proposed models can give real-time quantification of the organic matter maturity that can be linked with the real-time drilling operations and help identify the hotspots of mature organic matter in the drilled section.

1. INTRODUCTION
The depletion and fall of conventional oil and gas resources will lead to a shortage in the supply of the world’s energy needs. Therefore, unconventional resources, in particular shale gas, are gaining popularity in the recent era of oil and gas. Organic-rich shale is one of the most vital sources of unconventional oil and gas. The appropriate geochemical characterization of shale resources plays a critical role in defining the prospect and development of the economic model of the field. For example, the production from Barnett Shale is controlled mainly by the thermal maturity, total organic carbon (TOC), and thickness of the shale target. The geochemical analysis aims to evaluate the organic richness, thermal maturity, and hydrocarbon potential of the organic-rich shale. The wettability and pore structure of the organic-rich shale are also affected by geochemical parameters.

The most accurate estimation of the organic richness and thermal maturity of shale helps in reducing the risk carried by petroleum well drilling. Mineral heterogeneity, complex lithology, and natural fractures in shales have brought great challenges to the accurate estimation of the organic content. Core measurements and well logs are the main ways to obtain geochemical parameters. Direct measurement of organic richness in the laboratory on the core samples is the most accurate way to obtain thermal maturity parameters. However, retrieving core samples from each well of every field and carrying out laboratory experiments on them is quite a time-consuming and costly method. Consequently, core-based geochemical data is very scarce. On the other hand, well log data is a prime component of all well drilling plans and hence is readily available. Limited core sample data and the associated well logs are used to develop correlations that can be applied to the whole well. Therefore, empirical correlations and machine learning-based models are used to obtain these parameters indirectly.
from well logs (need reference). The accuracy and applicability of these models depend on the data set and the region from where the data set is collected (need reference).

Machine learning (ML) and artificial intelligence (AI) both are the captivating fields that integrate computational power with human intelligence to produce smart and reliable solutions of extremely nonlinear and highly complicated problems.10 In the past two decades, engineering journals have reported numerous articles utilizing AI and ML for regression, function approximation, and classification problems.11−13 With the advent of soft computing techniques, several correlations utilizing techniques from the field of AI have come to the fore, especially in reservoir characterization,14−16 reservoir engineering,17−20 and reservoir geomechanics.21,22 In petroleum geochemistry, such correlations can be seen in the works of Rahaman.23 In recent years, support vector machine (SVM) and least-squares support vector machine techniques were actively used by researchers to predict total organic carbon (TOC).24,25 Researchers have used artificial neural networks too to predict TOC and thermal maturity.26−28

Based on the literature survey, it was observed that much attention was given to predict only the TOC content of the organic shales. However, organic matter geochemical analysis of shale gas formations requires estimation of a suite of parameters such as $T_{\text{max}}$, $S_1$, $S_2$, $S_3$, and TOC. Therefore, the objective of this study is to explore the potential of the machine learning technique in predicting these five geochemical parameters ($T_{\text{max}}$, $S_1$, $S_2$, $S_3$, and TOC) for the Barnett Shale. This study has utilized both machine learning and evolutionary algorithms to arrive at the optimum model. In addition to that, five explicit empirical correlations derived from the PSO-ANN algorithm are proposed; these correlations do not require any ML-based software for the execution.

1.1. Background. Geochemical properties of shale such as thermal maturity, source potentials ($S_1−S_3$), total organic carbon content (TOC), hydrogen index (HI), and oxygen index (OI) are important parameters to evaluate its production potential.26,27 $S_1$ and $S_2$ are called as volatile hydrocarbon and remaining hydrocarbon or oil potential, respectively.30 The HI and OI are calculated from TOC and source potentials. Maximum temperature ($T_{\text{max}}$) is a chemical indicator of thermal maturity. $T_{\text{max}}$ is the temperature at which the $S_3$ attains its maximum hydrocarbon generation.31 It accounts for the hydrogen and oxygen richness of the shale. On the other hand, TOC is an important indicator of organic richness in shale plays. TOC is expressed in weight percentage (wt %). Usually, if TOC is less than 0.5 wt %, it means no organic matter exists in the shale rock. A wt % of TOC greater than 0.5 is a positive sign for the existence of organic matter in shale rock. Several cross plots such as between $T_{\text{max}}$ and hydrogen index (HI), $S_3$ and TOC, and HI and OI are used to evaluate the level of thermal maturity and kerogen type.32−34 Table 1 presents the range of TOC and $T_{\text{max}}$ according to the maturity level.

| parameter | no hydrocarbon | maturity range | postmaturity range | refs |
|-----------|----------------|----------------|--------------------|------|
| TOC, wt % | less than 0.5  | 0.5−2          | greater than 2     | 35, 36|
| $T_{\text{max}}$, °C | less than 435 | 435−465         | greater than 465   | 33, 37|

Single well log method38 and a composite well log method39 TOC can be measured directly in the laboratory in different ways such as filter acidification,40 nonfilter acidification,40 total minus coulometric,41 Rock-Eval,42 laser-induced pyrolysis,43 and diffuse reflectance infrared Fourier transformation spectroscopy (DRIFTS).34 Indirect methods involve the utilization of petrophysical well logs and seismic data. A large number of models are reported in the literature for the prediction of geochemical parameters using composite well logs.

Schmoker35 established the first correlation to predict TOC for Devonian shale formation. Schmoker correlation is expressed in eq 1, which gives results in volume percentage

$$\text{TOC (vol %)} = \frac{\rho_{\text{fl}} - \rho}{1.378}$$

(1)

Schmoker35 modified his correlation for Bakken formation as given by eq 2.

$$\text{TOC (wt %)} = \frac{(100\rho_f)(\rho_\text{fl} - 0.9922\rho_{\text{mat}} - 0.039)}{(R_\text{p})(\rho_\text{fl} - 1.135\rho_{\text{mat}} - 0.675)}$$

(2)

where $\rho_f$ is the organic matter density in g/cm$^3$, $\rho_{\text{mat}}$ is the average density of grain and pore fluid in g/cm$^3$, $R_p$ is the ratio of the organic matter to organic carbon in weight percentage.

Passey et al.35 suggested an easy-to-use model for TOC prediction, as summarized in eqs 3 and 4. Currently, this model is widely used for evaluating the unconventional resources reservoir.

$$\Delta \log R = \log_{10}(R/R_{\text{baseline}}) + 0.02 \times (\Delta t - \Delta t_{\text{baseline}})$$

(3)

$$\text{TOC} = \Delta \log R \times 10^{(2.297−0.1677\times\text{LOM})}$$

(4)

where $R_{\text{baseline}}$ and $R$ are the base formation and evaluated formation resistivities in $\Omega \cdot \text{m}$, respectively, $\Delta \log R$ represents the log separation, $\Delta t_{\text{baseline}}$ and $\Delta t$ are the base formation and evaluated formation sonic transit times both in $\mu\text{s}/\text{ft}$, respectively, and LOM represents the formation level of maturity. Sultan28 used a self-adaptive differential evolution algorithm to optimize artificial neural networks (ANN) and presented the empirical correlation to predict TOC of the Barnett shale. His correlation for TOC prediction is given by eqs 5−8

$$\text{TOC} = 0.4127 \times (\text{TOC}_n − 0.7) − 1$$

(5)

$$\text{TOC}_n = \sum_{j=1}^{N_j} w_j \frac{1}{1 + e^{−Y_j}} + b_j$$

(6)

$$Y_j = \sum_{j=1}^{N_j} w_{2j} \frac{1}{1 + e^{−X_j}} + b_{2j}$$

(7)

$$X_j = \frac{1}{1 + e^{−(n_{1j}G_{R_n} + n_{3j} D_{R_n} + n_{4j} D_{E_n} + n_{5j} R_{HOB} + b_{1j})}}$$

(8)

Table 2 presents some of the recent research works related to the prediction of organic matter in shale using machine learning and nonlinear regression approaches for the relevant geological fields.

2. RESULTS AND DISCUSSION

This section demonstrates the combined results for the prediction of five thermal maturity parameters such as $T_{\text{max}}$.
Table 2. Summary of the Research Related to the Prediction of Organic Matter in Shale

| refs          | study conducted | technique               | method type                          | input parameters\(^a\)                     | geological field study |
|---------------|-----------------|-------------------------|--------------------------------------|---------------------------------------------|------------------------|
| Tan et al.\(^{24}\) | prediction of TOC | artificial intelligence | epsilon-SVR, nu-SCR, SMO-SVR, and RBF | CNL, GR, AC, K, TH, U, PE, RHOB, and RT     | Huangping syncline, China |
| Rui et al.\(^{34}\)   | prediction of TOC | artificial intelligence | SVM                                 | wireline log data such as RHOB, GR, SP, RT, and DT | Beibu Gulf basin |
| Lawal et al.\(^{27}\) | prediction of TOC | artificial intelligence | ANN                                  | XRD data: SiO\(_2\), Al\(_2\)O\(_3\), MgO, and CaO | Devonian Shale |
| Sultan\(^{56}\)     | prediction of TOC | artificial intelligence | self adaptive differential evolution-based ANN | well logs: GR, DT, RT, and RHOB | Devonian Shale |
| Mahmoud\(^{16}\)    | prediction of TOC | artificial intelligence | ANN                                  | well logs: GR, DT, RT, and RHOB | Devonian Shale |
| Zhao et al.\(^{50}\) | prediction of TOC | regression              | nonlinear                            | CNL                                          | Ordos Basin in China and Bakken Shale of North Dakota |
| Wang et al.\(^{53}\) | prediction of TOC | regression              | nonlinear                            | DT and RT                                    | Sichuan Basin, Southern China |
| Alizadeh et al.\(^{46}\) | prediction of TOC and \(S_1\) | artificial intelligence | ANN                                  | DT and RT                                    | Dezful Embayment, Iran |
| Handhal et al.\(^{45}\) | prediction of TOC | artificial intelligence | SVR, ANN, KNN, random forest, and rotation forest | GR, RHOB, NPHI, RILD, and DT | Rumaila Oil Field, Iran |
| Wang et al.\(^{48}\) | prediction of TOC | artificial intelligence | ANN                                  | RHOB, NPHI, RT, and DT                        | Bohai Bay Basin, China |

\(^{a}\)GR = \(\gamma\)-ray, RHOB = bulk density, LLD = deep lateral log, LLS = shallow lateral log, MSFL = microspherical focused log, \(R_{\text{ILD}}\) = 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, SiO\(_2\) = silicon dioxide, Al\(_2\)O\(_3\) = aluminium dioxide, MgO = magnesium oxide, and CaO = calcium oxide.

Table 3. Step-by-Step Pseudocode for the Proposed PSO-ANN Algorithm for Thermal Maturity Parameter Prediction

| steps | working |
|-------|---------|
| 1     | start |
| 2     | set input variables |
| 3     | initialize parameters of ANN such as learning rate, activation functions, etc. |
| 4     | vary the number of hidden layers (sensitivity of hidden layers, 1–3) |
| 5     | vary the number of neurons in the hidden layer (sensitivity of neurons, 5–30) |
| 6     | select the learning rate \([0, 1]\) for the selected learning algorithm |
| 7     | train and test the ANN model |
| 8     | evaluate the objective function for a minimum convergence value |
| 9     | extract weights and biases from the trained model |
| 10    | initialize parameters of PSO algorithm such as the number of iterations, population of particles, cognitive and social accelerations, and initial and final inertia weights |
| 11    | set range for sample search space of each extracted weights and biases |
| 12    | feed extracted weights and biases in a PSO algorithm as the initial population |
| 13    | evaluate the objective function for a minimum convergence value |
| 14    | run the iterative process until the stopping criterion\(^b\) is achieved |
| 15    | pick the global best solution |
| 16    | set optimum weights and biases from the globally best model in the network for the prediction of thermal maturity parameters |
| 17    | end |

\(^{b}\)stopping criterion = a maximum number of iterations are attained or a maximum level of inactivity is reached.

For a \(T_{\text{max}}\) model, a total of 400 data points were obtained. On a training data set, the ANN model predicted the \(T_{\text{max}}\) with an \(R^2\) of 0.917, an average absolute percentage error (AAEPE) of 1.006%, and a root-mean-square error (RMSE) of 0.528. When testing data set, the ANN model predicted the \(T_{\text{max}}\) with an \(R^2\) of 0.918, an AAEPE of 1.137%, and an RMSE of 0.428. The training and testing scatter plots are shown in Figure 4. The learning algorithm utilized was the LM with a learning rate of 0.15. With this combination, an optimum model was stored, and
their weights and biases were extracted. The mathematical model for $T_{\text{max}}$ utilizing optimum weights and biases is given in Appendix A.

For an $S_1$ model, a total of 400 data points was obtained. On a training data set, the ANN model predicted the $S_1$ with an $R^2$ of 0.839, an AAPE of 13.8%, and an RMSE of 0.006, while on a testing data set, the ANN model predicted the $S_1$ with an $R^2$ of 0.827, an AAPE of 15.538%, and an RMSE of 0.010. The training and testing scatter plots are shown in Figure 5. The learning algorithm utilized was the RB with a learning rate of 0.15. With this combination, an optimum model was stored, and their weights and biases were extracted. The mathematical model for $S_1$ utilizing optimum weights and biases is given in Appendix B.

For an $S_2$ model, a total of 380 data points was obtained. On a training data set, the ANN model predicted the $S_2$ with an $R^2$ of 0.839, an AAPE of 13.8%, and an RMSE of 0.006, while on a testing data set, the ANN model predicted the $S_2$ with an $R^2$ of 0.827, an AAPE of 15.538%, and an RMSE of 0.010. The training and testing scatter plots are shown in Figure 6. The learning algorithm utilized was the RB with the learning rate of 0.15. With this combination, the optimum model was stored, and their weights and biases were extracted. The mathematical model for $S_2$ utilizing optimum weights and biases is given in Appendix C.

For an $S_3$ model, a total of 450 data points was obtained. On a training data set, the ANN model predicted the $S_3$ with an $R^2$ of 0.891, an AAPE of 5.4%, and an RMSE of 0.001, while on a testing data set, the ANN model predicted the $S_3$ with an $R^2$ of 0.868, an AAPE of 6.11%, and an RMSE of 0.010. The training and testing scatter plots are shown in Figure 7. The learning algorithm utilized was the LM with the learning rate of 0.15. With this combination, the optimum model was stored, and their weights and biases were extracted. The mathematical model for $S_3$ utilizing optimum weights and biases is given in Appendix D.

For a TOC model, a total of 360 data points were obtained. On a training data set, the ANN model predicted the TOC with an $R^2$ of 0.825, an AAPE of 7.863%, and an RMSE of 0.029, while on a testing data set, the ANN model predicted the TOC with an $R^2$ of 0.826, an AAPE of 8.713%, and an RMSE of 0.048. The training and testing scatter plots are shown in Figure 8. The learning algorithm utilized was GD with a learning rate of 0.15. With this combination, the optimum model was stored, and their weights and biases were extracted. The mathematical model for TOC utilizing optimum weights and biases is given in Appendix E.

To see the improvement in accuracy of the models using the proposed PSO-ANN-based algorithm, a comparison was made between conventional and PSO-ANN by comparing the $R^2$ obtained on overall data sets (training and testing). Figure 9
shows the bar chart that illustrates that in all five models (T\textsubscript{max}, S\textsubscript{1}, S\textsubscript{2}, S\textsubscript{3}, and TOC) the R\textsuperscript{2} values obtained using the PSO-ANN algorithm were much higher than the conventional ANN model. This proves that the proposed PSO-ANN algorithm-based
models have much higher accuracy than the models based on the conventional ANN technique.

3. CONCLUSIONS

A good estimation of shale geochemical properties requires a sophisticated approach. Minor variations in anticipated results lead to wastage of man-hours and huge investments. On the other hand, a small improvement in the estimation practices can improve the worth of the exploration project manifold. The development of robust and improved models for prediction of thermal maturity of organic shale was the focus of this study. To achieve the objective, an ANN tool coupled with a PSO algorithm is employed in this work. The evaluation of the proposed models was based on various statistical measures such as RMSE, AAPE, MAE, and R². A step-by-step comprehensive analysis to reach the optimum model selection along with the statistical and graphical metrics was also presented in this study. The generalization capability of the proposed models was tested using a blind data set. By correlating the predicted maturity index with the core-based one, the proposed models were found to be effective, faster, and more readily available than lab analysis. The proposed models are completely reproducible. The proposed PSO-ANN-based models can give reliable predictions in the absence of experimental data and therefore can be a good choice to be included in any software package for a complete analysis of geochemical data without going to the laboratory for carrying out the Rock-Eval pyrolysis experiment. The generalization capability of the proposed models was tested using a blind data set. By correlating the predicted maturity index with the core-based one, the proposed models were found to be effective, faster, and more readily available than lab analysis. The proposed models are completely reproducible. The proposed PSO-ANN-based models can give reliable predictions in the absence of experimental data and therefore can be a good choice to be included in any software package for a complete analysis of geochemical data without going to the laboratory for carrying out the Rock-Eval pyrolysis experiment.

4. MATERIALS AND METHODS

4.1. Studied Geological Field. The Mississippian Barnett Shale in Fort Worth Basin, North Texas is a classic world-class unconventional shale gas reservoir.57 It consists mainly of siliceous-rich mudstone interlaminated with clay- and calcareous-rich mudstone and deposited in a low-energy, relatively deep water environment.55,58 The subsurface thickness of Barnett Shale reaches up to about 1000 ft. (304.8 m) in the Newark sub-basin. In the Newark East field, from where the data comes, Barnett Shale is thermally mature, averaging 4–5 wt % total organic carbon TOC, and trapped between two impermeable limestone beds, which is the most favorable conditions for vertical well completion. However, the variation in the thickness, mineral composition, organic richness, and thermal maturity levels through the entire Fort Worth Basin raised the need to reduce the uncertainty of predicted reservoir properties in the rest of the basin and develop AI models that can be applied to another shale gas reservoirs.

4.2. Geochemical Analysis. Pyrolysis is a process of performing thermal decomposition of materials at higher temperatures. The geochemical analysis of an organic matter is comprised of Rock-Eval pyrolysis. Pyrolysis is used to evaluate the thermal maturity and organic richness of the source rock. From the pyrolysis experiment, the quality, quantity type, thermal maturity, hydrogen index, migration index, production index, and oxygen index of organic matter can be determined. Typically, the five parameters Tₘₐₓ, S₁, S₂, Sᵧ, and TOC are measured. S₁ accounts for free hydrocarbon released at 300°C measured in mg HC/g rock. S₂ accounts for hydrocarbon released from the cracking of kerogen at the temperature range between 300 and 600 °C measured in mg HC/g rock. S₃ accounts for carbon dioxide (CO₂) released from the breaking of carboxy groups and other oxygen-containing compounds measured in mg CO₂/g rock. The TOC is measured by oxidizing the residue left in the pyrolysis process at a fixed temperature of 600 °C.5,6 Figure 10 shows the schematic of the different fractions obtained from total organic matter.

![Figure 10](https://dx.doi.org/10.1021/acsomega.0c03751)

4.3. Data Analytics. The statistical description of the data set used to train AI models for the geochemical parameters prediction is given in Table 5. The ranges of the input parameters for each model are quite practically reasonable. The complete data set utilized for the training of each model is given in Figure 11.

The core data of the corresponding conventional wireline well logs are collected. The frequency distribution of the measured Tₘₐₓ, S₁, S₂, Sᵧ, and TOC from the geochemical analysis is shown in Figure 12. The Tₘₐₓ data is evenly distributed over a wide range between 420 and 540 °C. S₁ is mainly distributed between 0.1 and 1. S₂ is mainly distributed between 0.3 and 1.6. About 60% of the cores have an S₃ lower than 1, and only a few have permeability above 1.5. The S₆ data is uniformly distributed over a range between 0.1 and 0.3. The TOC data is mainly distributed between 2 and 6, with only fewer points above 6. The frequency histograms show that the core data values are distributed over a wide range of values and are quite heterogeneous.

4.4. Feature Selection. Feature selection was made by evaluating the relative importance of the input parameters with the output parameter using the Pearson correlation coefficient (CC) criterion, which is given by eq 10

\[
CC = \frac{k \sum xy - (\sum x)(\sum y)}{\sqrt{k(\sum x^2) - (\sum x^2)} \sqrt{k(\sum b^2) - (\sum b^2)}} \tag{10}
\]

where x and y are two variables and k is the sample size. The value of CC lies between −1 and +1. The values near to negative one show an inverse relationship between two variables, the values near to the positive one show a direct relationship between two variables, and the values near to zero show a poor relationship between the pair of two variables. Figure 13 shows the CC of input parameters such as GR, RHOB, NPHI, AT90, Δt, and SP log with the target parameters such as Tₘₐₓ, S₁, S₂, S₃, and TOC.

4.5. Accuracy Metrics. The models were evaluated based on the goodness-of-fit tests such as the average absolute percentage error (AAPE), mean absolute error (MAE), root-
mean-square error (RMSE), and coefficient of determination ($R^2$). The definition of these parameters is given in Table 6.

### 4.6. Machine Learning Method

**Artificial neural network (ANN)** is a machine learning (ML) technique, mostly used for function approximation purposes. It is comprised of a series of layers such as an input layer, a middle layer(s), and an output layer. The middle layer is also called a hidden layer and it can be single or multiple, depending on the training data set.60 The selection of the number of neurons in the middle layers depends on the overall model performance in terms of accuracy.60 A transfer function exists between the input layer and the middle layer, and another transfer function exists between the middle layer and the output layer. Various choices of transfer functions are available such as linear, sigmoidal, radial basis, and rectified linear unit (ReLU) type. The detailed description of the theory and utilization of ANN can be found in our previous publications.10,61,62

**Particle swarm optimization (PSO)** is utilized to optimize the weights and biases of a neural network. In the past, many researchers have found good results by coupling PSO with other AI techniques such as ANN, least-squares support vector machine (LSSVM), and adaptive neuro-fuzzy inference system (ANFIS).63

### Table 5. Ranges of the Data Used for AI Modeling

| models          | GR, API | RHOB (g/cc) | NPHI (vol/vol) | $R_{LLD}$ (Ω•m) | $\Delta t$ (μs/ft) | SP (mV) |
|-----------------|---------|-------------|----------------|------------------|-------------------|---------|
| $T_{max}$ (°C)  | 18.229  | 417.06      | 2.37 2.86      | 0.003 0.33       | 4.8 1283.34       | 45.27   |
| $S_1$ (mg HC/g rock) | 19.120  | 336.73      | 2.37 2.86      | 0.003 0.33       | 43.7 1027.90      | 45.27   |
| $S_2$ (mg HC/g rock) | 18.229  | 372.29      | 2.37 2.86      | 0.003 0.33       | 43.7 1027.90      | 45.27   |
| $S_3$ (mg CO₂/g rock) | 18.229  | 417.06      | 2.37 2.83      | 0.003 0.29       | 4.8 1088.66       | 45.27   |
| TOC (wt %)      | 55.500  | 359.76      | 2.20 2.68      | 0.03 0.33        | 6.0 148.87        | 62.05   |

**Figure 11.** Well logs’ input data (AT90 is a $R_{LLD}$ log).

**Figure 12.** Frequency distribution of $T_{max}$, $S_1$, $S_2$, $S_3$, and TOC.

**Figure 13.** Relative importance of the input parameters such as GR, $\rho$, NPHI, AT90 ($R_{LLD}$), $\Delta t$, and SP logs with the output parameters such as $T_{max}$, $S_1$, $S_2$, $S_3$, and TOC.
velocity, and cognitive parameters. A detailed discussion about the PSO algorithm can be found in the publication of Abido.\textsuperscript{68} Particles velocity term is given by eq \ref{eq:velocity}

\begin{equation}
\dot{v}_i(n + 1) = w\dot{v}_i(n) + \{c_1 \times \text{rand}(0, 1] \times (p_b^i - p(n))\} \\
+ \{c_2 \times \text{rand}(0, 1] \times (p^i_n - p(n))\}
\end{equation}

where \(w\) is the weight of the particle (0 \(\leq w \leq 1.2\), \(v_i\) is the particle velocity, \(c_1\) is the cognitive parameter (0 \(\leq c_1 \leq 1.2\)), \(c_2\) is the cognitive parameter (0 \(\leq c_2 \leq 1.2\)), \(n\) is the number of iteration, \(p_b^i\) is the local best solution of the particle, \(p^i_n\) is the global best solution of the particle, and \(p_i\) is the \(i\)th position of the particle at the \(n\)th iteration. The next position for each candidate solution in the search space is created by summation of the current particle position and particle velocity

\begin{equation}
p(n + 1) = p(n) + v(n)
\end{equation}

\section{APPENDIX A MATHEMATICAL MODEL TO PREDICT T_{\text{MAX}}}

The ANN-based mathematical model to predict the \(T_{\text{MAX}}\) of the organic shale is given by eq \ref{eq:ANN-Tmax}

\begin{equation}
T_{\text{MAX}} = 52 \times T_{\text{MAX}} + 417
\end{equation}

where

\begin{table}[h]
\centering
\caption{Statistical Indicators of Model Performance Evaluation\textsuperscript{a}}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{goodness-of-fit test} & \textbf{mathematical expression} \\
\hline
average absolute percentage error & \(\text{AAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{|Y_{\text{measured},i} - Y_{\text{predicted},i}|}{Y_{\text{measured},i}}\) & (11) \\
\hline
mean absolute error & \(\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_{\text{measured},i} - Y_{\text{predicted},i}|}{Y_{\text{measured},i}}\) & (12) \\
\hline
root-mean-square error & \(\text{RMSE} = \left( \frac{1}{n} \sum_{i=1}^{n} (Y_{\text{measured},i} - Y_{\text{predicted},i})^2 \right)^{1/2}\) & (13) \\
\hline
coefficient of determination & \(R^2 = \frac{k \sum y_i - (\sum x_i)(\sum y_i)}{\sqrt{k \sum (x_i^2) - (\sum x_i)^2} \sqrt{k \sum (y_i^2) - (\sum y_i)^2}}\) & (14) \\
\hline
\end{tabular}
\end{table}

\textsuperscript{a}\(Y_{\text{measured}}\) is the measured value of TOC, \(Y_{\text{predicted}}\) is the estimated value from the model, and \(n\) is the total number of samples.

\begin{table}[h]
\centering
\caption{Weights and Biases of the Proposed Model for T_{\text{MAX}} Prediction}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{hidden layer neurons (N_h)} & \textbf{weights between input and hidden layers (w_i)} & \textbf{weights between hidden and output layers (w_o)} & \textbf{hidden layer bias (h_i)} & \textbf{output layer bias (b_j)} \\
\hline
1 & 1 & & & & & & & \\
2 & 2 & & & & & & & \\
3 & 3 & & & & & & & \\
4 & 4 & & & & & & & \\
5 & 5 & & & & & & & \\
6 & 6 & & & & & & & \\
7 & 7 & & & & & & & \\
8 & 8 & & & & & & & \\
9 & 9 & & & & & & & \\
10 & 10 & & & & & & & \\
\hline
\end{tabular}
\end{table}

\begin{equation}
T_{\text{MAX}} = \sigma_0 \sum_{i=1}^{N_h} \{ w_i \sigma_i (w_{i1} GR_n + w_{i2} \rho_n + w_{i3} SP_n + w_{i4} R_{\text{ILD}}) \\
+ w_{i5} \Delta T_{\text{C}_n} + w_{i6} \Delta T_{\text{C}_n} + \rho_n + b_1 \} + b_2
\end{equation}

where \(\sigma_0(x) = (2/1 + e^{-2x}) - 1; \sigma_i(x) = x_i; w_{i1}, w_{i2}, b_1\) and \(b_2\) are the weights and biases of the \(T_{\text{MAX}}\) model, given in Table 7. \(GR_n\) is the normalized value of a \(GR\) log, \(\rho_n\) is a normalized value of a bulk density, \(SP_n\) is a normalized value of a neutron porosity, \(R_{\text{ILD}}\) is a normalized value of a \(R_{\text{ILD}}\) resistivity log, \(\Delta T_{\text{C}_n}\) is a normalized value of a compressional wave travel time, \(SP_n\) is a normalized value of an SP log. The equations to find \(GR, \rho, SP, R_{\text{ILD}}, \Delta T_{\text{C}_n}\) and \(SP_n\) are given by eqs AA3–AA7.

\begin{equation}
GR_n = 2 \times \left\{ \frac{GR - 18.3}{417 - 18.3} \right\} - 1
\end{equation}

where \(GR\) is the \(GR\) log in API.

\begin{equation}
\rho_n = 2 \times \left\{ \frac{\rho - 2.37}{2.86 - 2.37} \right\} - 1
\end{equation}

where \(\rho\) is the bulk density in g/cc.

\begin{equation}
\phi_n = 2 \times \left\{ \frac{\phi - 0.003}{0.333 - 0.003} \right\} - 1
\end{equation}

where \(\phi\) is the neutron porosity.

\begin{equation}
R_{\text{ILD}} = 2 \times \left\{ \frac{R_{\text{ILD}} - 4.8}{1283 - 4.8} \right\} - 1
\end{equation}
Table 8. Weights and Biases of the proposed Model for $S_1$ Prediction

| hidden layer neurons $N_h$ | GR  | $\rho$ | NPHI | $R_{ILD}$ | $\Delta t_C$ | SP  | weights between hidden and output layers ($w_i$) | hidden layer bias ($b_i$) | output layer bias ($b_2$) |
|---------------------------|-----|--------|------|-----------|-------------|-----|-----------------------------------------------|--------------------------|--------------------------|
| 1                         | -0.0872 | 2.3795 | 0.61 | 0.9346    | -1.0302     | 2.6013 | 2.2032                                      | 2.7471                    | -1.125                   |
| 2                         | 2.9553  | 1.196  | 0.9532 | -0.1497  | 2.0403      | -0.6852 | -1.5668                                     | 3.5566                    |                          |
| 3                         | 0.9799  | -2.0489 | -0.5193 | 1.0599    | 1.1969      | 3.2706 | -2.0522                                     | -0.5307                   |                          |
| 4                         | -0.3019 | 1.5661 | 0.1572 | 0.0954    | -0.0242     | -2.0559 | -3.1289                                     | 1.1359                    |                          |
| 5                         | 1.8638  | -0.8322 | -3.3966 | 0.6946    | 2.8564      | 3.4534 | -1.377                                      | -3.9418                   |                          |
| 6                         | -3.2506 | 0.8794 | 1.5293 | 1.2903    | -0.8682     | 0.4842  | -3.4481                                     | -0.2281                   |                          |
| 7                         | 1.7284  | -1.3916 | 4.3578 | -2.3109   | -5.3719     | 0.012   | -1.0532                                     | 0.6863                    |                          |
| 8                         | -2.3325 | 0.5338 | 1.3951 | -0.8393   | 0.687       | 0.3473  | 3.2165                                      | -2.2594                   |                          |
| 9                         | 0.6438  | 1.1076 | 3.1629 | 0.6633    | 0.8062      | 0.8584  | 1.4581                                      | 3.789                     |                          |
| 10                        | 1.7403  | -1.5715 | 1.5465 | -2.1928   | -0.0194     | -2.249  | -0.7571                                     | 1.6332                    |                          |

where $R_{ILD}$ is the resistivity log in $\Omega\cdot\text{m}$. 

$$
\Delta t_C = 2 \times \left( \frac{\Delta t_C - 45.27}{93.482 - 45.27} \right) - 1
$$

(A7)

where $\Delta t_C$ is the compressional wave travel time in $\mu$s/ft.

$$
SP_n = 2 \times \left( \frac{SP + 154.18}{28.625 + 154.18} \right) - 1
$$

(A8)

where SP is the spontaneous potential log in mV.

**APPENDIX B MATHEMATICAL MODEL TO PREDICT $S_1$**

The ANN-based mathematical model to predict the $S_1$ of the organic shale is given by eq BB1

$$
S_1 = 0.4248 \times S_{n1} + 0.6312
$$

(B1)

$$
S_{in} = \sigma_1 \sum_{i=1}^{N_h} w_{i1} \sigma_1(w_{i1} GR_n + w_{i1} \rho_n + w_{i1} \phi_n + w_{i1} R_{ILD}) + w_{i2} \Delta t_C + w_{i2} SP_n + b_1 + b_2
$$

(B2)

where $\sigma_1(x) = (2 / (1 + e^{-x})) - 1$; $\sigma_2(x) = x$; and $w_{i1}$, $w_{i2}$ $\rho_n$, $\phi_n$, and $b_i$ are the weights and biases of the $S_1$ model, given in Table 8. The equations to find $GR_n$, $\rho_n$, $\phi_n$, $R_{ILD}$, $\Delta t_C$, and SP are given by eqs BB3–BB7.

$$
GR_n = 2 \times \left( \frac{GR - 19}{336.8 - 19} \right) - 1
$$

(B3)

where GR is the $\gamma$-ray log in API.

$$
\rho_n = 2 \times \left( \frac{\rho - 2.37}{2.86 - 2.37} \right) - 1
$$

(B4)

where $\rho$ is the bulk density in g/cc.

$$
\phi_n = 2 \times \left( \frac{\phi - 0.003}{0.333 - 0.003} \right) - 1
$$

(B5)

where $\phi$ is the neutron porosity.

$$
R_{ILD} = 2 \times \left( \frac{R_{ILD} - 1.307}{1028 - 1.307} \right) - 1
$$

(B6)

where $R_{ILD}$ is the resistivity log in $\Omega\cdot\text{m}$.

$$
\Delta t_C = 2 \times \left( \frac{\Delta t_C - 45.27}{93.482 - 45.27} \right) - 1
$$

(B7)

where $\Delta t_C$ is the compressional wave travel time in $\mu$s/ft.

$$
SP_n = 2 \times \left( \frac{SP + 154.18}{28.625 + 154.18} \right) - 1
$$

(B8)

where SP is the spontaneous potential log in mV.

**APPENDIX C MATHEMATICAL MODEL TO PREDICT $S_2$**

The ANN-based mathematical model to predict the $S_2$ of the organic shale is given by eq CC1

$$
S_2 = 0.7988 \times S_{2n} + 1.0383
$$

(C1)
The ANN-based mathematical model to predict the $S_3$ of the organic shale is given by eq DD1

$$ S_3 = 0.10915 \times S_3 + 0.2085 \quad (D1) $$

where

$$ S_{2n} = \sigma_c \sum_{i=1}^{N_h} w_{i,1} \sigma_i (w_{i,1} GR_n + w_{i,2} \rho_n + w_{i,3} \phi_n + w_{i,4} R_{ILD_1} + w_{i,5} \Delta t_{C_1} + w_{i,6} SP_n + b_i) + b_2 \quad (C2) $$

where $\sigma_i (x) = (2/1 + e^{-x}) - 1$; $\sigma_i (x) = x$; and $w_{i,1}, w_{i,2}, b_i$, and $b_2$ are the weights and biases of the $S_2$ model, given in Table 9. The equations to find $GR_n, \rho_n, \phi_n, R_{ILD_1}, \Delta t_{C_1}$ and $SP_n$ are given by eqs CC3–CC7.

$$ GR_n = 2 \times \left( \frac{GR - 18.3}{372.3 - 18.3} \right) - 1 \quad (C3) $$

where $GR$ is the $\gamma$-ray log in API.

$$ \rho_n = 2 \times \left( \frac{\rho - 2.37}{2.86 - 2.37} \right) - 1 \quad (C4) $$

where $\rho$ is the bulk density in g/cc.

$$ \phi_n = 2 \times \left( \frac{\phi - 0.003}{0.333 - 0.003} \right) - 1 \quad (C5) $$

where $\phi$ is the neutron porosity.

$$ R_{ILD_1} = 2 \times \left( \frac{R_{ILD} - 1.307}{1283 - 1.307} \right) - 1 \quad (C6) $$

where $R_{ILD}$ is the resistivity log in $\Omega$-m.

$$ \Delta t_{C_1} = 2 \times \left( \frac{\Delta t_C - 45.27}{93.482 - 45.27} \right) - 1 \quad (C7) $$

where $\Delta t_C$ is the compressional wave travel time in $\mu$s/ft.

$$ SP_n = 2 \times \left( \frac{SP + 154.188}{-28.625 + 154.188} \right) - 1 \quad (C8) $$

where $SP$ is the spontaneous potential log in mV.

### APPENDIX D MATHEMATICAL MODEL TO PREDICT $S_3$

The ANN-based mathematical model to predict the $S_3$ of the organic shale is given by eq DD1

$$ S_3 = 0.10915 \times S_3 + 0.2085 \quad (D1) $$

where

$$ S_{3n} = \sigma_c \sum_{i=1}^{N_h} w_{i,1} \sigma_i (w_{i,1} GR_n + w_{i,2} \rho_n + w_{i,3} \phi_n + w_{i,4} R_{ILD_1} + w_{i,5} \Delta t_{C_1} + w_{i,6} SP_n + b_i) + b_2 \quad (D2) $$

where $\sigma_i (x) = (2/1 + e^{-x}) - 1$; $\sigma_i (x) = x$; and $w_{i,1}, w_{i,2}, b_i$, and $b_2$ are the weights and biases of the $S_3$ model, given in Table 10. The equations to find $GR_n, \rho_n, \phi_n, R_{ILD_1}, \Delta t_{C_1}$ and $SP_n$ are given by eqs DD3–DD7.

$$ GR_n = 2 \times \left( \frac{GR - 18.3}{417 - 18.3} \right) - 1 \quad (D3) $$

where $GR$ is the $\gamma$-ray log in API.

$$ \rho_n = 2 \times \left( \frac{\rho - 2.37}{2.835 - 2.37} \right) - 1 \quad (D4) $$

where $\rho$ is the bulk density in g/cc.

$$ \phi_n = 2 \times \left( \frac{\phi - 0.003}{0.298 - 0.003} \right) - 1 \quad (D5) $$

where $\phi$ is the neutron porosity.

$$ R_{ILD_1} = 2 \times \left( \frac{R_{ILD} - 4.8}{1088 - 4.8} \right) - 1 \quad (D6) $$

where $R_{ILD}$ is the resistivity log in $\Omega$-m.

$$ \Delta t_{C_1} = 2 \times \left( \frac{\Delta t_C - 45.27}{92.083 - 45.27} \right) - 1 \quad (D7) $$

where $\Delta t_C$ is the compressional wave travel time in $\mu$s/ft.

$$ SP_n = 2 \times \left( \frac{SP + 154.188}{-28.625 + 154.188} \right) - 1 \quad (D8) $$

where $SP$ is the spontaneous potential log in mV.

### APPENDIX E MATHEMATICAL MODEL TO PREDICT TOTAL ORGANIC CARBON

The ANN-based mathematical model to predict the $S_3$ of the organic shale is given by eq EE1

$$ TOC = 3.382 \times TOC_n + 5.468 \quad (E1) $$

where
Table 11. Weights and Biases of the Proposed Model for TOC Prediction

| hidden layer neurons (N_h) | weights between input and hidden layers (w_i) | weights between hidden and output layers (w_o) | hidden layer bias (b_i) | output layer bias (b_o) |
|-----------------------------|-----------------------------------------------|-----------------------------------------------|------------------------|------------------------|
| 1                           | 1.759 × 10^-1.040 1.233 2.680 −0.320 −1.225 | 1.372 −1.683 1.923 |
| 2                           | 0.875 0.288 0.309 −1.399 −0.701 0.682 | 4.431 −0.756 |
| 3                           | 0.990 −0.414 −1.286 −1.252 −0.283 3.069 | −1.250 −1.101 |
| 4                           | −1.100 −1.601 −1.629 −3.078 0.297 −0.882 | 1.538 −3.488 |
| 5                           | 2.962 −1.130 0.781 −2.217 −1.349 0.861 | −0.787 0.840 |
| 6                           | −0.886 1.227 0.386 6.639 2.619 −3.780 | 1.059 2.219 |
| 7                           | 0.417 −1.863 0.815 −0.082 1.125 −1.183 | −1.524 2.046 |
| 8                           | −0.847 −1.005 0.479 0.991 0.410 0.273 | −0.556 −2.068 |
| 9                           | −0.606 −1.045 −1.203 0.960 1.449 0.264 | 2.000 1.132 |
| 10                          | 2.624 −2.675 −0.025 −0.704 0.800 −1.333 | 1.013 3.426 |

where $\sigma_i(x) = (2/1 + e^{-2x}) - 1$; $\sigma_i(x) = x$; and $w_i, w_j, b_1,$ and $b_2$ are the weights and biases of the TOC model, given in Table 11. The equations to find $GR, \rho, \varphi, R_{ILD}, \Delta t_{C}$, and $SP$ are given by eqs EE3–EE7.

$$GR_n = 2 \times \left( \frac{GR - 55}{359 - 55} \right) - 1$$

where $GR$ is the $\gamma$-ray log in API.

$$\rho_n = 2 \times \left( \frac{\rho - 2.2}{2.68 - 2.2} \right) - 1$$

where $\rho$ is the bulk density in g/cc.

$$\varphi_n = 2 \times \left( \frac{\varphi - 0.03}{0.330 - 0.03} \right) - 1$$

where $\varphi$ is the neutron porosity.

$$R_{ILD} = 2 \times \left( \frac{R_{ILD} - 6}{148.870 - 6.0} \right) - 1$$

where $R_{ILD}$ is the resistivity log in $\Omega$-m.

$$\Delta t_{C} = 2 \times \left( \frac{\Delta t_{C} - 62}{93.390 - 62} \right) - 1$$

where $\Delta t_{C}$ is the compressional wave travel time in $\mu$s/ft.

$$SP_n = 2 \times \left( \frac{SP + 86.69}{-29.25 + 86.69} \right) - 1$$

where $SP$ is the spontaneous potential log in mV.

**AUTHOR INFORMATION**

**Corresponding Author**
Mohamed Mahmoud — King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; orcid.org/0000-0002-4395-9567; Email: mmahmoud@kfupm.edu.sa

**Authors**
Zeeshan Tariq — King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

**ACKNOWLEDGMENTS**

The authors would like to acknowledge the College of Petroleum Engineering & Geosciences at King Fahd University of Petroleum & Minerals for providing support to conduct this research.

**ABBREVIATIONS USED**

AAPE average absolute percentage error
ANFIS adaptive neuro fuzzy inference system
ANN artificial neural network
CC correlation coefficient
CNL compensated neutron log
DT compressional wave travel time
EA evolutionary algorithm
GR $\gamma$-ray, API
K potassium
LLD deep lateral log
LLS shallow lateral log
LSSVM least-squares support vector machine
MAE mean absolute error
ML machine learning
MLR multiple linear regression
MSFL microspherical focused log
NPHI neutron porosity, V/V
PE photoelectric index
PSO particle swarm optimization
$R$ correlation coefficient
$R^2$ coefficient of determination
RHOB bulk density
$R_{ILD}$ deep induction resistivity log
RMSE root-mean-square error
RT resistivity log
SD standard deviation
SP spontaneous potential
SVM support vector machine
SYMBOLS

\( \alpha \) learning rate

\( b_1 \) biases vector between the input and middle layers

\( b_2 \) bias value between the middle and output layers

\( c_1 \) cognitive parameter \( (0 \leq c_1 \leq 1.2) \)

\( c_2 \) cognitive parameter \( (0 \leq c_2 \leq 1.2) \)

\( i \) index used for the total number of neurons

\( j \) index used for the number of inputs

\( F \) total number of input parameters

\( N_h \) total number of neurons

\( N_p \) total number of input parameters

\( g^i \) position of the \( i \)th particle

\( p^i \) best solution of the particle

\( R^g_b \) global best solution

\( R^2 \) coefficient of determination

\( v_1 \) particle velocity

\( w_1 \) weights matrix between the input and middle layers

\( w_2 \) weights vector between the middle and output layers

\( x \) input parameters

\( y \) output variable

\( \sigma_x \) transfer function between the middle and output layers

\( \sigma_y \) transfer function between the input and middle layers

\( \omega \) weight \( (0 \leq \omega \leq 1.2) \)

REFERENCES

(1) Sharma, S.; Agrawal, V.; Akondi, R. N. Role of Biogeochemistry in Efficient Shale Oil and Gas Production. Fuel 2020, 259, No. 116207.

(2) Hosseini, H.; Tsau, J. S.; Shafer-Peltier, K.; Marshall, C.; Ye, Q.; Ghahfarokhi, R. B. Experimental and Mechanistic Study of Stabilized Dry CO2 Foam Using Polyelectrolyte Complex Nanoparticles Compatible with Produced Water to Improve Hydraulic Fracturing Performance. Ind. Eng. Chem. Res. 2019, 58, 9431–9449.

(3) Lee, K. S.; Kim, T. H. Integrative Understanding of Shale Gas Reservoirs; Springer, 2016.

(4) Ahmed, S.; Sri Hanamertani, A.; Rehan Hashmet, M. CO 2 Foam as an Improved Fracturing Fluid System for Unconventional Reservoir. In Exploitation of Unconventional Oil and Gas Resources - Hydraulic Fracturing and Other Recovery and Assessment Techniques; IntechOpen, 2019; pp 1–24.

(5) Zhao, H.; Givens, N. B.; Curtis, B. Thermal Maturity of the Barnett Shale Determined from Well-Log Analysis. AAPG Bull. 2007, 91, 535–549.

(6) Javie, D. M.; Hill, R. J.; Ruble, T. E.; Pollastro, R. M. Unconventional Shale-Gas Systems: The Mississippi Barnett Shale of North-Central Texas as One Model for Thermogenic Shale-Gas Assessment. AAPG Bull. 2007, 91, 475–499.

(7) Zhang, T.; Ellis, G. S.; Ruppel, S. C.; Milliken, K.; Yang, R. Effect of Organic-Matter Type and Thermal Maturity on Methane Adsorption in Shale-Gas Systems. Org. Geochem. 2012, 47, 120–131.

(8) Gale, J. F. W.; Reed, R. M.; Holder, J. Natural Fractures in the Barnett Shale and Their Importance for Hydraulic Fracture Treatments. AAPG Bull. 2007, 91, 603–622.

(9) Věrová, D.; Geršlová, E.; Díaz-Somoano, M.; Sfrková, I.; Suárez-Ruiz, I.; Havelcová, M.; Kuta, J. Distribution and Mobility Potential of Trace Elements in the Main Seam of the Most Coal Basin. Int. J. Coal Geol. 2018, 196, 139–147.

(10) Tariq, Z.; Mahmood, M.; Abdullaheem, A. Core Log Integration: A Hybrid Intelligent Data-Driven Solution to Improve Elastic Parameter Prediction. Neural Comput. Appl. 2019, 31, 8561–8581.

(11) Gao, W.; Chen, X.; Chen, D. Genetic Programming Approach for Predicting Service Life of Tunnel Structures Subject to Chloride-Induced Corrosion. J. Adv. Res. 2019, 20, 141–152.

(12) Samui, P.; Hariraharan, R. A Unified Classification Model for Modeling of Seismic Liquefaction Potential of Soil Based on CPT. J. Adv. Res. 2015, 6, 587–592.

(13) Ghanem, T. F.; Elkilani, W. S.; Abdul-kader, H. M. A Hybrid Approach for Efficient Anomaly Detection Using Metaheuristic Methods. J. Adv. Res. 2015, 6, 609–619.

(14) Anifowose, F. A.; Labadin, J.; Abdullaheem, A. Ensemble Machine Learning: An Untapped Modeling Paradigm for Petroleum Reservoir Characterization. J. Pet. Sci. Eng. 2017, 151, 480–487.

(15) Anifowose, F.; Abdullaheem, A. Fuzzy Logic-Driven and SVM-Driven Hybrid Computational Intelligence Models Applied to Oil and Gas Reservoir Characterization. J. Nat. Gas Sci. Eng. 2011, 3, 505–517.

(16) Anifowose, F.; Labadin, J.; Abdullaheem, A. Improving the Prediction of Petroleum Reservoir Characterization with a Stacked Generalization Ensemble Model of Support Vector Machines. Appl. Soft Comput. 2015, 26, 483–496.

(17) Al-Marhoun, M. A.; Osman, E. A. In Using Artificial Neural Networks to Develop New PVT Correlations for Saudi Crude Oils, Abu Dhabi International Petroleum Exhibition and Conference; Society of Petroleum Engineers, 2002.

(18) Gharbi, R. B.; Elsharkawy, A. M.; Karkoub, M. Universal Neural-Network-Based Model for Estimating the PVT Properties of Crude Oil Systems. Energy and Fuels 1999, 13, 454–458.

(19) Gharbi, R. B.; Elsharkawy, A. M. In Neural Network Model for Estimating the PVT Properties of Middle East Crude Oils, Proceedings of the Middle East Oil Show; Society of Petroleum Engineers, 1997; Vol. 1, pp 151–166.

(20) Khan, M. R.; Tariq, Z.; Abdullaheem, A. Application of Artificial Intelligence to Estimate Oil Flow Rate in Gas-Lift Wells. Nat. Resour. Res. 2020, 1–13.

(21) Tariq, Z.; Abdullaheem, A.; Mahmoud, M.; Ahmed, A. A Rigorous Data-Driven Approach to Predict Poisson’s Ratio of Carbonate Rocks Using a Functional Network. Petrophysics 2018, 59, 761–777.

(22) Tariq, Z.; Mahmoud, M. A.; Abdullaheem, A.; Al-Shehri, D. A. In On Utilizing Functional Network to Develop Mathematical Model for Poisson’s Ratio Determination, 52nd U.S. Rock Mechanics/Geomechanics Symposium; American Rock Mechanics Association: Seattle, Washington, 2018.

(23) Rahaman, M. S. A.; Vasant, P. Artificial Intelligence Approach for Predicting TOC From Well Logs in Shale Reservoirs; IGI Global, 2020; pp 46–77.

(24) Tan, M.; Song, X.; Yang, X.; Wu, Q. Support-Vector-Regression Machine Technology for Total Organic Carbon Content Prediction from Wireline Logs in Organic Shale: A Comparative Study. J. Nat. Gas Sci. Eng. 2015, 26, 792–802.

(25) Rui, J.; Zhang, H.; Zhang, D.; Han, F.; Guo, Q. Total Organic Carbon Content Prediction Based on Support-Vector-Regression Machine with Particle Swarm Optimization. J. Pet. Sci. Eng. 2019, 180, 699–706.

(26) Khoshsoodkia, M.; Mohseni, H.; Rahmani, O.; Mohammadi, A. TOC Determination of Gadwan Formation in South Pars Gas Field, Using Artificial Intelligent Systems and Geochemical Data. J. Pet. Sci. Eng. 2011, 78, 119–130.

(27) Lawal, L. O.; Mahmoud, M.; Saheed Alade, O.; Abdulraheem, A. Total Organic Carbon Characterization Using Neural-Network Analysis of XRF Data. Petrophysics 2019, 60, 480–493.

(28) Sultan, A. In New Artificial Neural Network Model for Predicting the TOC from Well Logs, SPE Middle East Oil and Gas Show and Conference; Society of Petroleum Engineers, 2019.

(29) Mendhe, V. A.; Mishra, A. K.; Kamble, A. D.; Bannereje, M.; Singh, B. D.; Sutay, T. M.; Singh, V. P. Geochemical and Petrophysical Characteristics of Permian Shale Gas Reservoirs of Raniganj Basin, West Bengal, India. Int. J. Coal Geol. 2018, 188, 1–24.

(30) Yu, F.; Sun, P.; Zhao, K.; Ma, L.; Tian, X. Experimental Constraints on the Evolution of Organic Matter in Oil Shales during Heating: Implications for Enhanced in Situ Oil Recovery from Oil Shales. Fuel 2020, 261, No. 116412.
(31) Ahmed, M. A.; Hassan, M. M. A. Hydrocarbon Generating-Potential and Maturity-Related Changes of the Khatatba Formation, Western Desert, Egypt. *Pet. Res. 2019*, 4, 148−163.

(32) Adeoye, J. A.; Akande, S. O.; Adekeye, O. A.; Sonibare, W. A.; Ondrak, R.; Dominik, W.; Erdtmann, B. D.; Neeka, J. Source Rock Maturity and Petroleum Generation in the Dahomey Basin SW Nigeria: Insights from Geologic and Geochemical Modelling. *J. Pet. Sci. Eng. 2020*, 195, No. 107844.

(33) Tissot, B. P.; Welte, D. H.; Tissot, B. P.; Welte, D. H. From Kerogen to Petroleum. In *Petroleum Formation and Occurrence*; Springer, 1978.

(34) Herron, M. M.; Loan, M. E.; Charsky, A. M.; Herron, S. L.; Pomerantz, A. E.; Polyakov, M. In *Kerogen Content and Maturity, Mineralogy and Clay Typing from Drifts Analysis of Cuttings or Core*; SPWLA 55th Annual Logging Symposium 2014; Society of Petrophysicists and Well-Log Analysts, 2014.

(35) Bowker, K. A. Barnett Shale Gas Production, Fort Worth Basin: Issues and Discussion. *AAPG Bull. 2007*, 91, 523−533.

(36) Ekwoezor, C. M.; Gomri, J. R. Petroleum Geochemistry Of Late Cretaceous And Early Tertiary Shales Penetrated By The Akukwa-2 Well In The Anambra Basin, Southern Nigeria. *J. Pet. Geol. 1983*, 6, 207−215.

(37) Philip, R. P. Petroleum Formation and Occurrence. *Eos, Trans. Am. Geophys. Union 1985*, 66, No. 643.

(38) Schmoker, J. Determination of Organic Content of Appalachian Devonian Shales from Formation-Density Logs: GEOLOGIC NOTES. *AAPG Bull. 1979*, 63, 1504−1509.

(39) Aziz, H.; Ehsan, M.; Ali, A.; Khan, H. K.; Khan, A. Hydrocarbon Source Rock Evaluation and Quantification of Organic Richness from Correlation of Well Logs and Geochemical Data: A Case Study from the Sembar Formation, Southern Indus Basin, Pakistan. *J. Nat. Gas Sci. Eng. 2020*, 81, No. 103433.

(40) Peters, K. E.; Simonett, B. T. Rock-Eval Pyrolysis of Quaternary Sediments from Leg 64, Sites 479 and 480, Gulf of California. In *Initial Reports of the Deep Sea Drilling Project 64*; U.S. Government Printing Office, 1982.

(41) Engleman, E. E.; Jackson, L. L.; Norton, D. R. Determination of Carbonate Carbon in Geological Materials by Coulometric Titration. *Chem. Geol. 1985*, 53, 125−128.

(42) Lafargue, E.; Marquis, F.; Pillot, D. Rock-Eval 6 Applications in Hydrocarbon Exploration, Production, and Soil Contamination Studies. *Rev. Inst. Fr. Pet. 1998*, 53, 421−437.

(43) Elias, R.; Duclerc, D.; Le-Van-Loi, R.; Gelin, F.; Dessort, D. In *A New Geological Tool for the Assessment of Organic-Rich Shales, Unconventional Resources Technology Conference; Society of Exploration Geophysicists, American Association of Petroleum Geologists and Society of Petroleum Engineers: Denver, Colorado*, 12−14 August, 2013; pp 2002−2006.

(44) Kadkhodaie-Ilkhchi, A.; Rahimpour-Bonab, H.; Rezaee, M. A. Committee Machine with Intelligent Systems for Estimation of Total Organic Carbon Content from Petrophysical Data: An Example from Kangan and Dalan Reservoirs in South Pars Gas Field, Iran. *Comput. Geosci. 2009*, 35, 459−474.

(45) Handhal, A. M.; Al-Abadi, A. M.; Chaifeet, H. E.; Ismail, M. J. Prediction of Total Organic Carbon at Rumaila Oil Field, Southern Iraq Using Conventional Well Logs and Machine Learning Algorithms. *Mar. Pet. Geol. 2020*, 116, No. 104347.

(46) Alizadeh, B.; Maroufi, K.; Heidarifard, M. H. Estimating Source Rock Parameters Using Wireline Data: An Example from Dezful Embayment, South West of Iran. *J. Pet. Sci. Eng. 2018*, 167, 857−868.

(47) Wang, P.; Chen, Z.; Pang, X.; Hu, K.; Sun, M.; Chen, X. Revised Models for Determining TOC in Shale Play: From Devonian Duvernay Shale, Western Canada Sedimentary Basin. *Mar. Pet. Geol. 2016*, 70, 304−319.

(48) Wang, H.; Wu, W.; Chen, T.; Dong, X.; Wang, G. An Improved Neural Network for TOC, S1 and S2 Estimation Based on Conventional Well Logs. *J. Pet. Sci. Eng. 2019*, 176, 664−678.

(49) Tan, M.; Liu, Q.; Zhang, S. A Dynamic Adaptive Radial Basis Function Approach for Total Organic Carbon Content Prediction in Organic Shale. *Geophysics 2013*, 78, D445−D459.

(50) Zhao, P.; Ostadkhassan, M.; Shen, B.; Liu, W.; Abarghani, A.; Liu, K.; Luo, M.; Cai, J. Estimating Thermal Maturity of Organic-Rich Shale from Well Logs: Case Studies of Two Shale Plays. *Fuel 2019*, 235, 1195−1206.

(51) Zhao, P.; Ma, H.; Rasouli, V.; Liu, W.; Cai, J.; Huang, Z. An Improved Model for Estimating the TOC in Shale Formations. *Mar. Pet. Geol. 2017*, 83, 174−183.

(52) Rui, J.; Zhang, H.; Ren, Q.; Yan, L.; Guo, Q.; Zhang, D. TOC Content Prediction Based on a Combined Gaussian Process Regression Model. *Mar. Pet. Geol. 2020*, 118, No. 104429.

(53) Wang, J.; Gu, D.; Guo, W.; Zhang, H.; Yang, D. Determination of Total Organic Carbon Content in Shale Formations With Regression Analysis. *J. Energy Resour. Technol. 2019*, 141, No. 012907.

(54) Schmoker, J. Organic Content of Devonian Shale in Western Appalachian Basin. *AAPG Bull. 1980*, 64, 2156−2165.

(55) Passey; Creaney, J. A Practical Model for Organic Richness from Porosity and Resistivity Logs. *AAPG Bull. 1990*, 74, 1777−1794.

(56) Mahmoud, A. A. A.; Elkattany, S.; Mahmoud, M.; Abouelrash, M.; Abdulraheem, A.; Ali, A. Determination of the Total Organic Carbon (TOC) Based on Conventional Well Logs Using Artificial Neural Network. *Int. J. Coal Geol. 2017*, 179, 72−80.

(57) Pollastro, R. M. Total Petroleum System Assessment of Unconventional Resources in the Giant Barnett Shale Continental (Unconventional) Gas Accumulation, Fort Worth Basin; Texas. *AAPG Bull. 2007*, 91, 551−578.

(58) Abouelrash, M. O.; Slatt, R. M. Lithofacies and Sequence Stratigraphy of the Barnett Shale in East-Central Fort Worth Basin, Texas. *AAPG Bull. 2012*, 96, 1−22.

(59) Tariq, Z.; Mahmoud, M.; Abdulraheem, A. Real-Time Prognosis of Flowing Bottom-Hole Pressure in a Vertical Well for a Multiphase Flow Using Computational Intelligence Techniques. *J. Pet. Explor. Prod. Technol. 2020*, 10, 1411−1428.

(60) Mohaghegh, S. D. Shale Analytics: Data-Driven Analytics in Unconventional Resources; Springer International Publishing: Cham, 2017.

(61) Mahmoud, M.; Tariq, Z.; Kamal, M. S.; Al-Naser, M. Intelligent Prediction of Optimum Separation Parameters in the Multistage Crude Oil Production Facilities. *J. Pet. Explor. Prod. Technol. 2019*, 9, 2979−2995.

(62) Khan, M. R.; Tariq, Z.; Abdulraheem, A. In *Machine Learning Derived Correlation to Determine Water Saturation in Complex Lithologies, SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition; Society of Petroleum Engineers*, 2018.

(63) Vasumathi, B.; Moorthi, S. Implementation of Hybrid ANNSPSO Algorithm on FPG for Harmonic Estimation. *Eng. Appl. Artif. Intell. 2012*, 25, 476−483.

(64) Wang, J.; Zhou, Q.; Jiang, H.; Hou, R. Short-Term Wind Speed Forecasting Using Support Vector Regression Optimized by Cuckoo Optimization Algorithm. *Math. Probl. Eng. 2015*, 2015, 1−13.

(65) Chatterjee, S.; Sarkar, S.; Hore, S.; Dey, N.; Ashour, A. S.; Balas, V. E. Particle Swarm Optimization Trained Neural Network for Structural Failure Prediction of Multistoried RC Buildings. *Neural Comput. Appl. 2017*, 28, 2005−2016.

(66) Catalão, J. P. S.; Pousinho, H. M. I.; Mendes, V. M. F. Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. *IEEE Trans. Sustainable Energy 2010*, 2, 50−59.

(67) Ethabi, S.; Omar, R.; Mazlina, M. K. S.; Radiah, A. B. D.; Syafie, S. Development of a Hybrid PSO−ANN Model for Estimating Glucose and Xylose Yields for Microwave-Assisted Pretreatment and the Enzymatic Hydrolysis of Lignocellulosic Biomass. *Neural Comput. Appl. 2018*, 30, 1111−1121.

(68) Abido, M. A. Optimal Design of Power-System Stabilizers Using Particle Swarm Optimization. *IEEE Trans. Energy Convers. 2002*, 17, 406−413.