ACS OMEGA

http://pubs.acs.org/journal/acsodf

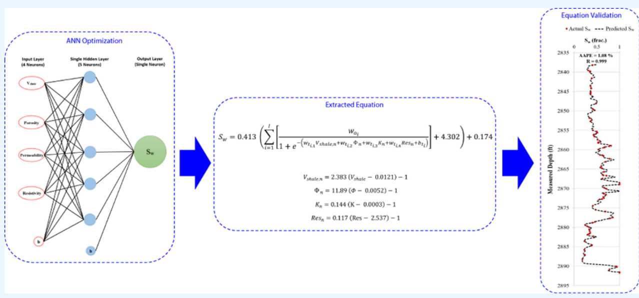
Article

Predicting Water Saturation in a Greek Oilfield with the Power of Artificial Neural Networks

Mohammed Gad, Ahmed Abdulhamid Mahmoud,* George Panagopoulos, Paschalia Kiomourtzi, Panagiotis Kirmizakis, Salaheldin Elkatatny, Umair bin Waheed, and Pantelis Soupios

Cite This: ACS Omega 2025, 10, 557−566

ABSTRACT: Water saturation plays a vital role in calculating the volume of hydrocarbon in reservoirs and defining the net pay. It is also essential for designing the well completion. Innacurate water saturation calculation can lead to poor decision-making, significantly affecting the reservoir's development and production, potentially resulting in reduced hydrocarbon oil recovery. Various techniques to estimate the water saturation in both clean and shaly formations. However, the most widely used approaches in the petroleum industry rely on petrophysical models, including Archie's equation, Waxman-Smits, Simandoux, Indonesia, and dual-water models. Most of these methods are only valid for clean sands or carbonate, while the presence of clay significantly limits the accuracy of these models. On the other hand, the estimation of the water saturation through core analysis does not usually cover a large interval of the well, is highly costly, and requires much time. In this study, an empirical equation for predicting water saturation based on the weight and biases of the artificial neural networks (ANN) was developed. 334 data points of the shale volume, formation deep resistivity, porosity, and permeability and their corresponding water saturation collected from the Epsilon Field in Greece were considered for optimizing the ANN model. The ANN model was trained on 252 data sets, where the water saturation was predicted with an average absolute percentage error (AAPE) of 0.90%. Then, an empirical equation was developed based on the optimized ANN model and its weights and biases. The developed equation predicted the water saturation for the remaining 82 data sets (testing data) with an AAPE of 1.08%. The newly established empirical correlation enhances the precision of water saturation prediction and provides a cost-effective means to acquire a continuous water saturation profile, a critical asset for oilfield management and hydrocarbon exploration.



1. INTRODUCTION

Most hydrocarbon reservoirs have two different fluids within their pore spaces, and some have all three phases of gas, oil, and water.[1−3] Therefore, knowing water saturation (S_w) is essential to determine the hydrocarbon content. Kamel and Mabrouk[4] assumed that all reservoirs' void spaces consist of water and hydrocarbon; therefore, $S_h = 1\text{-}S_w$. Determining water saturation is complex; it started in 1942 by incorporating some well logs in clean sandstones.[5] Then, many scientists proposed various equations to validate this procedure in shaly sands and carbonates. Later, many scientists proposed different relations to validate the proposed correlation between shaly sands and carbonates. The main challenge in this process is its dependency on laboratory core analysis, which takes a lot of time and is costly.[6] In addition, it does not provide a continuous recording through the well but only a discrete measurement at specific depths. To tackle the dependency of water saturation estimation on core analysis, researchers suggested using well logs and performing formation evaluation to calculate the water saturation as continuous values with depths. The relationship between water saturation, formation resistivity, and porosity logs is described by Archie's equation,[5] eq 1 and shown below:

$$S_w = \left(\frac{aR_w}{\Phi^m R_t}\right)^{1/n} \tag{1}$$

where S_w is the water saturation, R_t is the true formation resistivity, Φ^m is the formation porosity R_w is the formation water resistivity, and a, m, and n are constants (a: tortuosity constant, m: cementation exponent, n: saturation exponent) that are based on the pore geometry, cementation, and rock type.

It should be mentioned that the formation porosity (Φ^m), can be estimated from neutron (NPHI), density,[7] nuclear magnetic

Received: August 5, 2024
Revised: December 18, 2024
Accepted: December 23, 2024
Published: January 3, 2025

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

Table 1. Comparison of Application, Advantages, and Disadvantages of the Conventional Correlations for Prediction of Water Saturation

Method	Application	Advantages	Disadvantages
Archie's equation[5]	Applicable for homogeneous and Clean formation	●Simplicity	●Not suitable for complex lithologies
			●Sensitivity to cementation and saturation exponents
			●It needs to be calibrated and corrected
Indonesia equation	Can handle complex lithologies	●Suitable for a wide range of reservoir conditions	●Limited historical use
		●Incorporates salinity effects	●Calibration may be necessary to achieve optimal accuracy in specific reservoirs, which can be time-consuming
			●It relies on specific input parameters, including salinity, which must be accurately measured
Simandoux equation	Improved accuracy in shaly sands	●Suitable for reservoirs with significant clay content	●Requiring the incorporation of clay-specific parameters, which can be challenging
			●Primarily intended for shaly sands and may not be the best choice for other lithologies
dual-water models	Widely used in mixed-wet reservoirs	●Physically realistic	●Increased complexity
		●Improved accuracy in mixed-wet reservoirs	●Accurate capillary pressure and wettability data are needed for robust application
			●Primarily suited for mixed-wet reservoirs and may not be necessary or applicable in water-wet or oil-wet reservoirs

resonance (NMR),[8] or acoustic logs.[9] The self-potential log (SP) can be used to calculate the formation water resistivity (Rw).

Archie's equation is restricted to the clean formation, which is not the case in most hydrocarbon reservoirs since the existence of clay causes some errors in resistivity measurements because clay increases conductivity. Moreover, it is not common in oil and gas fields to have SP logs for the R_w estimation. In addition, if the rock matrix is unknown, then a low accuracy of porosity estimation is expected. Furthermore, m and n constants cannot be measured. Consequently, the estimation of S_w from Archie's eq 1 needs to be calibrated and corrected.

Other techniques, such as the Waxman−Smits, Indonesia, eq 2 and Simandoux, eq 3, and dual-water models, were proposed to modify and revise Archie's equation by considering the shale factor into Archie's equation.

$$S_w = \left(\frac{\sqrt{1/R_t}}{\left(\frac{V_{sh}^{1-0.5V_{sh}}}{\sqrt{R_{sh}}} \right) + \sqrt{\frac{\Phi^m}{aR_w}}} \right)^{2/n} \tag{2}$$

$$S_w = \frac{aR_w}{2\Phi^m}\left(\sqrt{(\frac{V_{sh}}{R_{sh}})^2 + \frac{4\Phi^m}{aR_wR_t}} - \frac{V_{sh}}{R_{sh}} \right) \tag{3}$$

where V_{sh} is the shale volume, R_{sh} is shale resistivity, Φ is the formation porosity, and R_t is the true formation resistivity.

Also, the existence of nonconductive kerogen underestimates the water saturation. Table 1 compares the conventional correlations' application, advantages, and disadvantages. As the comparison in this table indicates, each model applies to specific formations and has low accuracy in others. Changing the formation also requires calibration of some of the parameters used in these equations to ensure high accuracy.

The conventional way to estimate the water saturation is by applying the Special Core Analysis (SCAL) at specific depths, which is accurate but requires time and is highly expensive[10,11] Although conventional well logs provide the water saturation as a continuous profile, not discrete as SCAL, estimating water saturation is still challenging due to the uncertainty of the constants' calculations.

Artificial intelligence (AI) is probably the most general-purpose technology rapidly entering industries, triggering and leading to significant improvements in multiple sectors (i.e., energy, healthcare, transportation, retail, media, and finance) in the international market and changing the competition rules.[12] AI is a game-changer since companies are moving from traditional and human-centered business processes to more automated and real-time processes companies, minimizing the risk and optimizing and/or maximizing the benefits of any company using AI solutions. Advanced algorithms are trained in large and multiparametric data sets. The algorithms can automatically produce more accurate and reliable models by adding more data.

Oil and gas (O and G) and mining companies are the latecomers to digitalization. Still, they are rapidly becoming more dependent on AI solutions to improve the efficiency of production and exploration phases by utilizing advanced digital technologies. In practice, enhancing the efficiency in O and G typically means accelerating processes and reducing risks now and in the coming years during the development of an oil field.[13−15] Machine learning techniques can play a vital role in solving such difficulties properly.[16,17] Various machine learning methods such as artificial neural network (ANN),[18−21] adaptive neuro-fuzzy inference systems,[16] functional neural networks,[22−24] support vector machine,[25,26] and random forests[27,28] were successfully applied for predicting various parameters related to petroleum engineering and other fields as well.[29−32] Hence, they may have a potential application for predicting water saturation without additional cost. An ANN is a computing system designed to imitate the organizing principles of the nervous system processes. For example, ANN can handle a mapping problem by discovering a close approximation of the connection between input and output data by learning automatically from provided training patterns. This separates it from other typical expert systems.[16] An ANN computing system comprises artificial neurons that act as fundamental components and mimic the parallel process of a biological brain to get the solution.

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

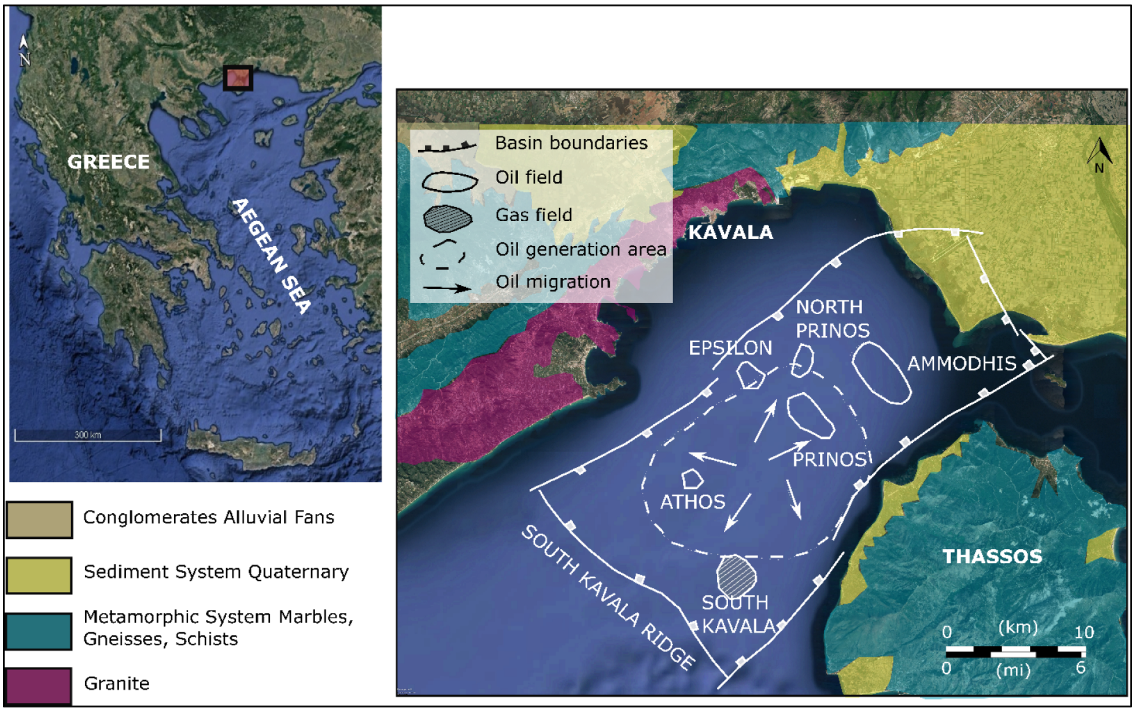


Figure 1. Location of the Epsilon oil field in the Northern Aegean Sea. The Epsilon oil field, positioned within the northern sector of the Prinos basin, lies between the Greek mainland and Thassos island. The oil and the migration paths are shown.

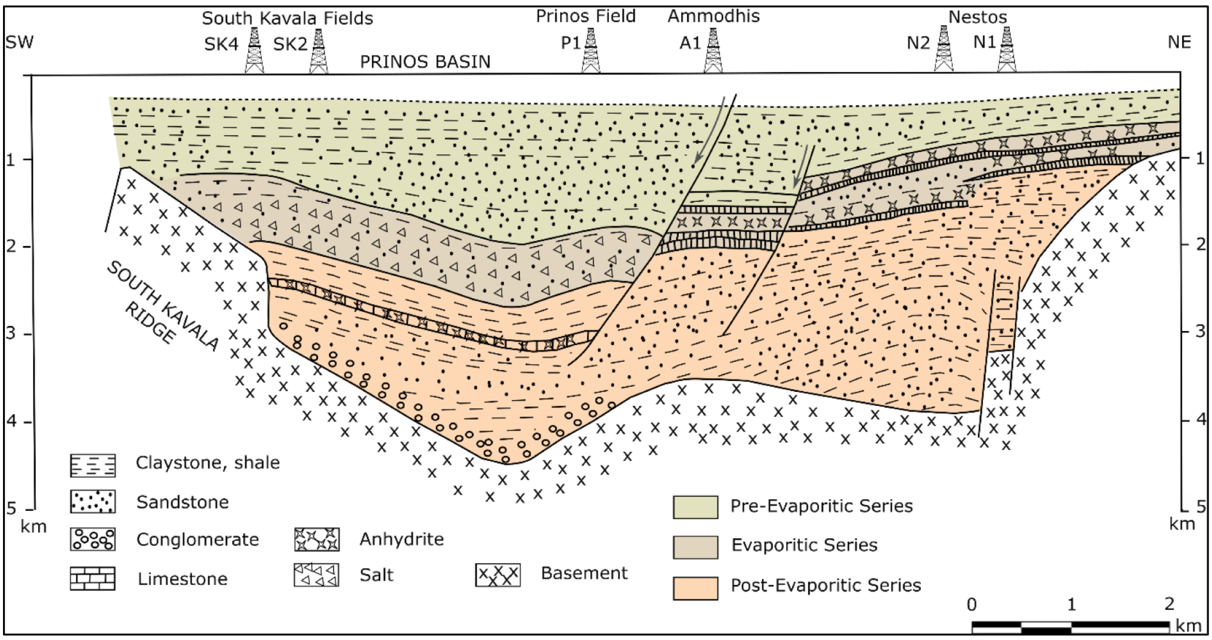


Figure 2. A SW-NE geological section of the broader Prinos Basin is presented. The extensive faulting formed an anticline. The stratigraphic sequence is also presented in different colors. Depths are in kilometers below mean sea level.

This paper developed an empirical correlation for water saturation evaluation as a function of the weights and biases of the optimized ANN model based on the well log data of the shale volume, formation deep resistivity, porosity, and permeability.

2. STUDY AREA

The Prinos basin is currently the only active hydrocarbon-producing oilfield in Greece, with the Epsilon oil field being the most recently developed.[33] The Epsilon oil field lies in the northern Aegean Sea, situated between the Greek mainland and the island of Thassos, about 11 km south−southeast of Kavala (Figure 1).[34,35] Production from this region began in 1976, following a series of exploratory efforts. The initial discovery well drilled in 1971 about 20 km east of Thassos island, encountered oil, but it was extremely low-gravity. The following two wells in 1972−1973 focused on the area west of Thassos, which resulted in the discovery of the South Kavala gas field at a water depth of 52 m. In late 1973, the fourth well was successfully drilled targeting the central part of the Prinos basin, this leaded to successfully uncovering the Prinos oil field in waters about 100

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

feet deep. This field remains significant in Greece's domestic energy production.

The stratigraphy model of the taphrogenetic basin, as shown in Figure 1, has been detailed by Pollak,[36] Proedrou,[37] and Proedrou and Sidiropoulos[38] (Figure 2). The model divides the generalized stratigraphic column into three main series: the Pre-Evaporitic Series, which consists of conglomerates, sandstones, siltstone/shales, and localized limestone formations; the Evaporitic Series, made primarily of salt and anhydrite deposits; and the Post-Evaporitic Series, which is characterized by siltstone/shales and sandstones. The taphrogenetic basin's overall structure is a dome-shaped anticline created by NW-striking and SW-dipping syngenetic faults connected to the trapping process.[33,35] The primary reservoir rocks are sandstones and siltstones deposited during the late Miocene, sourced from deltaic, marine, and turbiditic settings,[34,39,40] The entire basin is capped by evaporites, which acts as impermeable seal for the trapped hydrocarbons, except in areas like the South Kavala and Ammodhis fields, where fault activation has likely facilitated the upward migration of hydrocarbons.[35]

3. METHODOLOGY

Herein, we describe the methodology for training the ANN model, extracting the weights and biases of the optimized ANN model, developing the empirical equation, and testing the developed equation.

3.1. Data Preparation. For the water saturation prediction, 364 data points of shale volume (V_{shale}), formation porosity, permeability, deep resistivity measurement, and corresponding water saturation were used from Well-1. It is essential to mention that the permeability values considered are calculated values based on well-log data and not measured permeability. In the first stage, the data were preprocessed to remove all nonreal (unexpected/outliers) values; for example, all inputs with water saturation values less than zero or greater than 1.0 were considered nonreal inputs and removed from the input data set. For outliers removal, a statistical analysis was applied to all input parameters, removing all values outside ±3.0 standard deviation range. We also performed clustering and density analysis of the data to ensure no outliers were present.

3.2. Training the ANN Model. After data preprocessing, 30 points were removed, leaving 334 valid data sets for model training. The artificial neural network (ANN) was trained to predict water saturation, with key design parameters like learning function, number of layers, neurons per layer, and transfer functions optimized. The training data percentage was adjusted between 40% and 90% for best results. Several training functions, including Levenberg–Marquardt,[41,42] resilient backpropagation, and Bayesian regularization, were assessed. Additionally, transfer functions like tangential sigmoidal, logarithmic sigmoidal, and linear were evaluated for their effectiveness in predicting water saturation. The study aimed to optimize these parameters to improve the ANN model's accuracy in predicting water saturation in reservoirs.

For optimization and evaluation of all possible combinations of these design parameters, inserted *for* loops were built using MATLAB software; each loop is to optimize one of the design parameters. The use of single, two, or three layers for learning the model with 4 to 30 neurons per layer was also studied before choosing and finalizing the optimum ANN model.

It is important to mention here that since this is the first paper to optimize ANN model for water saturation prediction, our objective was to try all possible combinations of the ANN model, this is to help us understanding more the model, since the automatic hyperparameter optimization methods will only select specific combination of parameters, we just avoid it in this first study. Incorporating methods like Bayesian optimization or grid search in future studies would streamline the process and potentially improve the model's performance.

The learning stage results concluded that 70% of the data could be used for training using a single learning layer associated with five neurons. The Bayesian regularization backpropagation function for training and the logarithmic sigmoidal function were the optimum learning and transferring functions, respectively, for water saturation. Figure 3 shows a schematic

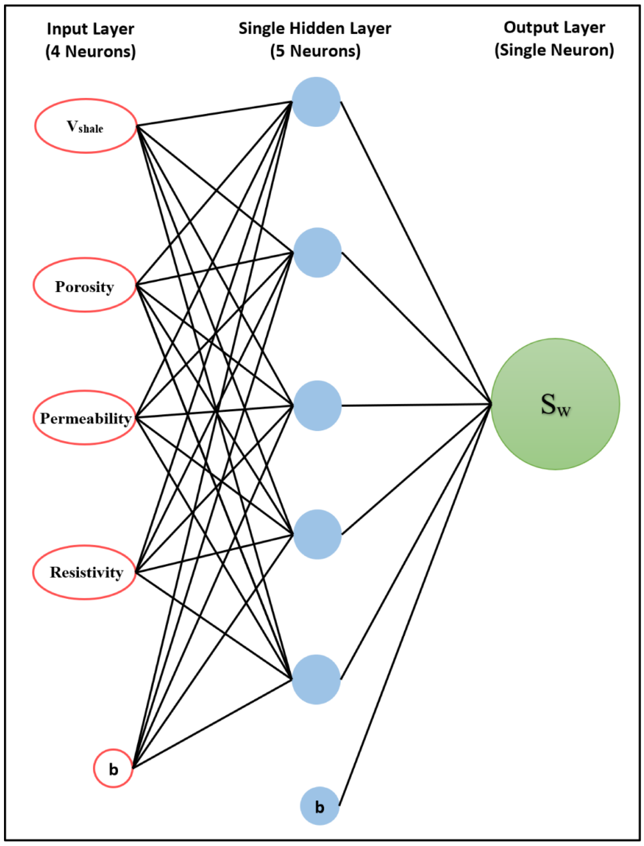


Figure 3. Schematic of the optimized ANN model for water saturation prediction. The letter b denotes the biased nodes.

of the optimized ANN model proposed for water saturation. Other properties of this optimized model are summarized in Table 2, whereas the statistical features and ranges of the training data sets are listed in Table 3.

Since limited data was used to develop this model (252 data sets for training and 82 for testing), it is very important to monitor the changes in the training and testing errors to ensure there was no overfitting issue. During the training process and

Table 2. Optimized Parameters for Water Saturation Prediction

Parameter	Optimum Value
Training layers	Single
Number of neurons	Five
Training function	Bayesian regularization backpropagation
Transferring function	Logarithmic sigmoidal function

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

Table 3. Statistical Features of the 252 Datasets used to Train the ANN for Water Saturation Estimation

	V_{shale}(frac.)	Porosity (frac.)	Permeability (mD)	Resistivity (Ohm.m)	S_w(frac.)
Minimum	0.012	0.005	0.0003	2.537	0.174
Maximum	0.852	0.173	13.86	19.63	1.000
Average	0.291	0.096	3.494	6.624	0.489
Mean	0.227	0.087	2.443	5.790	0.448
Median	0.231	0.099	2.952	5.535	0.427
Sample variance	0.037	0.001	5.989	14.80	0.041
Standard deviation	0.193	0.036	2.447	3.847	0.203

Table 4. Extracted Weights and Biases for water Saturation estimation

		Input Layer					Output Layer	
		Weights (w_t)						
		j = 1	j = 2	j = 3	j = 4	Biases (b_t)	Weights (w_O)	Biases (b_O)
Number of neurons	I = 1	1.765	−0.061	−0.113	−1.094	−2.308	−3.828	3.302
	I = 2	0.569	7.871	0.855	−1.174	5.078	−0.579	
	I = 3	−0.269	−1.088	−1.479	−0.515	−2.930	3.360	
	I = 4	−2.487	−2.741	1.120	0.899	2.414	−3.865	
	I = 5	−0.346	−2.111	0.574	−3.528	−4.541	2.913	

until iteration 42, both errors were decreasing. After 42 iterations, the testing error started increasing while the training error was still decreasing. This indicated that the optimum number of iterations to avoid overfitting should be 42.

3.3. Analytical eq Derivation from Optimized ANN. The ANN model shown in Figure 3, with the properties shown above in Table 2, was then considered for developing a new equation for water saturation estimation. This model could be generally expressed as shown in eq 4.

$$S_{w,n} = \sum_{i=1}^{I} w_{o_i} \log \operatorname{sig}\left(\sum_{j=1}^{J} w_{t_{i,j}} Y_j + b_{t_i}\right) + b_o = \sum_{i=1}^{I} w_{o_i} \frac{1}{1 + e^{-\left(\sum_{j=1}^{J} w_{t_{i,j}} Y_j + b_{t_i}\right)}} + b_o \tag{4}$$

where $S_{w,n}$ is the normalized water saturation as estimated from the input Y parameter, and I and J correspond to the total number of neurons and input parameters, which in this case are five and four, respectively. The variables w and b stand for the weights and biases of the network, crucial for determining the strength of connections between neurons. The symbol t refers to the training layer, which handles the iterative adjustment of weights through the learning process, while o signifies the output layer[43] responsible for producing the final prediction or classification based on the trained network.

It is important to note that the parameters used in the ANN model are typically normalized automatically using a two-point slope method, as described by eqs 5 and 6. This normalization process adjusts the parameter values to fall within the range of −1 to 1, ensuring consistency and improving the model's performance by standardizing the input data for more effective learning and prediction.

$$\frac{Y_n - Y_{n,\min}}{Y_{n,\max} - Y_{n,\min}} = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \tag{5}$$

$$Y_n = \left(\frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}}\right)(Y_{n,\max} - Y_{n,\min}) + Y_{n,\min} \tag{6}$$

In this context, Y represents the input parameter, which could be any of the following parameters: V_{shale}, porosity, permeability, resistivity, or water saturation. The subscript n indicates the normalized version of the parameter, while min and max refer to the minimum and maximum values for each input parameter, as outlined in Table 3. For instance, since the input parameters are normalized to fall within the range of −1 to 1, the minimum normalized value $Y_{n,min}$ will be −1, and the maximum normalized value $Y_{n,max}$ will be 1. Thus, eq 6 can be reformulated and expressed in the same format as eq 7), reflecting the normalized relationship between the original and normalized input parameters, ensuring all values align with the normalized scale.

$$Y_n = 2\left(\frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}}\right) - 1 \tag{7}$$

To derive the general normalization expressions for the four input parameters used in this study, one can substitute the corresponding minimum and maximum values for each parameter, as presented in Table 3. By doing so, the resulting equations are formulated, leading to the expressions outlined in eqs 8–q 1111).

$$V_{\text{shale},n} = 2.383(V_{\text{shale}} - 0.0121) - 1 \tag{8}$$

$$\Phi_n = 11.89(\Phi - 0.0052) - 1 \tag{9}$$

$$K_n = 0.144(K - 0.0003) - 1 \tag{10}$$

$$\text{Res}_n = 0.117(\text{Res} - 2.537) - 1 \tag{11}$$

where $V_{shale,n}$, Φ_n, K_n, and Res_n are the normalized shale volume, porosity, permeability, and resistivity, respectively.

eq 4 can be expressed using normalized parameters. This expansion and rewriting of the equation result in eq 12, which now incorporates these normalized variables for further analysis or calculation.

$$S_{w,n} = \sum_{i=1}^{I} w_{o_i} \frac{1}{1 + e^{-\left(w_{t_{i,1}} V_{\text{shale},n} + w_{t_{i,2}} \Phi_n + w_{t_{i,3}} K_n + w_{t_{i,4}} \text{Res}_n + b_{t_i}\right)}} + b_o \tag{12}$$

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

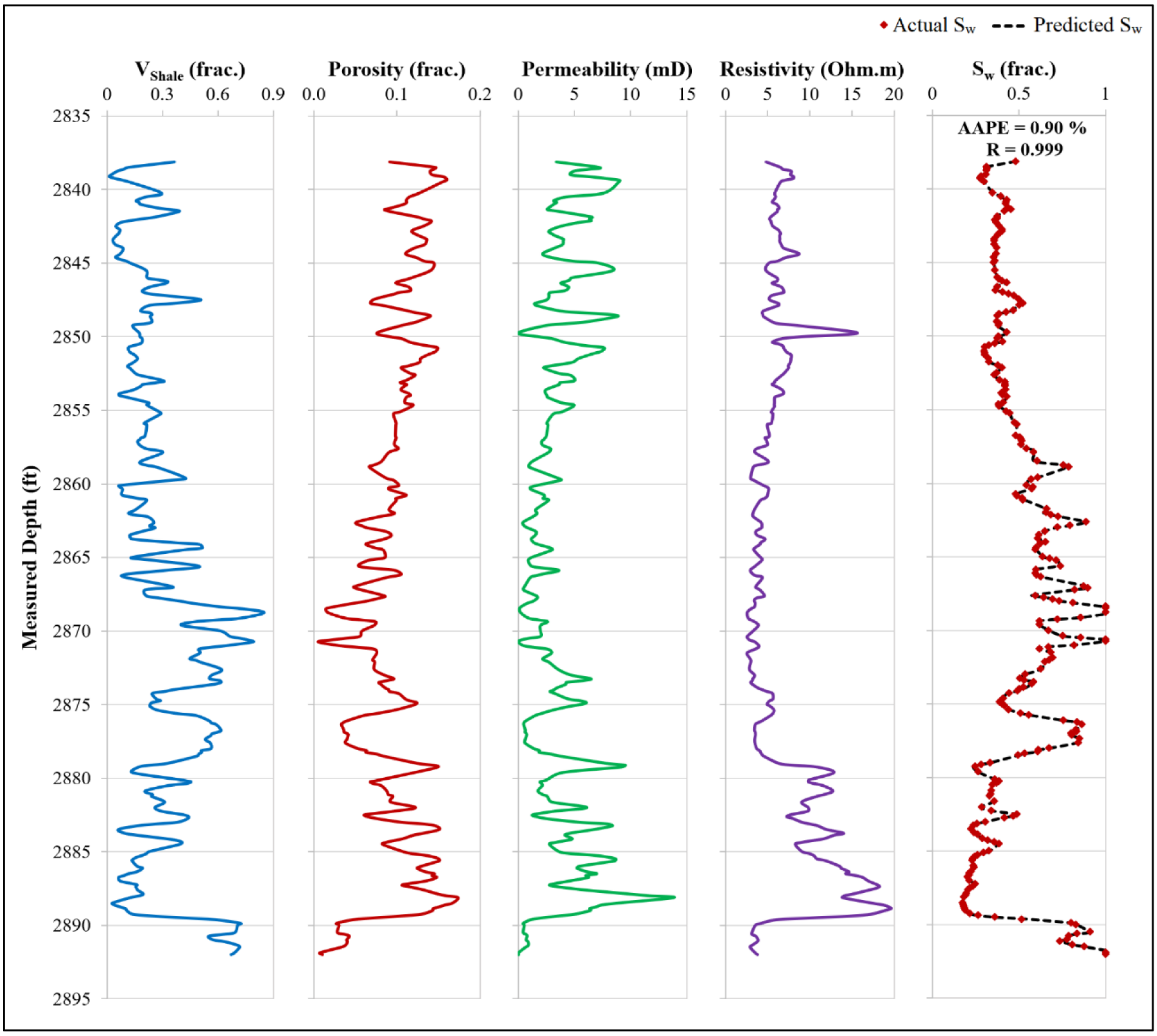


Figure 4. Training the ANN model to predict the water saturation from the shale volume, porosity, permeability, and resistivity logging data. The permeability is calculated based on well log data and not measured permeability.

The normalized parameters in eq 12 can be computed by applying the formulas provided in eq 8 to eq 11. The corresponding weights and biases required for this calculation were obtained from the optimized ANN model and are detailed in Table 4.

eq 12 provides a normalized water saturation, which needs to be denormalized to determine the actual water saturation. To achieve this, eq 7 is rearranged to solve for water saturation, substituting it in place of the variable Y. This manipulation results in eq 13, which can then be used to calculate the denormalized water saturation.

$$S_w = \frac{(S_{w,n} + 1)}{2}(S_{w,\max} - S_{w,\min}) + S_{w,\min} \tag{13}$$

By substituting the minimum and maximum water saturation values from Table 3, along with the normalized water saturation from eqs 12 ,13, eq 14 can be derived.

$$S_w = 0.413\left(\sum_{i=1}^{I}\left[\frac{w_{o_i}}{1 + e^{-(w_{t_{i,1}}V_{\text{shale},n} + w_{t_{i,2}}\Phi_n + w_{t_{i,3}}K_n + w_{t_{i,4}}\text{Res}_n + b_{t_i})}}\right] + 4.302\right) + 0.174 \tag{14}$$

3.4. Testing the Developed Empirical Equation for Water Saturation Estimation. The developed equation for estimating water saturation, eq 14, was tested using 82 data points from Well-1. The accuracy of the equation was assessed using several evaluation metrics, including the average absolute percentage error (AAPE), mean squared error (MSE), root mean squared error (RMSE), and the correlation coefficient. Additionally, a visual comparison was conducted to examine the difference between actual and predicted water saturation values. These metrics provided a comprehensive analysis of the model's performance, ensuring that the predictions made by eq 14 were reliable and accurate.

4. RESULTS AND DISCUSSION

ANN model design parameters were optimized using 334 data sets of the well logs input parameters of shale volume, deep formation resistivity, porosity, and permeability. 70% of the data (252 data sets were used for learning the ANN model, while the remaining 30% of the data was considered for testing the developed empirical equation. The following sections will discuss the results of training the ANN model and testing the modeled equation for water saturation prediction.

4.1. Training the Artificial Neural Networks Model. Figure 4 illustrates the key input parameters used to train the

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

ANN model to predict water saturation. From left to right, these parameters include shale volume, porosity, calculated permeability, and deep formation resistivity. The rightmost track of this figure compares the observed water saturation (depicted as dark red diamonds) and the water saturation profile predicted by the optimized ANN model, which is a function of the well log inputs, represented by the black dotted line. This predictive model is underpinned by the fundamental understanding that these input parameters collectively influence water saturation.

Figure 5 shows the water saturation track, demonstrating the remarkable concordance between the actual and predicted water saturation values. This alignment underscores the robustness of the ANN model's predictions, which are rooted in the scientific principles governing the interplay of shale volume, porosity, permeability, and deep formation resistivity. The accompanying cross plot in Figure 5 provides further validation, illustrating that all data points closely adhere to the 45° line. This is a testament to the model's accuracy and ability to account for the complex interactions between the input parameters and water saturation. The low AAPE of 0.90% for this training data set (comprising 252 data points) underscores the reliability of the water saturation predictions. At the same time, the high correlation coefficient of 0.999 signifies the strong relationship between the input logs and water saturation, thus affirming the scientific soundness of the ANN model's results. This figure also shows that the MSE and RMSE of the predicted values were 0.57 and 0.75%, respectively.

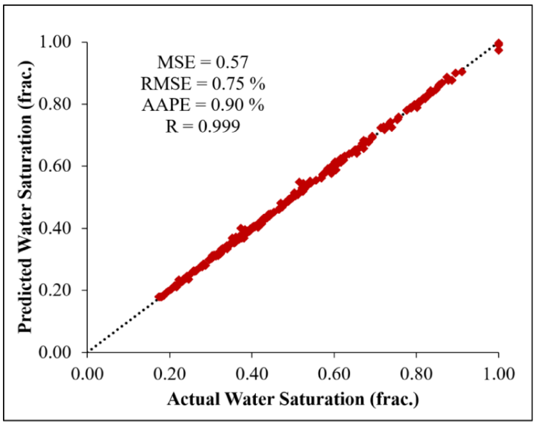


Figure 5. Crossplot of the training data set's actual and predicted water saturation (252 data points).

4.2. Testing the Developed Empirical Equation. The empirical equation developed for predicting water saturation (referred to as eq 14 was subjected to rigorous testing on 30% of the available data sets, totaling 82 data sets. This evaluation was rooted in sound scientific principles to ascertain the equation's effectiveness in capturing variations in water saturation.

Figure 6 presents the input log data, revealing the parameters influencing water saturation. These inputs were carefully selected to encompass the key geological and petrophysical factors that significantly affect water saturation. As can be discerned from this figure, eq 14 produced highly accurate predictions of water saturation. The ability of the equation to do so is underpinned by the underlying scientific rationale that these input parameters represent the complex processes governing water saturation in geological formations. This accuracy is reaffirmed by the remarkably low AAPE of 1.08%, which reflects the minimal deviation of predicted values from observed values. The high correlation coefficient of 0.999 further strengthens the scientific credibility of eq 14, signifying a robust and precise relationship between the input parameters and water saturation. Notably, there is a perfect alignment between the actual and predicted water saturation, providing visual evidence of the equation's reliability and consistent adherence to scientific principles.

Figure 7, with its cross plot, serves as an additional scientific validation, depicting the agreement between observed and estimated water saturation values. Here, all data points tightly adhere to the 45° line, further establishing the equation's high accuracy and ability to accommodate the complex interplay between the input parameters and water saturation. This alignment substantiates the scientific foundation of eq 14 and underscores its efficacy in predicting water saturation with a high degree of precision. As indicated in this figure also the MSE is 1.08 and the RMSE was 1.17%.

The future scope of this study holds significant potential for advancing reservoir management and the broader field of oil and gas exploration. By delving deeper into this research, the development of more refined prediction models for water saturation in diverse geological settings is expected. These models can have transformative implications for the oil and gas industry by offering heightened resource management and production optimization precision.

Furthermore, the successful application of ANNs in predicting water saturation in a Greek oilfield opens doors to adopting ANNs in similar geological contexts worldwide. Researchers may also investigate implementing different artificial intelligence and machine learning tools for water saturation estimation. This expansion of applicability may lead to a paradigm shift in hydrocarbon exploration practices, making them more adaptable and accurate. Additionally, the potential integration of emerging technologies and data sources could further enhance the predictive power of ANNs, presenting exciting opportunities for future research in reservoir characterization and management. The findings from this study, coupled with continued research and innovation, have the potential to shape the future of hydrocarbon exploration, yielding profound benefits for the energy industry and global sustainability.

5. CONCLUSIONS

This study presents the implementation of the artificial neural network (ANN) technique to estimate water saturation using conventional logging data, including well logs of shale volume, porosity, permeability, and deep formation resistivity; this makes water saturation very fast and cheap since none of the inputs require any laboratory measurement. A new empirical correlation for estimating water saturation has been formulated, utilizing the weights and biases derived from the optimized ANN model. The following are the main findings:

●Sensitivity analysis was performed during the ANN learning stage. As a result, the optimum design parameter of the ANN was selected as a combination of the parameters that were able to predict the water saturation with the minimum error.

●For training the data, the correlation coefficient value between the actual and estimated water saturation based on the ANN model was 0.999, which is the same as the coefficient between actual and estimated water saturation using the developed ANN-based empirical correlation.

●Water saturation was predicted using the optimized ANN model with an AAPE of 0.90% for training data.

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

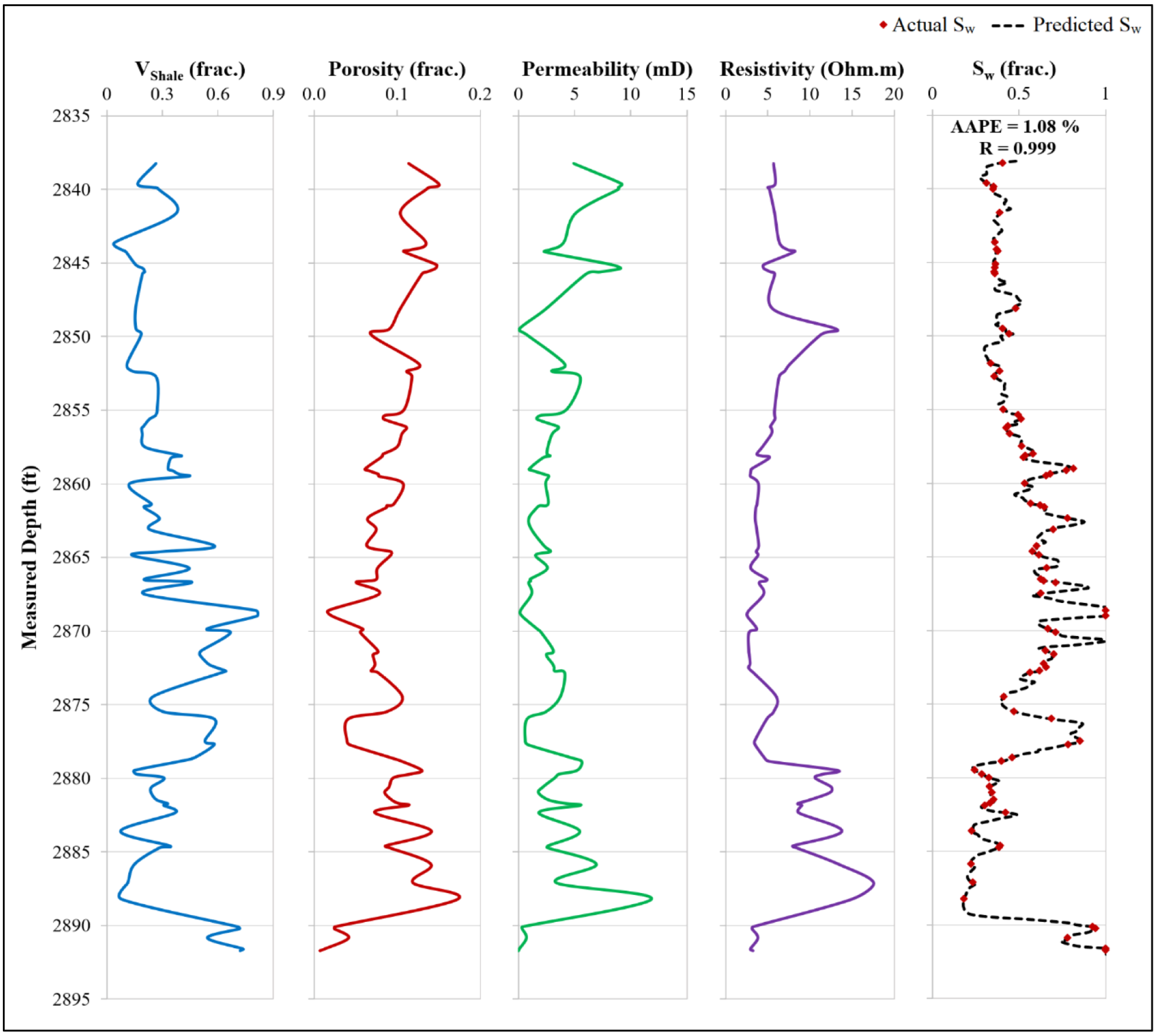


Figure 6. Testing the ANN model to predict the water saturation from the logging data. The rightmost track compares the actual water saturation with those predicted by eq 14 developed in this study.

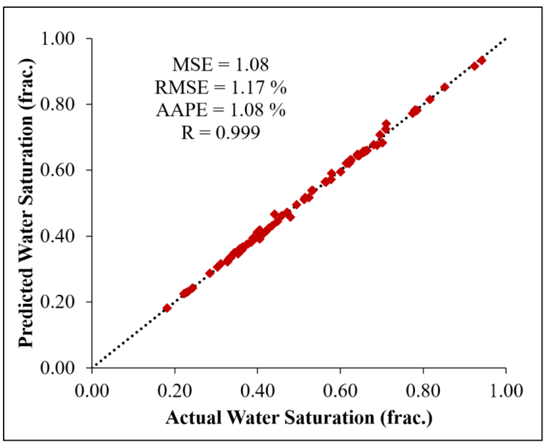


Figure 7. Cross plot of actual and predicted water saturation for the testing data set (82 data points).

●The newly developed empirical equation for water saturation estimation was validated using a separate data set that was not involved in the training process. The resulting correlation coefficient between the actual and the predicted water saturation values was an impressive 0.999, with an AAPE of 1.08%.

●This research highlights the effectiveness of the ANN-based approach and the resulting empirical correlations in accurately predicting water saturation from conventional well log data. The near-perfect correlation coefficient of 0.999 and the minimal error of 1.08% demonstrate the excellent fit and reliability of the model in practical applications.

●The developed correlation can help geologists and reservoir engineers to accurately predict the water saturation and its distribution along the depth of the drilled formations. This is very important since water saturation is an input parameter needed by geologists and reservoir engineers to define the net pay, and estimate the original oil in place and oil reserves. The accurate estimation of water saturation is also critical for completion engineers since it is important for optimizing the design of the reservoir completion (i.e., optimum selection of the preformation intervals). Hence, this will help the production engineers to maximize the oil recovery by reducing water encroachment.

●The developed empirical correlation will not only improve the accuracy of predicting the water saturation, but it is also a

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

noncostly method of obtaining a continuous water saturation profile, which is important for oilfield management and hydrocarbon exploration.

●Incorporating methods like Bayesian optimization or grid search in future studies would streamline the process and potentially improve the model's performance.

AUTHOR INFORMATION

Corresponding Author

Ahmed Abdulhamid Mahmoud − *Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; orcid.org/0000-0002-6703-9001; Phone: 966 (013) 860-8946; Email: ahmed.mahmoud@kfupm.edu.sa

Authors

Mohammed Gad − *Department of Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

George Panagopoulos − *Energean Oil & Gas S.A., Kifissias 11111, Greece*

Paschalia Kiomourtzi − *Energean Oil & Gas S.A., Kifissias 11111, Greece*

Panagiotis Kirmizakis − *Center for Integrative Petroleum Research, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

Salaheldin Elkatatny − *Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; orcid.org/0000-0002-7209-3715

Umair bin Waheed − *Department of Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

Pantelis Soupios − *Department of Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c07175

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge CPG for technical and financial support and Energean Oil & Gas S.A. for sharing the data and accepting publishing the results. Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

(1) Dandekar, A. Y. *Petroleum Reservoir Rock and Fluid Properties*; CRC Press, 2013. .

(2) Ahmed, T. *Reservoir Engineering Handbook*; Gulf Professional Publishing, Linacre House: Jordan Hill, Oxford OX2 8DP, UK, 2006.

(3) Dake, L. P. *Fundamentals of Reservoir Engineering*; Elsevier Science B.V.: Amsterdam, The Netherlands, 1998.

(4) Kamel, M. H.; Mabrouk, W. M. Estimation of shale volume using a combination of the three porosity logs. *J. Pet. Sci. Eng.* **2003**, *40*, 145−157.

(5) Archie, G. E. The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Trans. AIME* **1942**, *146*, 54−62.

(6) McPhee, C.; Reed, J.; Zubizarreta, I. *Core Analysis: a Best Practice Guide*; Elsevier, 2015.

(7) Tariq, Z.; Hassan, A.; Waheed, U. B.; Mahmoud, M.; Al-Shehri, D.; Abdulraheem, A.; Mokheimer, E. M. A.; et al. A Data-Driven Machine Learning Approach to Predict the Natural Gas Density of Pure and Mixed Hydrocarbons. *J. Energy Resour Technol.* **2021**, *143* (9), 092801.

(8) Xiao, L.; Mao, Z.; Li, G.; Jin, Y. Estimation of Saturation Exponent from Nuclear Magnetic Resonance (NMR) Logs in Low Permeability Reservoirs. *Appl. Magn. Reson.* **2013**, *44*, 333−347.

(9) Pickett, G. R. Acoustic Character Logs and Their Applications in Formation Evaluation. *J. Pet. Technol.* **1963**, *15*, 659−667.

(10) Kokkedee, J. A.; Boom, W.; Frens, A. M.; Maas, J. G. Improved Special Core Analysis: Scope for a reduced residual oil saturation. In *Society of Core Analysis ConferencePaper*, Jgmaas: 1996; Vol: *9601*, 1-13.

(11) Stiles, J. H., Jr; Hutfilz, J. M. The Use of Routine and Special Core Analysis in Characterizing Brent Group Reservoirs, U.K. North Sea. *J. Pet. Technol.* **1992**, *44*, 704−713.

(12) Elzain, H. E.; et al. Modeling of aquifer vulnerability index using deep learning neural networks coupling with optimization algorithms. *Environ. Sci. Pollut. Res.* **2021**, *28*, 57030−57045.

(13) Sircar, A.; Yadav, K.; Rayavarapu, K.; Bist, N.; Oza, H. Application of machine learning and artificial intelligence in oil and gas industry. *Pet. Res.* **2021**, *6*, 379−391.

(14) Siddig, O.; Gamal, H.; Soupios, P.; Elkatatny, S. Utilization of adaptive neuro-fuzzy interference system and functional network in prediction of total organic carbon content. *SN Appl. Sci.* **2022**, *4* (1), 16.

(15) Siddig, O.; Mahmoud, A. A.; Elkatatny, S.; Soupios, P. Utilization of Artificial Neural Network in Predicting the Total Organic Carbon in Devonian Shale Using the Conventional Well Logs and the Spectral Gamma Ray. *Comput. Intell Neurosci.* **2021**, *2021* (1), 2486046.

(16) Armaghani, D. J.; Asteris, P. G. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Comput. Appl.* **2021**, *33*, 4501−4532.

(17) Mahmoud, A. A.; Elkatatny, S.; Abdulraheem, A. Machine Learning Applications in the Petroleum Industry. In *Handbook of Energy Transitions*; CRC Press: Boca Raton, Florida, United States, 2022.

(18) Alkinani, H. H; Al-Hameedi, A. T.; Dunn-Norman, S.; Flori, R. E.; Alsaba, M. T.; Amer, A. S.Applications of Artificial Neural Networks in the Petroleum Industry: A Review. In *SPE Middle East Oil and Gas Show and Conference*; OnePetro: SPE, Manama, Bahrain, 2019;

(19) Al-AbdulJabbar, A.; Mahmoud, A. A.; Elkatatny, S. Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations. *Arab J Geosci.* **2021**, *14* (2), 1−14.

(20) Al-Bulushi, N. I.; King, P. R.; Blunt, M. J.; Kraaijveld, M. Artificial neural networks workflow and its application in the petroleum industry. *Neural Comput. Appl.* **2012**, *21*, 409−421.

(21) Mahmoud, A. A.; Elkatatny, S.; Ali, A.; Moussa, T. Estimation of static Young's modulus for sandstone formation using artificial neural networks. *Energies* **2019**, *12*, 2125.

(22) Castillo, E.; Cobo, A.; Gutiérrez, J. M.; Pruneda, E. Functional Networks: A New Network-Based Methodology. *Computer-Aided Civil And Infrastructure Engineering* **2000**, *15*, 90−106.

(23) Alsaihati, A.; Elkatatny, S.; Mahmoud, A. A.; Abdulraheem, A. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities while Drilling Horizontal Wells, with Real Case Study. *J. Energy Res. Technol.* **2021**, *143* (4), 043201.

(24) Mahmoud, A. A.; Elkatatny, S.; Alsabaa, A.; Al Shehri, D. Functional Neural Networks-Based Model for Prediction of the Static Young's Modulus for Sandstone Formations. In *The 54th US Rock Mechanics/Geomechanics Symposium*; ARMA: Golden, Colorado, USA, 2020.

(25) Acar, M. C.; Kaya, B. Models to estimate the elastic modulus of weak rocks based on least square support vector machine. *Arab J Geosci.* **2020**, *13* (14), 590.

(26) Durgesh, S.; Lekha, B. Data classification using support vector machine. *J. Theor Appl. Inf Technol.* **2010**, 1−7.

(27) Hegde, C.; Wallace, S.; Gray, K.Using Trees, Bagging, and Random Forests to Predict Rate of Penetration During Drilling. In *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*, OnePetro: SPE, Abu Dhabi, UAE, 2015; SPE-176792. .

https://doi.org/10.1021/acsomega.4c07175
ACS Omega 2025, 10, 557−566

(28) Osman, H.; Ali, A.; Mahmoud, A. A.; Elkatatny, S. Estimation of the rate of penetration while horizontally drilling carbonate formation using random forest. *J. Energy Resour. Technol.* **2021**, *143* (9), 093003.
(29) Akkem, Y.; Biswas, S. K.; Varanasi, A. Smart farming using artificial intelligence: A review. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105899.
(30) Akkem, Y.; Biswas, S. K.; Varanasi, A. Smart Farming Monitoring Using ML and MLOps. In *International Conference on Innovative Computing and Communications*; Springer: Singapore, 2023; 665-675.
(31) OĞuz, A.; ErtuĞrul, Ö. F. A survey on applications of machine learning algorithms in water quality assessment and water supply and management. *Water Suppl.* **2023**, *23*, 895−922.
(32) Sanjeev, D.; Singh, K.; Craciun, E.-M.; Rabaea, A.; Batra, A. Next-Cart Recommendation by Utilizing Personalized Item Frequency Information in Online Web Portals. *Neural Process Lett.* **2023**, *55*, 9409.
(33) Kiomourtzi, P. *Depositional environments and hydrocarbon potential basins in the North Aegean Sea*; University of Patras, 2016. .
(34) Kiomourtzi, P.; Pasadakis, N.; Zelilidis, A. Source Rock and Depositional Environment Study of Three Hydrocarbon Fields in Prinos - Kavala Basin (North Aegean). *Open Petr. Engineer. J.* **2008**, *1*, 16−29.
(35) Proedrou, P.; Papaconstantinou, M. C. Primus basin-a model for oil exploration. *Bull. Geol. Soc. Greece* **2004**, *36*, 327−333.
(36) Pollak, W. H. *Structural and lithological development of the Prinos-Kavala basin, sea of Thrace, Greece*; Annales Geologique des Pays Hellenic: Athens, Greece, 1979.
(37) Proedrou, P. The Evaporite formation in the Nestos-Prinos graben in the northern Aegean sea. *Ann. Ge¤ ol. Pays Hell* **1979**, 1013−1020.
(38) Proedrou, P.; Sidiropoulos, T.*Prinos field − Greece Aegean basin*AAPG Special Volumes1992275−291
(39) Choustoulakis, E. *Detection of gas in sandstone reservoir using AVO analysis in Prinos basin*; Technical University of Crete, 2015.
(40) Kiomourtzi, P.; Pasadakis, N.; Zelilidis, A. Geochemical characterization of satellite hydrocarbon formations in Prinos-Kavala basin (North Greece). *Bull. Geol. Soc. Greece.* **2007**, *40*, 839.
(41) Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431−441.
(42) Hagan, M. T.; Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **1994**, *5*, 989−993.
(43) Masapanta, J. *Machine and Deep Learning for Lithofacies Classification from Well Logs in the North Sea*; University of Stavanger, 2021.