Novel Correlation for Calculating Water Saturation in Shaly Sandstone Reservoirs Using Artificial Intelligence: Case Study from Egyptian Oil Fields

Reda Abdel Azim* and Ghareb Hamada*

Cite This: ACS Omega 2022, 7, 29666−29674

ACCESS

ABSTRACT: The accurate determination of water saturation in shaly sandstone reservoirs has a significant impact on hydrocarbons in place estimation and selection of possible hydrocarbon zones. The available numerical equations for water saturation estimation are unreliable and depend on laboratory core analysis. Therefore, this paper attempts to use artificial intelligence methods in developing an artificial neural network model (ANN) for water saturation (Sw) prediction. The ANN model is developed and validated by using 2700 core measured points from the fields located in the Gulf of Suez, Nile Delta, and Western Desert of Egypt, with inputs including the formation depth, the caliper size, the sonic time, gamma rays (GRs), shallow resistivity (Rxo), neutron porosity (NPHI), the photoelectric effect (PEF), bulk density, and deep resistivity (Rt). The study results show that the optimization process for the ANN model is achieved by distributing the collected data as follows: 80% for training and 20% for testing processes, with an $R^2$ of 0.973 and a mean square error (MSE) of 0.048. In addition, a mathematical equation is extracted out of the ANN model that is used to estimate the formation water saturation in a simple and direct approach. The developed equation can be used incorporating with the existing well logs commercial software to increase the accuracy of water saturation prediction. A comparison study is executed using published correlations (Waxman and Smits, dual water, and effective models) to show the robustness of the presented ANN model and the extracted equation. The results show that the proposed correlation and the ANN model achieved outstanding performance and better accuracy than the existing empirical models for calculating the formation water saturation with a high correlation coefficient ($R^2$) of 0.973, lowest mean-square error (MSE) of 0.048, lowest average absolute percent relative error (AAPRE) of 0.042, and standard deviation (SD) of 0.24. To the best of our knowledge, the current study and the proposed ANN model establish a novel base in the estimation of formation water saturation.

1. INTRODUCTION

Clays are considered significant among the constituents of rocks by log analysis, as they form 40−50% of the mineral components of shale. Clay minerals are structured as sheets of silica tetrahedron lattices. Within the clay sheets, there are usually excessive negative electrical charges. This resulted in a local electrical imbalance within clay particles. The Archie\(^1\) water saturation equation considers the formation water as the only electrically conductive material in the formation. The presence of conductive clay materials requires modifying the Archie equation or generating new models to incorporate the rock resistivity. Over the years, different models have been proposed to relate the fluid saturation and resistivity, assuming that shale exists in specific geometric forms (laminated, dispersed, and structural). These model parameters contain a clean sand term defined by Archie\(^1\) and a clay term, as shown in eq 1.

\[
\frac{1}{R_o} = \frac{1}{FR_w} + \text{clay term}
\]

where $R_o$ is the rock resistivity saturated with water, ohm m; $F$ is the formation factor, dimensionless; and $R_w$ is the water resistivity, ohm m.

All these models follow the clean sand Archie equation when the clay fraction is zero; for small amounts of shale (5−10%), most models yield quite similar results.

The presence of clay minerals in a sandstone reservoir leads to reduction in the reservoir’s storage and reduces the reservoir’s ability to transmit fluids by reducing the porosity and permeability, respectively.

The clay minerals’ occurrence in sandstone with freshwater leads to an overly pessimistic water saturation value unless...
corrections are made. Moreover, the presence of clay minerals on salty formation water makes the recorded deep resistivity (Rt) too low, and this led to an increase in the values of formation water saturation. Both cases lead to bypassed production as these zones will be considered erroneous, as noncommercial. Therefore, to determine the accurate water saturation in shaly sand oil reservoirs, the Archie water saturation equation must be modified, but unfortunately no adequate model exists that accounted for the fundamental electrical behavior in shaly sands. Consequently, the entire water saturation equations available for shaly sands are of empirical nature.

The well-known shaly sand models for water saturation are the Waxman and Smits model, effective medium model, and Simandoux model. These models are selected to be used in the comparative study in this work based on the data collected from Egypt oil fields, as these models are widely used by petrophysicists in water saturation calculation in such fields.

The Simandoux model is based on laboratory experiments for conductivity estimation.

\[
C_t = C_w F_s S_w \cdot V_{sh} C_{sh} \tag{2}
\]

where \( C_{sh} \) is the shale conductivity, and \( C_w \) is the water conductivity.

The main drawback of this model is that it yields water saturation values quite low in shaly sand formation containing clay minerals.

The Waxman and Smits model is based on the laboratory measurements of cation exchange capacity (CEC), and the main drawback of this model is not available during the log analysis. This model has been modified by Juhasz. Juhasz used the formation density log to estimate porosity, while the dual water model uses the neutron density cross plot. The type of clay is included in the Waxman and Smits model in the estimation of effective porosity.

The general equation for water-saturated sands is then obtained

\[
C_0 = \frac{1}{F} \cdot (BQ_v + C_w) \tag{3}
\]

with

\[
B = [1 - a \cdot \exp(-C_w')] \cdot 0.001 \lambda_{Na} \tag{4}
\]

where \( C_o \) is the specific conductance of sand, \( C_w \) is the specific conductance of the brine, \( \lambda_{Na} \) is the maximum equivalent ionic conductance of sodium exchange ions, \( \gamma \) is the empirical constant, and \( Q_v \) is the effective concentration of clay exchange cations.

A simple method to determine \( Q_v \):

\[
Q_v = (CEC \times (1 - \Phi)) / 100 \times \Phi \tag{5}
\]

The dual water model is indirectly based on the cation exchange capacity. Basically, the pore volume is divided into bound water (Swb) or shale water and free water, also called sand water (Swf). Bound and free waters contributed to the conductivity of the shaly sand. Both these waters have their own resistivity, Rw and Rwf, respectively. The amount of bound water is directly related to the volume of clay in the shaly sand formation. With an increase in clay volume, more bound water will occupy the total pore space (\( \Phi \)).

Archie’s formula can be written as follows:

\[
C_t = (S_{wT} - \Phi)^2 \cdot C_{wM} \tag{6}
\]

where \( C_t = 1/R_t \) is the formation conductivity, and \( C_{wM} \) is the conductivity of mixed water (bound and free).

\[
C_{wM} = \frac{S_{wb} \Phi_t}{S_{wt}} + \frac{(S_{wt} - S_{wb}) \Phi_t}{S_{wt}} \cdot C_{wF} \tag{7}
\]

\[
C_{wM} = \frac{S_{wb} \cdot C_{wb} + (S_{wt} - S_{wb}) \cdot C_{wF}}{S_{wt}} \tag{8}
\]

where \( \Phi_t \) is calculated from LDT-CN [\( \Phi_t = (\Phi_b + \Phi_{Na})/2 \)]; \( S_{wb} \) and \( S_{wF} \approx V_d \) from clay indicators; \( R_{wF} \) from the water zone; and \( R_{wB} \) from the 100% water zone.

The effective medium model is a theoretical one used to calculate water saturation based on porosity and resistivity measurements. The model can be used for different types of clay minerals including laminated and dispersed shale distributions. The model based on the assumption that the matrix and hydrocarbons can be treated together is as follows:

\[
Sw \Phi = \left( \frac{R_w}{R_t} \right)^{1/m} \left( \frac{R_t - R_d}{R_w - R_d} \right) \tag{9}
\]

where \( R_d \) is the dispersed phase resistivity (the combination of the matrix and hydrocarbons).

\[
R_d = \left( \frac{R_t}{1 - \Phi} \right) (1 - Sw \Phi) \tag{10}
\]

The literature review showed that several water saturation models are developed to estimate the formation water saturation in the presence of clay minerals. These models have drawbacks, and none of them has the ability to predict precisely the water saturation in the hydrocarbon zones. The main weakness in the published correlations is their strong dependency on the experimental core analyses that requires a long experimental time. Hence, this study aims to develop a new ANN model with a supervised algorithm to accurately estimate the water saturation profile in shaly sand oil fields located in Egypt from conventional well log data.

In order to fulfill this purpose, more than 2700 datasets of core points from numerous Egyptian oil fields located in Western Desert, Nile Delta, and Gulf of Suez are collected. The data include the measured petrophysical properties to be used in the developed model. The new presented correlation incorporates new parameters including gamma rays, the caliper size, and photoelectric effect (PEF).

2. ARTIFICIAL NEURAL NETWORK

In this study, the back-propagation algorithm (BP) is used as a learning process to supervise the neural net. The forward step is used to send signals to the input layers, and the backward one is used to calculate the proposal error between the field and target outputs. Weights are used for each layer to adjust the mean-square error (MSE) during the backward operation, as follows

\[
\text{MSE} = \frac{1}{n_1 \times n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (x_p - y_p)^2 \tag{11}
\]

where \( n_1 \) and \( n_2 \) are the number of training and output neurons, respectively, and \( x_p \) and \( y_p \) are the target and estimated outputs,
Table 1. Statistical Analysis of the Data Used To Validate the ANN Model

| parameter                     | formation depth (DEPPTH), ft | caliper size (inch) CALI | sonic time (μs/ft) DTR | gamma ray (GR), API | shallow resistivity (LLS), ohm m | neutron porosity (NPHI) | photo electric effect (PEF) | bulk density (RHOB), g/cc | deep resistivity (Rt), ohm m |
|-------------------------------|-----------------------------|--------------------------|-------------------------|---------------------|-------------------------------|-------------------------|--------------------------|-----------------------------|----------------------------|
| min                           | 7870                        | 8.04                     | 45.364                  | 7.88                | 0.3756                        | 0.006                   | 0.036                    | 1.321                       | 0.33                       |
| max                           | 9430                        | 25.836                   | 109.78                  | 138.25              | 255.8                         | 65.14                   | 10                       | 5.70                        | 231.01                     |
| standard deviation            | 460.179                     | 2.6786                   | 12.755                  | 29.460              | 17.708                        | 3.99                    | 1.94                     | 0.5491                      | 15.941                     |
| skewness                      | 1.1862                      | 3.0877                   | -0.1815                 | 0.0967              | 6.956                         | 7.84                    | 0.794                    | 1.729                       | 7.470                      |
| mean                          | 8267                        | 8.642                    | 74.715                  | 56.696              | 3.7466                        | 0.209                   | 3.1718                   | 2.602                       | 2.8117                     |

 respected. The employed function used in the back-propagation algorithm in this study is a sigmoid curve called the logistic function.

The network speed of the convergence process can be improved by adding an acceleration technique as follows:

$$w(t + 1) = w(t) + \beta[\Delta w(t)] + \alpha[w(t - 1)]$$  \hfill (12)

where $\alpha$ is the energy constant, $w$ is the weight, and $\Delta w$ is the difference weight. The learning and momentum constants are set in the range of 0 and 1.

3. APPLICATION OF ANN TO PREDICT WATER SATURATION

Neural network systems have become increasingly popular in engineering applications. This is partly due to the fact that intelligent animals can solve problems which are impossible for even the most powerful modem computers and partly because of the desire by engineers and computer scientists to explore and exploit parallel hardware systems and apply them to solve practical problems. In petroleum engineering, successful applications include drill bit diagnosis, practical problems. In petroleum engineering, successful applications include drill bit diagnosis, seismic processing, identification of well test interpretation model, flow measurements, identification of well productivity, and wireline log analysis. Also, McCormack and Day and Fogelman-Soulie provided some introductory articles on the use of neural networks in the petroleum industry. Artificial intelligence provides numerous benefits for petrophysical evaluation. Several researchers have used ANN models, particularly feed-forward back-propagation neural networks (FFNNs), to develop more accurate predictions of the reservoir rock properties that include water saturation and porosity. An AI algorithm is developed for shaly sandstone reservoirs to predict water saturation, with the mean-square error (MSE) of 0.064. Hamada et al. used PSONN as an optimization algorithm for water saturation estimation for clean sandstone formations, and their results show that the new hybrid PSONN model outperforms some available methods with a lower root-mean-square error of 0.009 and an $R^2$ of 0.95. Aydin et al. proposed a model for forecasting coal consumption in Turkey. The data used in Aydin’s model are divided into two groups for training and testing processes, and the results show MSE achieved of 0.025.

An artificial intelligence model is proposed to calculate water saturation for two reservoirs in the Middle East. The model is based on a three-layer neural net to predict saturation in the formation and yields a correlation factor of 0.91 with an error of 0.025. Amiri et al. (2015) proposed an ANN model with different well log data. The results showed that the model is precise for the forecasting process with the correlation factor $R^2$ of 0.97. An artificial neural network model is proposed to evaluate the saturation of water in gas reservoirs based on competitive algorithm. A total of 2200 data points taken from 12 wells have been collected to build the model. The results indicated that the model is efficient. The developed ANN framework consists of four different structures based on the tan sigmoid function to predict water saturation from the well log data. The results show that the proposed model is more robust that the dual water model. Kamalyar et al. (2011) proposed an ANN methodology for water saturation prediction for oil wells in southern Iran. Permeability, porosity, and other well logs were used as the input data, and saturation of water profile was the target. Helle and Bhatt (2002) presented an ANN model based on nine trained neural networks. The input data used in their model include density, resistivity, and sonic logs, and saturation was the target. Numerous studies have been presented for water saturation calculation based on the regression method. Gomaa et al. (2022) proposed an ANN model for SW estimation using 383 core samples. The input data used in the model include porosity, permeability, and resistivity index. The results show that the ANN model gives a coefficient of determination of 0.99 and an average relative error of 0.13, and MSE = 0.066.

Most of the previous attempts rarely provide a single equation to determine water saturation from well logs to overcome the drawbacks in the existing empirical models. Consequently, the main aim of this study is to provide a novel correlation using the ANN model. Furthermore, in this study, a new convergence technique is provided to rapidly predict the target data by adding two new parameters in the ANN methodology: step size and momentum.

The traditional ANN works that have been presented in the literature exhibit a drawback in the convergence process. Thus, this study presents two new parameters used to speed up and overcome the abovementioned problems, namely, step size and momentum. Furthermore, the learning rate is incorporated with the BP algorithm.

4. COLLECTED DATA ANALYSIS

4.1. Description. Groups of datasets from different fields in Egypt were used in developing the ANN model. The data comprise nine inputs used for the training: the formation depth (DEPPTH), the caliper size (CALI), the sonic time (DTR), gamma rays (GRs), shallow resistivity (LLS), neutron porosity (NPHI), the photoelectric effect (PEF), bulk density (RHOB), and deep resistivity (Rt). Data are normalized in a range of 0 and 1 (see Table 1).

4.2. Data Acquisition and Analysis. 4.2.1. Distribution of Input Data. The data are divided into training (80%) and testing groups (20%) by using a randomization function. The distribution is used to train the ANN model to create a correlation between inputs and formation water saturation. The learning algorithm that optimizes the training data by reducing

29668 https://doi.org/10.1021/acsomega.2c01945
ACS Omega 2022, 7, 29666−29674
the error between the target and actual water saturation is back-
propagation (BP). The BP learning algorithm provides exceptional results with an $R^2$ of 0.973 and MSE = 0.048 compared to other algorithms, including scaled conjugate gradient (SCG) and one-step secant (OSS), as shown in Figure 1.

4.2.2. Optimization for ANN Neurons. Levenburg Mar-
quardt algorithm \textsuperscript{32} is used to optimize the number of neurons and hidden layers. It can be seen from Figure 2 that the optimal neuron number is 15 with MSE = 0.048.

Figure 2. Mean-square error versus number of neurons tested during the training and testing processes.

Figure 3 shows a flowchart of the steps involved in the ANN model used in this study to estimate the water saturation. First, the data are collected from conventional well logs. Next, numerous ratios are tested for the training and testing processes. Moreover, the ANN model parameters are optimized, including the number of hidden layers, number of neurons, learning constant, and training functions. Table 2 summarizes the optimized parameters used in this study.

Another test is executed to test how strongly output data (Sw) are related to input data (including depth, caliper size, sonic time, GR, LLS, NPHI, PEF, RHOB, and Rt) by using the correlation coefficient (CC). Figure 4 shows that Sw is strongly dependent on GR, NPHI, RHOB, PEF, and Rt, with $CC = 0.578$, 0.465, 0.385, 0.294, and $-0.142$, respectively. It can be seen from Figure 4 that Sw has a direct relationship with GR, NPHI, RHOB, and PEF and an inverse relationship with Rt. At

Table 2. Optimized Parameters for the ANN Model

| parameter         | tested range | optimized parameters |
|-------------------|--------------|----------------------|
| number of neurons | 2–35         | 15                   |
| hidden layer number | 1–3          | 1                    |
| algorithm function | tan sig/log sig | tan sig             |
| learning rate     | 0.001–0.8    | 0.03                 |

Figure 4. Correlation coefficients of water saturation vs inputs.
high GR, which indicates shale zones, water saturation increases with low Rt values.

5. RESULTS AND DISCUSSION

To optimize the ANN model parameters, a sensitivity analysis is performed including different runs to show the deviation between the actual measured core data and the predicted water saturation data. The nine inputs mentioned before give minimum errors with the highest correlation factor $R^2 = 0.973$.

Approximately 2000 points are used for the training process, and 700 data points are used for testing. During the training process, an iterative operation is executed using different number of neurons. The results show that using 10 neurons in the neural net gives an $R^2$ of 0.91 (see Figure 5a), while using 15 neurons gives an optimum $R^2$ of 0.973 (see Figure 5b), with MSE = 0.021. The results of the testing process show that the optimal $R^2$ of 0.965 and MSE = 0.035 are found with the use of 15 neurons, while 10 neurons give an $R^2$ of 0.91, as shown in Figure 6a,b, with MSE = 0.068. Figure 7 shows an exceptional match between the measured saturation of core points and ANN water saturation at the same depth for the testing process, which confirms the consistency of the created neural model and water saturation correlation.

Perceiving the encouraging outcomes out of the ANN operations, a mathematical equation is extracted to be used with a very simple approach to calculate the saturation of water at different depths. The weights and biases for the generated equation are given in Table 3.

The novel correlation generated using ANN for water saturation estimation in shaly sand reservoirs is given by

$$
Sw_w = \sum_{i=1}^{N} w_{2i} \tan \left( \sum_{j=1}^{J} w_{1i_j} x_j + b_1 \right) + b_2
$$

(13)
Table 3. Weights Used in the Extracted Correlation Eq 14

| neuron number | depth | caliper size | sonic time | gamma ray | LLS | NPHI | PEF | RHOB | Rt |
|---------------|-------|-------------|------------|-----------|-----|------|-----|------|----|
| 1             | $8.60 \times 10^{-2}$ | $-1.64366$ | $1.364198$ | $4.488277$ | $0.823984$ | $-3.76761$ | $-0.71577$ | $6.80 \times 10^{-2}$ | $2.37 \times 10^{-2}$ |
| 2             | $-0.33533$ | $0.69407$ | $-0.37064$ | $2.452354$ | $-1.78967$ | $0.410602$ | $0.570176$ | $-21.7014$ | $-0.46693$ |
| 3             | $-0.72663$ | $-0.7733$ | $1.05871$ | $2.158347$ | $0.124696$ | $2.74917$ | $0.236778$ | $-4.02574$ | $0.399239$ |
| 4             | $-1.28712$ | $-1.36386$ | $4.136454$ | $-7.21623$ | $-1.53894$ | $1.860875$ | $0.74278$ | $0.471668$ | $3.147989$ |
| 5             | $-0.30975$ | $1.913667$ | $-1.04754$ | $5.220827$ | $2.326796$ | $-3.79047$ | $-4.28611$ | $1.29184$ | $-0.35187$ |
| 6             | $-4.70811$ | $-0.19207$ | $1.949784$ | $-5.15739$ | $1.25263$ | $2.13902$ | $5.58273$ | $1.97966$ | $11.8251$ |
| 7             | $-4.2743$ | $2.073048$ | $-5.45202$ | $-2.09825$ | $2.974399$ | $-1.36232$ | $-1.17286$ | $-0.57869$ | $2.070786$ |
| 8             | $-2.78686$ | $0.850972$ | $0.604677$ | $0.37774$ | $-3.36826$ | $3.01271$ | $3.099444$ | $0.905125$ | $-12.0279$ |
| 9             | $0.756154$ | $2.578654$ | $1.50889$ | $3.33637$ | $1.529856$ | $0.134132$ | $0.751849$ | $-2.46792$ | $3.57177$ |
| 10            | $-0.76985$ | $0.816966$ | $1.435553$ | $3.299912$ | $-1.25607$ | $-16.8058$ | $-0.18043$ | $-1.91115$ | $-2.22478$ |
| 11            | $0.28348$ | $1.90887$ | $-0.57044$ | $-0.90216$ | $2.53042$ | $-2.54324$ | $-1.48602$ | $-0.46482$ | $1.13 \times 10^{-2}$ |
| 12            | $1.203564$ | $-0.40776$ | $-0.1572$ | $-0.37021$ | $-0.96029$ | $-0.25028$ | $-2.22131$ | $-0.90248$ | $3.32748$ |
| 13            | $1.370686$ | $4.117492$ | $-0.48873$ | $3.742478$ | $1.837794$ | $18.25063$ | $-1.79264$ | $-0.77339$ | $1.265298$ |
| 14            | $2.113377$ | $-1.51161$ | $15.37353$ | $-22.0491$ | $1.22313$ | $-2.77885$ | $1.126326$ | $23.15521$ | $-16.2084$ |
| 15            | $1.076056$ | $0.507737$ | $0.16187$ | $1.27 \times 10^{-2}$ | $-1.39666$ | $3.772604$ | $-1.12231$ | $2.583637$ | $0.196369$ |

weights ($w_1$)  bias ($b_1$)  bias ($b_2$)

-2.45044  -3.51022  8.92208
-0.322973  3.251176
3.860151  -10.325
-10.0577  -1.16598
-2.22 \times 10^{-3}  -2.07675
4.390237  5.96349
7.388246  15.15375
9.32 \times 10^{-2}  -4.54714
-10.7205  -15.1273
5.523283  -0.55698
-1.23776  5.067025
-0.13705  3.599462
-5.68593  7.51383
-0.66437  3.364801
-7.24734  -2.56502

5.1. Validation of the Developed ANN Model

In order to validate the newly proposed correlation for water saturation, the known correlations are used to predict the water saturation values at different formation depths. The published correlations are the bias vectors for the output layer, $b$, and $b_1$ are the bias vectors for the input and output layers, respectively. The gamma ray, LLS, NPHI, GR, PEF, RHOB, and Rt are the neutron porosity, photoelectric effect, bulk density, and deep resistivity, respectively. The gamma ray is the formation depth, LLS is the shallow resistivity, PEF is the photoelectric effect, RHOB is the bulk density, and Rt is the deep resistivity. The extraction of Sw is achieved by denormalizing Sw as follows:

$$ Sw = \frac{\sum w_i (\text{input}) + b}{\sum w_i (\text{output}) + b_1} $$

where Sw is the normalized water saturation, and Sw is the water saturation. The correlation is used to predict the water saturation values at different formation depths, and then a comparison against well-known correlations is performed. The published correlations are used for validation process. First, the unconfined data sets are used for the training process. Then, the correlation is used to predict the water saturation values at different formation depths, and then a comparison against well-known correlations is performed.
used in this comparison study are those of Waxman and Smits, dual water model, and effective model. The log data for a well including porosity and resistivity measurements are shown in Figure 8a,b.

Figure 9 shows the cross plots of the calculated water saturation using the proposed ANN model, and empirical correlations include Waxman and Smits, dual water, and effective models. The generated correlation is able to predict accurately the water saturation with an $R^2$ of 0.93, as shown in Figure 9a. Proceeding to validate the generated correlation in this study, the available models are used to estimate $S_w$ as well.

![Figure 9](image)

**Figure 9.** Scatter diagram comparing the predicted $S_w$ and actual $S_w$ using (a) ANN model, (b) Waxman and Smits model, (c) dual water model, and (d) effective model.

Figure 10. Core $S_w$ vs ANN-predicted values for validation data.

**Figure 10.** Core $S_w$ vs ANN-predicted values for validation data.

![Figure 10](image)

**Figure 11.** MSE for the ANN correlation, Waxman and Smits model, dual water model, and effective model.

**Figure 11.** MSE for the ANN correlation, Waxman and Smits model, dual water model, and effective model.

![Figure 11](image)

**Table 4. Statistical Analysis for the Used Correlations and Neural Model**

|          | AAPRE $^{15}$ | MSE | SD $^{15}$ | correlation coefficient |
|----------|---------------|-----|------------|-------------------------|
| Waxman and Smits | 0.1303        | 0.95 | 0.35       | 0.64                    |
| dual water    | 0.6012        | 1.52 | 1.51       | 0.457                   |
| effective model | 0.1419        | 3.52 | 0.324      | 0.065                   |
| Hamada et al. 2020 | 0.15          | 0.064 | 1.32       | 0.92                    |
| Sayed et al. 2022 | 0.13          | 0.066 | 21.5       | 0.99                    |
| this study: ANN | 0.042         | 0.042 | 0.24       | 0.973                   |

including porosity and resistivity measurements are shown in Figure 8a,b.

**Figure 9** shows the cross plots of the calculated water saturation using the proposed ANN model, and empirical correlations include Waxman and Smits, dual water, and effective models. The generated correlation is able to predict accurately the water saturation with an $R^2$ of 0.93, as shown in Figure 9a. Proceeding to validate the generated correlation in this study, the available models are used to estimate $S_w$ as well.
The results show that the match between the predicted and actual Sw is poor using the Waxman and Smits model, with an $R^2$ of 0.64 (see Figure 9b) and AAPRE = 0.13. The dual water model gives an $R^2$ of 0.457, while the effective model provides an $R^2$ of 0.065 (see Figure 9c,d). Out of the presented results, it has been resolved that the published correlations are not able to predict Sw precisely compared to the newly generated correlation. Figure 10 presents the performance of an ANN empirical correlation versus the published empirical correlations. From Figure 10, it can be seen that the ANN empirical correlation is able to detect the changes of water saturation with depth, with AAPRE of 0.042, while the Waxman and Smits model gives AAPRE = 0.13. The dual water model gives AAPRE of 0.60, and the effective model provides AAPRE of 0.14. Table 3 shows a comparison of the statistical information of the proposed correlation with other empirical correlations. It can be seen from Table 3 that the proposed ANN empirical correlation gives AAPRE of 0.048 and MSE of 0.042, less than that obtained from Sayed et al. (2022)$^{19}$ (AAPRE of 0.13 and MSE of 0.066) and Hamada et al.$^{19}$ (AAPRE of 0.15 and MSE of 0.064). Statistically, the proposed ANN model is consistent and robust, based on the presented statistical analysis in Table 4.

Figure 11 shows the MSE values between the real and predicted water saturation for the new ANN correlation: it shows the lowest MSE of 0.048 compared to that of the Waxman and Smits (MSE =0.95), dual water (MSE = 1.52), and effective (MSE = 3.52) models. The results show that the proposed ANN model is able to model a complex nonlinear relationship between the input and output variables. Based on the published literature and information collected from the Egypt fields, it is observed that the Waxman and Smits model is the most significant correlation used to estimate the saturation of water in shaly sand formation. Therefore, on testing the strength of the developed ANN correlation against the Waxman and Smits model, it is observed that using the ANN empirical correlation makes the selection of hydrocarbon zones and the accurate estimation of hydrocarbons in place more efficient.

6. CONCLUSIONS

1. In this study, the ANN technique is presented to propose a new correlation for water saturation estimation with an accuracy of 97%–98% in shaly oil reservoirs without using the ANN software. This correlation is generated for the reservoirs located in Gulf of Suez, Western Desert, and Nile Delta of Egypt, and other fields in the same locations having the same data range can use the generated correlation.

2. The ANN technique in this study uses the back-propagation learning algorithm with the new acceleration method for improving the convergence scheme. ANN results provide the optimum water saturation values with the lowest error and highest $R^2$ value.

3. The comparison performed in this study shows that literature correlations display severe lags in predicting the measured core water saturation. Accordingly, this study provides a solution for oil companies in Egypt to forecast precisely water saturation and fluid in place consequently.

4. The presented ANN model in this study is promising, and it should be evaluated further using large number of oil fields with various lithologies.

- **AUTHOR INFORMATION**

**Corresponding Authors**

Reda Abdel Azim — Petroleum Engineering Department, American University of Kurdistan, Duhok 40025, Kurdistan Region of Iraq; orcid.org/0000-0003-3793-3560; Email: reda.abdulrasoul@auk.edu.krd

Ghareb Hamada — Petroleum Engineering Department, American University of Kurdistan, Duhok 40025, Kurdistan Region of Iraq; Email: Ghareb.hamada@auk.edu.krd

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01945

**Author Contributions**

Reda Abdel Azim: conceptualization; developed the simulation code, data collection, and writing. Ghareb Hamada: review and editing.

**Notes**

The authors declare no competing financial interest.

- **ACKNOWLEDGMENTS**

The authors would like to thank American University of Kurdistan for providing the laboratory and computational facilities.

- **REFERENCES**

  1. Archie, G. E. The electrical resistivity log as an aid in determining some reservoir characteristics. *Trans. AIME* 1942, 146, 54–62.
  2. Waxman, M. H.; Smits, L. J. M. Electrical conductivities in oil-bearing shaly sands. SPE J. 1968, 8, 107–122.
  3. Clavier, C.; Coates, G.; Dumanoir, J. Theoretical and experimental bases for the dual-water model for interpretation of shaly sands. SPE J. 1984, 24, 153–168.
  4. Berg, C. R. Effective-medium resistivity models for calculating water saturation in shaly sands. *Log Analyst* 1996, 37.
  5. Simandoux, P. Dielectric Measurements on Porous Media; Application to the Measurement of Water-Saturations: Revue de l’Institut Francais du Petrole 18. Translated text in shaly sand reprint volume. Soc. Well Log Analysts, Houston 1963, p. IV, 97–124.
  6. Juhasz, I. Normalised Qty-the key to shaly sand evaluation using the Waxman-Smits equation in the absence of core data. In *SPWLA 22nd Annual Logging Symposium*, 1981, 23.
  7. Dayhoff, J. E. Computational properties of networks of synchronous groups of spiking neurons. *Neural Comput.* 2007, 19, 2433–2467.
  8. Wessels, L. F.; Barnard, E. Avoiding false local minima by proper initialization of connections. *IEEE Trans.* 1992, 3, 899–905.
  9. Nguyen, D.; Widrow, B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Int. Jt. Conf. Neural Networks* 1990, 3, 21–26.
  10. Abdel Azim, R. Application of artificial neural network in optimizing the drilling rate of penetration of western desert Egyptian wells. *SN Appl. Sci. J.* 2020, 2, 1177.
  11. Veezhinathan, J.; Wagner, D. A neural network approach to first break picking. *Int. Jt. Conf. Neural Networks* 1990, 3, 235–240.
  12. Al-Kaabi, A. U.; Lee, W. J. Using artificial neural networks to identify the well test interpretation model. *SPE Form. Eval. J.* 1993, 8, 233–240.
  13. Azim, R. A. Prediction of multiphase flow rate for artificially flowing wells using rigorous artificial neural network technique. *Flow Meas. Instrum.* 2020, 76, No. 101835.
  14. Garcia, G.; Whitman, W. W. Inversion of a lateral log using neural networks. In *Petroleum Computer Conference*. 1825, 1992; p 4.
  15. Rogers, S. J.; Fang, J. H.; Karr, C. L.; Stanley, D. A. Determination of lithology from well logs using a neural network. *AAPG Bull.* 1992, 5, 731–739.
(16) McCormack, M. D.; Day, R. How artificial intelligence impacts E&P productivity. *World Oil* 1993, 10, 81–86.
(17) Fogelman-Soulie, F. Neural networks and their applications for the oil industry. *Rev. Inst. Fr. Pet.* 1992, 3, 383–392.
(18) Adeniran, A. A. Artificial Intelligence Techniques in Reservoir Characterization. Doctoral Dissertation, King Fahd University of Petroleum and Minerals, 2009.
(19) Hamada, G.; Sakka, A.; Nyein, C. Prediction of Porosity and Water Saturation Using Neural Networks in Shaly Sand Reservoirs, Western Desert, Egypt. *J. Pet. Min. Eng.* 2020, 22, 80–91.
(20) Elshafei, M.; Hamada, G. M. Neural network identification of hydrocarbon potential of shaly sand reservoirs. In *SPE Saudi Arabia Section Technical Symposium*, 2007.
(21) Hamada, G. M.; Al-Gathe, A. A.; Al-Khudafi, A. M. Hybrid Artificial Intelligent Approach for Determination of Water Saturation using Archie’s Formula in Carbonate Reservoirs. *J. Pet. Environ. Biotechnol.* 2015, 6, 250.
(22) Aydin, G. The application of trend analysis for coal demand modeling. *Energy Sources, Part B* 2015, 10, 183–191.
(23) Al-Bulushi, N.; King, P. R.; Blunt, M. J.; Kraaijveld, M. Development of artificial neural network models for predicting water saturation and fluid distribution. *J. Pet. Sci. Eng.* 2009, 68, 197–208.
(24) Amiri, M.; Ghiasi-Freez, J.; Golkar, B.; Hatampour, A. Improving water saturation estimation in a tight shaly sandstone reservoir using artificial neural network optimized by imperialist competitive algorithm—A case study. *J. Pet. Sci. Eng.* 2015, 127, 347–358.
(25) Mardi, M.; Nurozi, H.; Edalatkhah, S. A water saturation prediction using artificial neural networks and an investigation on cementation factors and saturation exponent variations in an Iranian oil well. *Pet. Sci. Technol.* 2012, 30, 425–434.
(26) Kamalyar, K. Using artificial neural network for predicting water saturation in an Iranian oil reservoir. In *10th EAGE International Conference on Geoinformatics-Theoretical and Applied Aspects*, 2011, 4, p 24.
(27) Helle, H. B.; Bhatt, A. Fluid saturation from well logs using committee neural networks. *Pet. Geosci.* 2002, 8, 109–118.
(28) Yan, J. Reservoir parameters estimation from well log and core data: a case study from the North Sea. *Pet. Geosci.* 2002, 8, 63–69.
(29) Azim, R. A. Estimation of fracture network properties from FMI and conventional well logs data using artificial neural network. *Upstream Oil Gas Technol.* 2021, 7, No. 100044.
(30) Gorjaei, R. G.; Songolzadeh, R.; Torkaman, M.; Safari, M.; Zargar, G. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. *J. Nat. Sci. Eng.* 2015, 24, 228–237.
(31) Gomaa, S.; Soliman, A. A.; Mohamed, A.; Emara, R.; Attia, A. M. New Correlation for Calculating Water Saturation Based on Permeability, Porosity, and Resistivity Index in Carbonate Reservoirs. *ACS Omega* 2022, 7, 3549–3556.
(32) Moré, J. J. The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, 1976, vol 21; pp 105–116.
(33) Azim, R. A. Estimation of Bubble Point Pressure and Solution Gas Oil Ratio Using Artificial Neural Network. *Int. J. Thermofluid* 2022, 14, No. 100159.