Journal of Geomine (2025), 3(2), 93-105
DOI: 10.22077/jgm.2025.9704.1052

Use of well logging data to estimate fluid saturation based on artificial neural network algorithms

Adel Eidpoor[1], Mojtaba Rahimi[1,2], Ali Heidari[1]

Recived: 2025 Jul. 08, Revised: 2025 Aug. 11, Published: 2025 Aug. 18

ABSTRACT

Among all the methods used for determining fluid saturation of the reservoir rock, the ability of neural networks to predict fluid saturation in reservoir rock is of great interest to researchers. This study gathers the necessary data for estimating this important reservoir parameter and the variables involved in the process. Afterward, artificial neural networks (ANNs) and particle swarm optimization (PSO) algorithms are combined to provide a proper and accurate model for estimating water saturation. This combination provides an outstanding model in which fluid saturation distribution at any point in one of Iran's carbonate oil reservoirs can be obtained. To predict the water saturation value as the model output, several input parameters, including depth, gamma ray, resistance, neutron, micro-spherical resistance, and spontaneous potential logs, are employed. The multi-layer perceptron neural network (MLP) and radial basis function neural network (RBF) are the two models used, and the accuracy of each model is examined. Although the relationship between fluid saturation in the reservoir and logging information is completely nonlinear, these two artificial intelligence (AI) models can very well recognize this nonlinear relationship and provide great predictions with high correlation coefficient (R) values and low average absolute relative deviation (AARD) and root mean square error (RMSE) values. The values of R, AARD, and RMSE for the MLP model are obtained as 0.9739, 33.24, and 0.0824, respectively. Those for the RBF model are 0.9986, 7.47, and 0.0024, respectively, reflecting that the RBF model is superior to the MLP model due to its higher R value and lower AARD and RMSE values.

KEYWORDS

Carbonate oil reservoirs, Fluid saturation, Artificial intelligence, Artificial neural network, Multi-layer perceptron neural network, Radial basis function neural network

I. INTRODUCTION

One of the most important issues facing Iran's oil industry today is the continuous production and exploration of oil and gas at minimal cost. According to the latest information, over 60% of the world's oil reserves and 40% of gas reserves are located in carbonate reservoirs. The majority of reservoirs in the Middle East are also predominantly carbonate (Riazi et al., 2007). Accurate estimation of fluid saturation within rock pores is essential for calculating in-place reserves in a reservoir. Changes in fluid saturation during production from the reservoir indicate which sections are more efficient for production. These observations assist petroleum engineers in predicting production changes and optimizing the reservoir's production and depletion strategies. Any miscalculation in determining saturation levels can lead to the loss of productive oil zones. Since determining fluid saturation at the surface through core tests is challenging, time-consuming, and often lacks sufficient accuracy, the majority of experts in the oil sector prefer to use well logging data. On the other hand, carbonate rocks, due to their complex structure and heterogeneous properties, often pose challenges in interpreting logging data. Consequently, conventional modeling with fixed computational procedures may not yield acceptable answers in many cases. Given the importance of carbonate reservoirs, as most of the world's reservoirs are carbonate, methods that can accurately predict this petrophysical parameter have attracted considerable attention from researchers (Shedid and Saad, 2017).

New computational tools have provided solutions to these problems. Artificial neural networks (ANNs) are one of these new, efficient, and powerful computational tools that allow for the high-accuracy correlation of a wide range of data without getting involved in detailed computational models. Therefore, by utilizing intelligent systems, the development plan of an oil field can be significantly improved by generating the necessary data and reducing costs. An ANN, with its unmatched mathematical capabilities, can relate data obtained from well logging to fluid saturation. A notable feature of this

[1] Department of Petroleum Engineering, Kho. C., Islamic Azad University, Khomeinishahr, Iran, [2] Stone Research Center, Kho. C., Islamic Azad University, Khomeinishahr, Iran
✉ M. Rahimi: mrahimi@iau.ac.ir

model is its high flexibility, ease of validation, and simple interpolation and extrapolation of data. In this study, this tool and its capabilities in the field of extracting fluid saturation data are analyzed, and suitable computational networks are developed to estimate fluid saturation in the reservoir using well logging data and ANNs. Previous studies have been conducted on the application of intelligent methods to obtain reservoir rock properties through well logging data, including the following: Al-Bulushi et al. (2009) used ANNs to predict water saturation values from well logging data. The model was examined on one of the oil fields in Oman, showing a strong ability to estimate water saturation with a correlation coefficient of 0.91. Kaydani et al. (2014) studied permeability estimation in one of the oil fields in Iran based on a multi-gene genetic algorithm. The results obtained from this model were compared with values achieved from an adaptive neuro-fuzzy inference system and genetic algorithm, demonstrating that the model could serve as a quick and suitable method for estimating permeability. Azizi et al. (2017) utilized ANNs and support vector machines (SVMs) to estimate permeability in one of the heterogeneous carbonate reservoirs. Initially, well logging data were categorized into electrical facies using principal component analysis and model-based cluster analysis. Each electrical facies was then considered as input for the ANN and SVM. The ANN was trained with ten hidden layers using the backpropagation and momentum gradient descent algorithms. This study indicated that the radial basis function (RBF) of the SVM had lower error rates compared to the ANN. Soleimani et al. (2020) analyzed the porosity of an oil field in Iran using ANNs, with their study indicating that estimating porosity using ANNs is preferable compared to other methods. Okon (2021) employed a developed feedforward backpropagation neural network model with multiple inputs and outputs to predict the petrophysical properties of the reservoir, including porosity, permeability, and water saturation in the Niger Delta region. The results reflected that this model could serve as a suitable method for estimating petrophysical parameters, achieving a correlation coefficient of 0.99. Gamal and Elkatatny (2022) investigated a predictive model for rock porosity based on ANNs, noting that rock porosity significantly impacts hydrocarbon reserve estimation and is considered an important petrophysical feature. Most porosity measurements are conducted in laboratories, which are costly and time-consuming. This study utilized a new model based on ANNs, with drilling data as input parameters and rock porosity as the output parameter. The results demonstrated that the model possesses high accuracy and capability. García-Benítez and Arana-Hernández (2023) employed ANNs to estimate permeability using well logging and core data in heterogeneous environments. The input data included depth, porosity, water and oil saturations, grain density, resistivity, and gamma-ray data, while the network output was rock permeability. The results were compared with three semi-empirical models used to describe the reservoir, concluding that the ANN provided the best overall prediction and the highest correlation coefficients compared to the three conventional methods in heterogeneous formations.

Although particle swarm optimization (PSO) has been applied in various optimization problems within petroleum engineering, its application in optimizing RBF network parameters for fluid saturation prediction is limited in the literature. Previous studies on fluid saturation estimation have mainly used neural networks with fixed parameters, genetic algorithms, or other heuristic approaches, without employing PSO to fine-tune both the spread value and the number of hidden neurons. The integration of PSO with RBF in this study provides a robust framework that combines the nonlinear approximation capability of RBF with the global search efficiency of PSO, leading to automated parameter selection and improved predictive accuracy. This represents a distinctive contribution of the present work compared to prior studies. Considering that the use of ANNs for feature extraction has recently emerged in research, and that there is no longer operator intervention in feature determination, this has substantially enhanced identification and classification accuracy. The novelty in this study aims to extract features using the integration of multiple neural networks to estimate fluid saturation. This field is new, and given the noteworthy accuracy of well logging data for oil carbonate reservoirs in recent years and the lack of a concentrated resource on this topic, it can be said that this research has substantial potential for further exploration. We intend to employ different models and algorithms, and as a result, a comparison between them will be conducted, validating the results of the model's predictions against the measured data.

II. METHODS

A. Multi-layer Perceptron Neural Network (*MLP*)

MLPs have a good capability for producing a specific output corresponding to an input vector. These networks are particularly suitable for solving simple classification problems. MLP is the most well-known model of ANN. A multilayer neural network consists of three parts: the input layer, the hidden layer, and the output layer (Fausett, 2006). The perceptron is the simplest type of neuron modeling. Since examining multiple perceptrons in different layers can be somewhat complex, we will focus on a single perceptron. A perceptron has a series of external inputs, an internal input known as bias, and an output. Each perceptron represents a neuron, and the output of a perceptron is always Boolean. This means the output can take on two

values: 1 or 0. If the output of a perceptron is 1, it is referred to as an active perceptron. All inputs, including the bias, have an associated weight, which is multiplied by the input value. Typically, the weight of the bias is set to 1. One of the most important factors for any neuron is its activation function. The activation function determines how the neuron's output will behave based on its inputs. In perceptrons, we use one of the simplest activation functions. This function sums all the perceptron's inputs after multiplying them by their respective weights. If the sum is greater than or equal to a certain threshold, the output will be 1, meaning the perceptron will be activated; otherwise, the perceptron will be inactive (Barlow, 1989). One of the main characteristics of perceptrons is their ability to learn. This learning in perceptrons is supervised, meaning we need to have a set of inputs along with the correct outputs for the perceptron to be able to mimic them.

B. Radial Basis Function Neural Network (*RBF*)

RBFs are widely used for non-parametric estimation of multidimensional functions through a limited set of training data. RBFs are particularly interesting and efficient due to their rapid and comprehensive training, which has garnered special attention. It has been proven that RBFs are extremely powerful approximators, capable of approximating any continuous function to any degree of accuracy, provided they have a sufficient number of neurons in the hidden layer (Dayhoff, 1990; Huang and Zhang, 1994). Notably, these networks possess this property with just a single hidden layer. RBFs are primarily inspired by statistical techniques for pattern classification and have essentially emerged as a vital type of neural network, with their significant advantage being the classification of patterns that exist in non-linear spaces. Although the number of these techniques is limited, they have been widely utilized. These networks are often compared to error backpropagation neural networks. The basic architecture of RBF consists of a three-layer network, as illustrated in Fig. 1 (Du & Swamy, 2006).

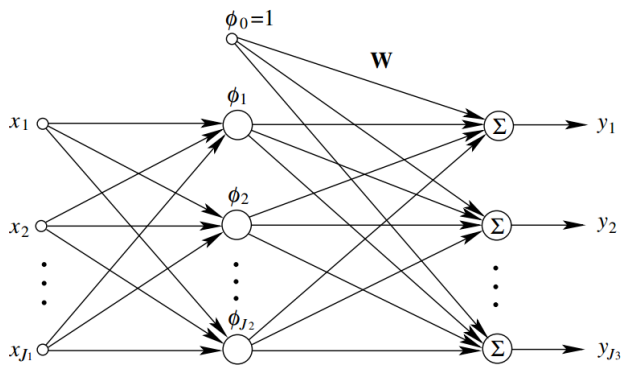

Fig. 1. Schematic representation of an RBF (Du & Swamy, 2006)

The input layer consists of only a single input layer, where no processing occurs. The second layer, or hidden layer, establishes a nonlinear mapping between the input space and a typically higher-dimensional space, playing a crucial role in transforming nonlinear patterns into linearly separable patterns. Finally, the third layer produces a weighted sum along with a linear output. When RBF is employed for function approximation, such an output is useful; however, if classification of patterns is required, a hard limiter or a sigmoid function can be applied to the output neurons to generate output values of 0 or 1. Each neuron in layer j of the hidden layer implements a specific algorithm that consists of two steps: in the first step, the difference of each input P_i (i=1,2,...,R) with respect to its associated weight (W_{ij}) is calculated, and this value is subjected to the Euclidean distance function or norm. Ultimately, the value of n is obtained by multiplying the resulting value by the threshold value.

$$b_i n_j = \|p_i - w_{ij}\| \tag{1}$$

Thus, one of the differences between RBFs and MLPs lies in the definition of the value n. In RBFs, n is defined as the difference of the Euclidean distance between the input vector and the weight vector, whereas in MLPs, it is the dot product of these two vectors. In the second step, the activation function f is applied to n to obtain the neuron's output. The most commonly used activation functions in RBFs for the hidden layer are exponential Gaussian functions, while the identity function is used for the output layer. The exponential Gaussian function is defined as follows (Devijver and Kittler, 1982):

$$F(n_j) = \exp(-n_j)^2 \tag{2}$$

The positive features of RBFs include fast learning, the ability to train with minimal initial datasets, the determination of the optimal network size by the algorithm itself, and the absence of local minimum problems.

C. Particle Swarm Optimization (*PSO*)

The PSO algorithm is a type of optimization algorithm that operates based on the random generation of an initial population. This algorithm is inspired by and simulates the collective flight of birds or the collective movement of fish. Each member of this group is defined by a velocity vector and a position vector in the search space. In each iteration, the new position of the particles is defined based on the velocity vector and the position vector in the search space. During each iteration, the new position of the particles is updated according to the current velocity vector, the best position found by that particle, and the best position found by the best particle in the group. This algorithm was initially defined for continuous parameters; however, since some applications involve discrete parameters, the algorithm

has also been extended to discrete cases. The discrete version of the PSO algorithm is referred to as Binary PSO (BPSO), where the position of each particle is defined by binary values of one and zero. In this algorithm, the position of each particle is represented by binary values of either zero or one. In BPSO, the value of each particle can change from zero to one or from one to zero.

Since PSO also begins with a randomly generated population matrix, it is similar to many other evolutionary algorithms, such as genetic algorithms. Unlike genetic algorithms, PSO does not have evolutionary operators like mutation and crossover. Each population element is called a particle, which is equivalent to a chromosome in genetic algorithms. In fact, the PSO algorithm consists of a specified number of particles that randomly receive initial values. For each particle, two values, namely position and velocity, are defined and modeled respectively by a position vector and a velocity vector. These particles iteratively move in the n-dimensional space of the problem to search for new possible options by calculating the optimal value as a measure of fitness. The dimensionality of the problem space corresponds to the number of parameters present in the function to be optimized. A memory is allocated to store the best position of each particle in the past, and another memory is allocated to store the best position encountered among all particles. Using the experiences from these memories, particles decide how to move in the next iteration. In each iteration, all particles move in the *n*-dimensional space of the problem until the global optimal point is found. The particles update their velocities and positions based on the best global and local solutions.

The PSO algorithm was configured in MATLAB with carefully selected parameters to ensure a balance between global search capability and convergence speed. Table 1 summarizes the complete parameter settings used in this study for optimizing the spread and the number of neurons in the RBF network. These settings allowed the algorithm to efficiently explore the search space while avoiding premature convergence.

III. Target Oil Field

The Bibi Hakimeh oil field is located in the Dezful embayment, Southwest of Iran. This oil field is situated in the city of Gachsaran, within the coordinates of 50° 12' to 50° 53' East longitude and 30° 16' to 29° 54' North latitude. The field was discovered with the drilling of well BH-1 in 1961, and production commenced in the same year. The Bibi Hakimeh oil field is composed of an asymmetric subsurface anticline, 70 km in length and 7 km in width, located approximately 250 km southeast of Ahvaz. The reservoirs of this field include Asmari, Bangestan, and Khami. The Pabdeh and Gurpi layers, approximately 400 m thick, separate the producing reservoirs of Asmari and Bangestan. The Asmari reservoir horizon in this field has a length of 75 km and a width of 5 km. Based on the petrophysical and lithological characteristics of this reservoir horizon, it is divided into four distinct zones. The Bangestan horizon present in this oil field has a length of 72 km and a width of 4.5 km. This reservoir horizon is also divided into nine distinct zones based on its petrophysical and lithological characteristics. The Khami oil system of the Bibi Hakimeh oil field consists of the Fahliyan, Gadvan, and Darian formations. The Bakhtiari, Aghajari, and Mishan formations are the surface outcrops of this field (Faghih et al., 2022). Fig. 3 shows the location of the Bibi Hakimeh oil field in the Dezful embayment.

IV. Data Gathering

To predict the saturation of fluids using AI algorithms, 350 log data points from one of Iran's carbonate oil reservoirs, located in the Bibi Hakimeh oil field, were utilized. The input parameters for various models include depth, gamma ray log (GR), resistivity log (RT), spontaneous potential log (SP), neutron log (CNL), and micro-spherical focused log (MSFL), with water saturation considered as the output parameter of the models. It is worth noting that by predicting water saturation, oil saturation can also be easily calculated. Some static parameters of the input and output data, such as minimum, maximum, and average values for each parameter, are presented in Table 2. For model construction, 85% of the data was used for training each model, while 15% of the data was used for testing the models.

Fig. 4 illustrates the Pearson correlation coefficients among the input well log parameters and water saturation. As shown, the resistivity log (RT) exhibits the strongest negative correlation with water saturation (−0.63), indicating that higher resistivity values are generally associated with lower water saturation, which is consistent with petrophysical principles (i.e., the lower saline water saturation, the higher formation resistivity). In contrast, the neutron log (CNL) shows the highest positive correlation with water saturation (0.42), followed by spontaneous potential (SP) with a moderate positive correlation (0.36). Additionally, moderate to strong correlations exist between specific input parameters, such as the negative correlation between RT and CNL (−0.62) and the positive correlation between MSFL and CNL (0.44). These relationships highlight the complementary nature of the selected input logs while also suggesting potential redundancy among specific measurements.

Table 1. PSO parameter settings used for RBF network training

Parameter	Value / Setting	Description
Swarm size	500	Number of particles in the swarm
Maximum iterations	1000	Maximum allowed iterations before termination
Maximum stall iterations	1000	Iterations allowed without improvement in fitness
Initial swarm range	−8.5 to +8.5	Range for randomly generating initial particle positions
Cognitive learning coefficient (c1)	1.49 (MATLAB default)	Weight for the individual particle's best-known position
Social learning coefficient (c2)	1.49 (MATLAB default)	Weight for the global best-known position in the swarm
Termination criteria	Max iterations or minimal improvement in fitness	Stopping condition

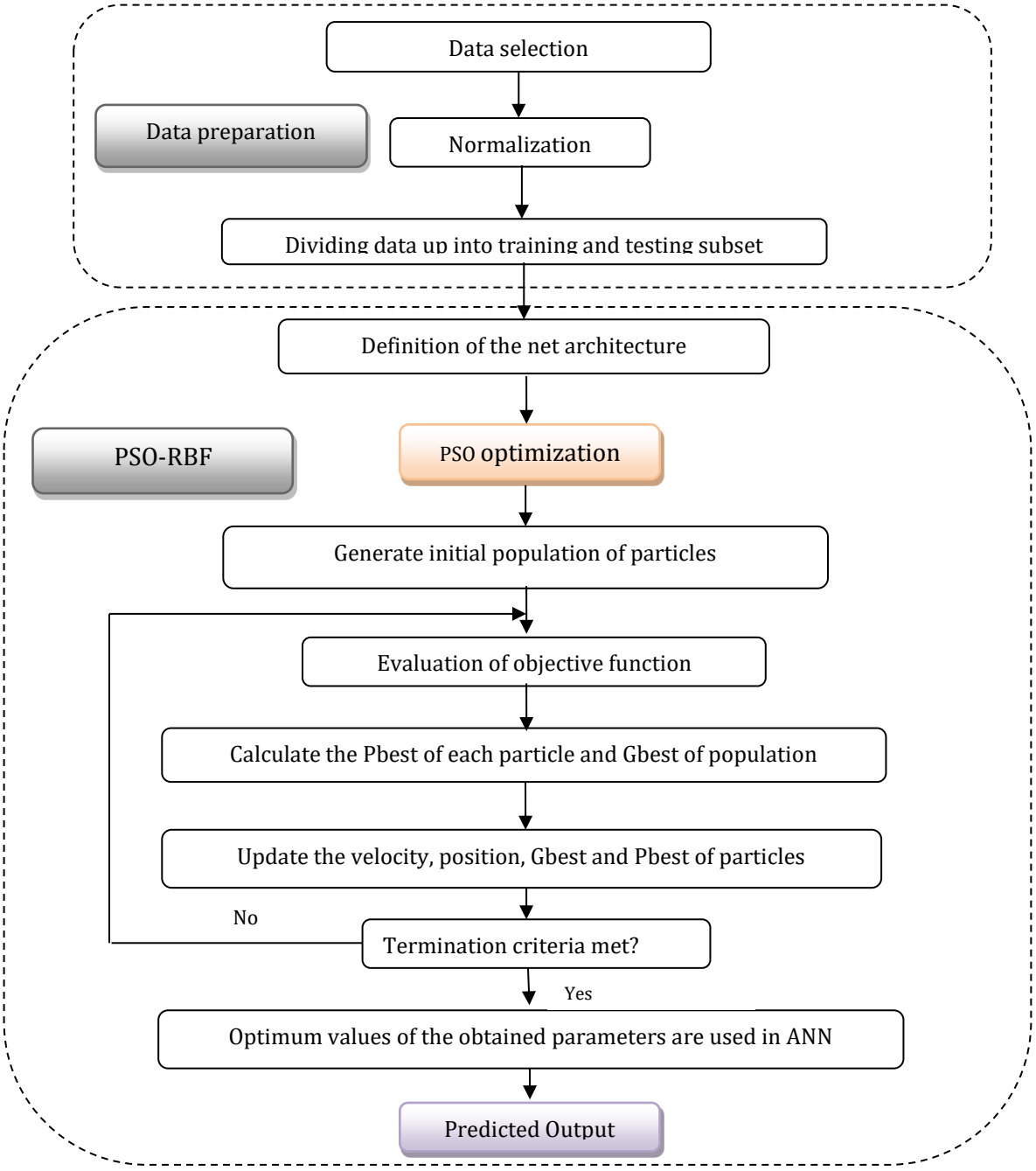

Fig. 2. presents the flowchart of the suggested RBF-PSO algorithm.

Fig. 3. Location of the Bibi Hakimeh oil field in the Dezful embayment (Abbasi and Tabatabaei, 2019)

Table 2. Static parameters of the input and output data used in this study

Parameter	Type	Minimum	Maximum	Average
Depth (m)	Input	1150	1200	1175
Resistivity Log (RT) (Ωm)	Input	10.51	563.84	73.85
MSFL Log (Ωm)	Input	1.96	104.27	6.23
Neutron Log (CNL) (%)	Input	0.81	15	5.41
SP Log (mv)	Input	5.3	90.7	56.6
GR Log (API)	Input	14.14	130.01	42.2
Water Saturation (fraction)	Output	0	0.3	0.2

V. MODEL DEVELOPMENT

A. MLP Model

This model consists of three layers, with the middle layer, or hidden layer, potentially having one or more sub-layers. However, based on previous studies and research, it has been proven that selecting a single sub-layer for the middle layer is sufficient for predicting data with acceptable accuracy. In this research, a single middle layer is also used for data prediction. The hyperbolic tangent function is employed in the middle layer, while a linear function is used in the output layer. The number of neurons in the middle layer is a crucial parameter affecting the accuracy of the model. In this study, the number of neurons in the middle layer varies from 4 to 16, and the accuracy of the networks is evaluated in each case. The accuracy of the networks is calculated by computing the mean squared error (MSE) between the actual data and the data predicted by the model for both the training and testing sections. Fig. 5 illustrates MSE for the training and testing data as a function of the number of neurons in the middle layer. As shown, the lowest MSE occurs with 14 neurons, indicating that the network with this number of neurons in the middle layer has the highest accuracy. Therefore, this model is selected for data prediction and comparison with other models.

B. RBF Model

The number of layers in this type of network, similar to multilayer neural networks, consists of an input layer, a middle or hidden layer, and an output layer. In these networks, the middle layer cannot have sub-layers and consists of only a single fixed layer. Two parameters, namely the spread number and the number of neurons in the middle layer, significantly affect the accuracy of the model. Therefore, to present a model with acceptable accuracy, it is essential to optimize the numerical values of these two parameters. The PSO algorithm is utilized to optimize these two parameters. Initially, a population of 100 members is selected for the values of these two parameters. Based on these parameters, 100 models are constructed, and the accuracy of each model is calculated by determining its MSE. Then, by using the operators related to the PSO algorithm, a new population of 100 members is created based on the best solutions from the previous population. This process continues with the PSO algorithm until optimal solutions are found. Fig. 6 shows the minimum MSE against the iterations of the PSO algorithm.

As indicated, after 142 iterations, the MSE stabilizes and no longer changes, indicating that the algorithm has converged to the minimum parameter values. The optimal values for the spread parameter and the maximum number of neurons in the middle layer are determined to be 0.38 and 24, respectively. Using these two parameters, a radial basis model is constructed for data prediction, and its accuracy is compared with that of other models.

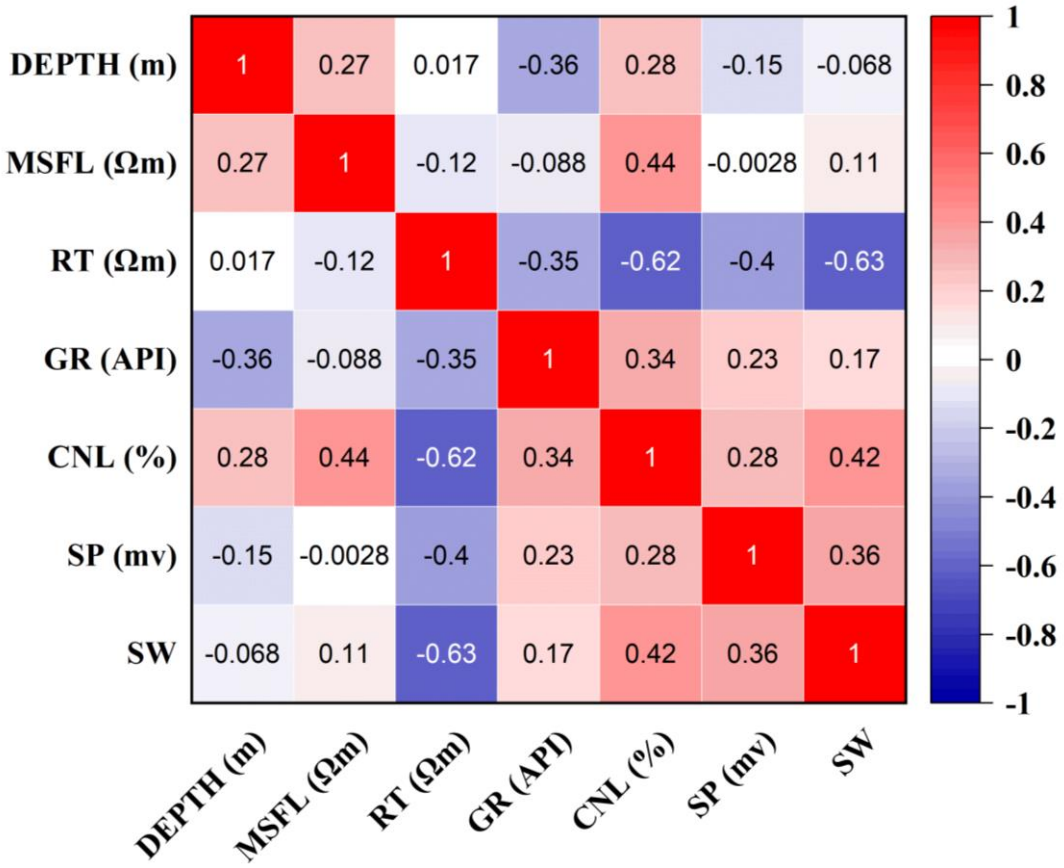

Fig. 4. Pearson correlation heatmap between input well log parameters and water saturation (SW), illustrating the strength and direction of relationships (−1 to +1) among variables.

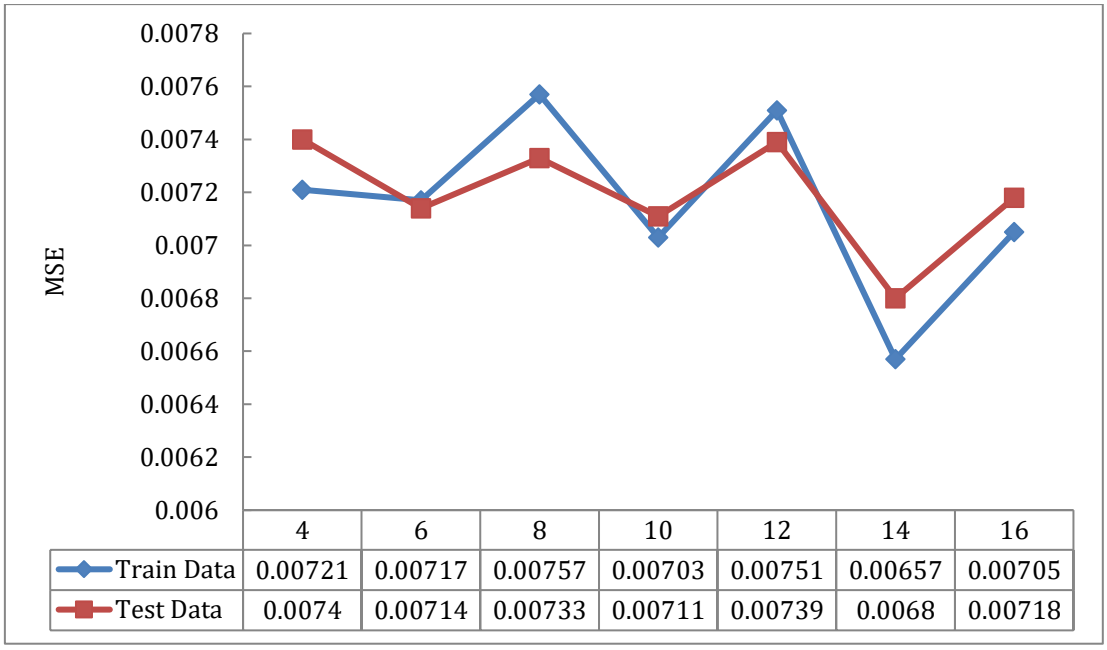

Fig. 5. Change in MSE as a function of the number of neurons in the middle layer for testing and training Data

C. Evaluation Criteria for Model Performance

These statistical parameters include the average absolute relative deviation (AARD) and the root mean squared error (RMSE). The following formulas are used to calculate these parameters:

$$\%AARD = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{(\lambda_{pred}(i)-\lambda_{Exp}(i))}{\lambda_{Exp}(i)}\right| \quad (3)$$

$$RMSE = \left(\frac{\sum_{i=1}^{N}(\lambda_{pred}(i)-\lambda_{Exp}(i))^2}{N}\right)^{0.5} \quad (4)$$

Where $\lambda_{pred}(i)$ denotes the predicted value by the model and $\lambda_{Exp}(i)$ stands for the actual value of the parameter, and N represents the number of data points.

VI. RESULTS

A. MLP

After determining the number of neurons in the input layer, the training data are used to construct this model. Following the model creation, its performance and accuracy are further evaluated using the testing data. The cross-plots in Figs. 7 and 8 depict the correlation coefficient (parameter R) between the saturation obtained from the logs and the saturation predicted by the MLP for both the training and testing data. The horizontal axis represents water saturation data obtained from the logs, while the vertical axis represents the data predicted by the model. The closer the R is to one, the higher the accuracy of the model. Fig. 9 shows the relative error for the training and testing data related to the designed MLP network, indicating the number of data points and the vertical axis representing the calculated relative error. The closer the data points are to the horizontal line corresponding to a relative error of zero, the more accurate the model is.

In Figs. 10 and 11, a comparison is made between the saturation values predicted by the MLP and those obtained from the well logs. The horizontal axis represents the number of water saturation data points. In contrast, the vertical axis shows the water saturation data obtained from the well logs and the data predicted by the model. The greater the overlap between the circular data points and the dashed lines, the closer the log-derived data and the model's predicted data are to each other, indicating that the model has higher accuracy. The training data are shown in Fig. 10, and the testing data are exhibited in Fig. 11.

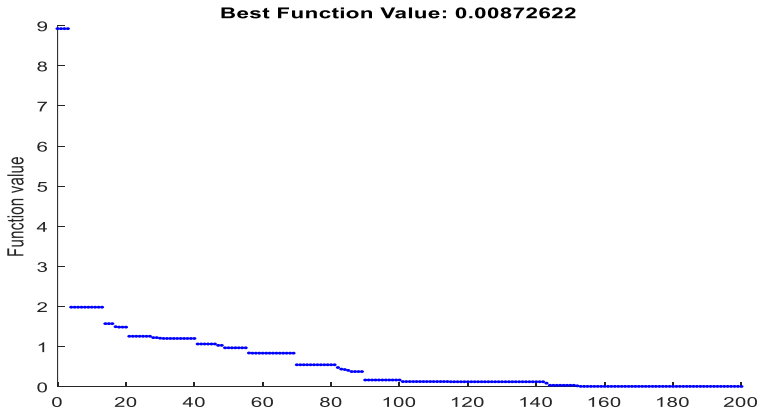

Fig. 6. Performance of the PSO algorithm for optimizing parameters of the radial basis network

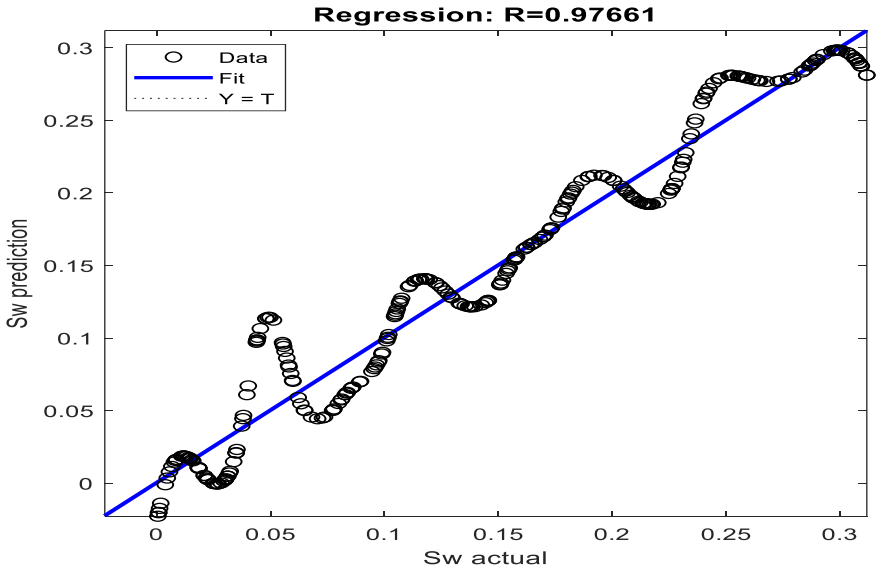

Fig. 7. Correlation coefficient between water saturation values obtained through well logs and those using MLP for training data

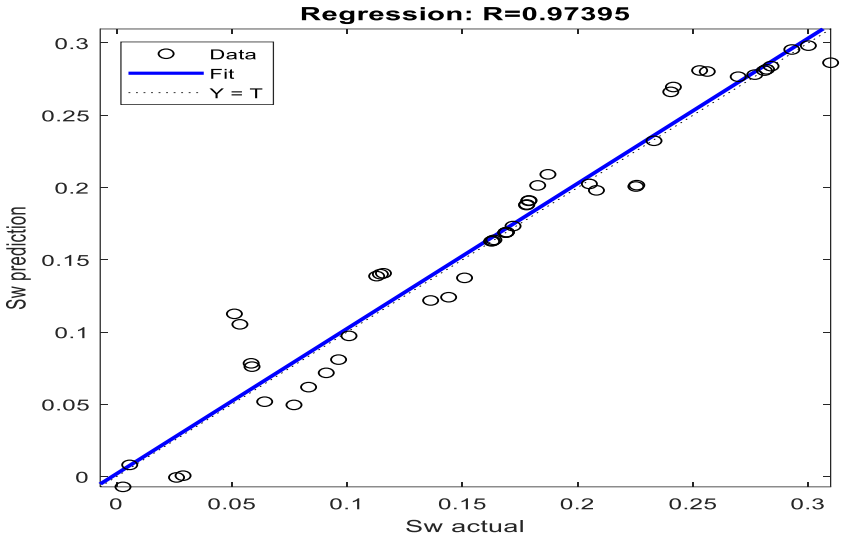

Fig. 8. Correlation coefficient between water saturation values obtained through well logs and those using MLP for testing data

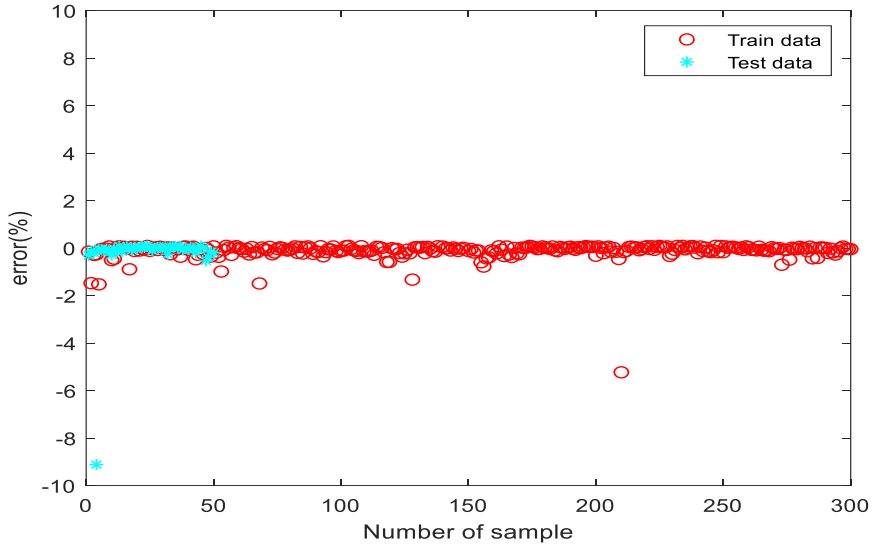

Fig. 9. Relative error for training and testing data of MLP

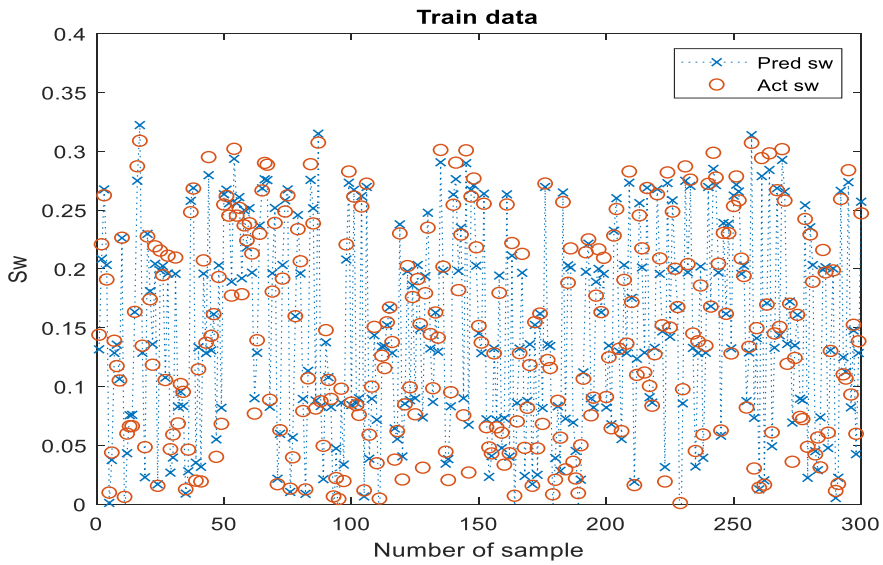

Fig. 10. Comparison of predicted values of water saturation by MLP with water saturation data obtained through well logs for training data

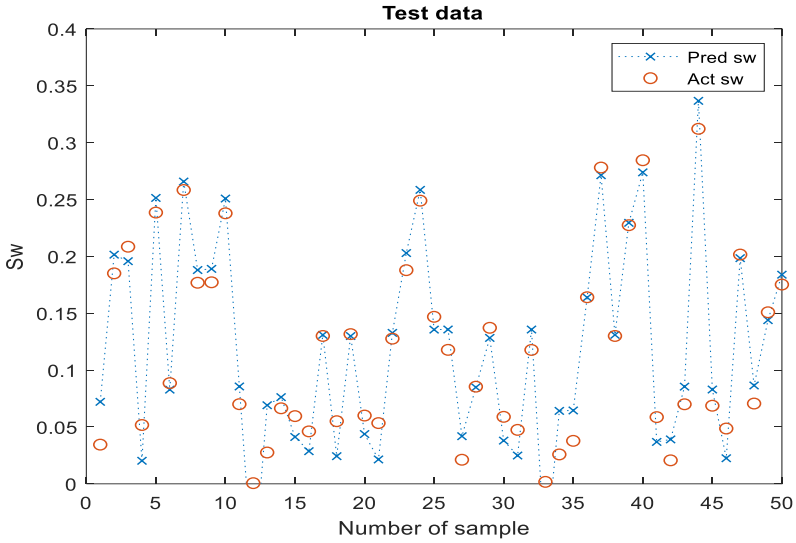

Fig. 11. Comparison of predicted values of water saturation by MLP with water saturation data obtained through well logs for testing data

B. RBF with PSO Algorithm

As previously mentioned, this model has two parameters: the maximum number of neurons and the spread number, for which optimal values of 24 and 0.38 are obtained by using the PSO algorithm. Using these two values, the model is trained and constructed based on the training data, and the predictions made by this model are evaluated using the testing data, resulting in the following outcomes. The cross-plots in Figs. 12-13 exhibit R between the water saturation obtained from the well logs and that predicted by the RBF for both the training and testing data. The horizontal axis represents the water saturation data obtained from the logs, while the vertical axis shows the data predicted by the model. The closer the R is to one, the higher the accuracy of the model. Fig. 14 shows the relative error for the training and testing data related to the designed RBF network. The horizontal axis represents the number of data points, while the vertical axis indicates the calculated relative error. The closer the data points are to the horizontal line corresponding to a relative error of zero, the more accurate the model is.

In Figs. 15 and 16, a comparison is made between the saturation values predicted by the RBF network and the saturation data obtained from the well logs. The horizontal axis represents the number of water saturation data points. In contrast, the vertical axis exhibits the water saturation data obtained from the well logs and those predicted by the model. The greater the overlap between the circular data points and the dashed lines, the closer the log-derived data and the model's predicted data are to each other, reflecting that the model has higher accuracy.

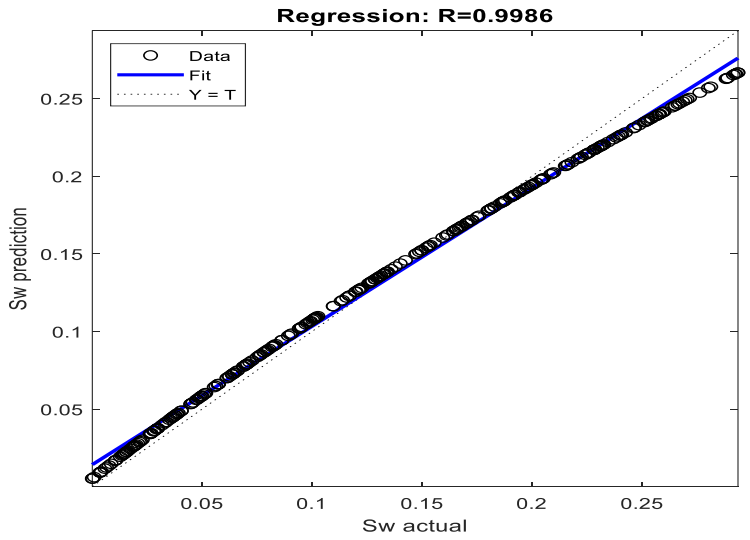

Fig. 12. Correlation coefficient between water saturation values obtained through well logs and those using RBF for training data

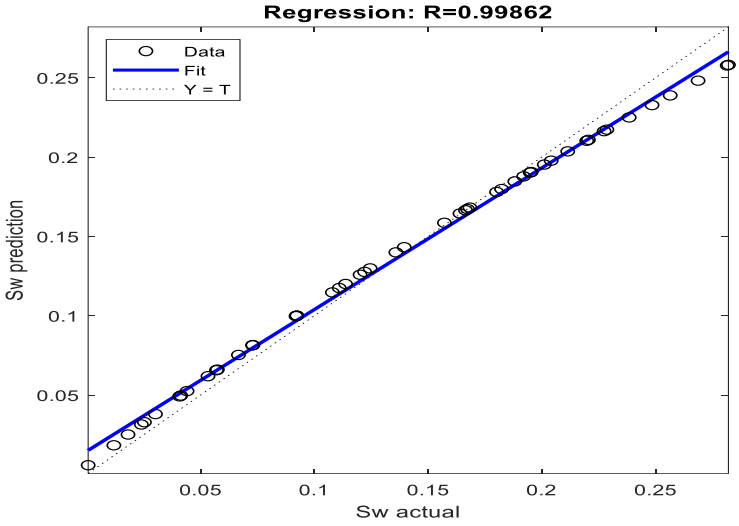

Fig. 13. Correlation coefficient between water saturation values obtained through well logs and those using RBF for testing data

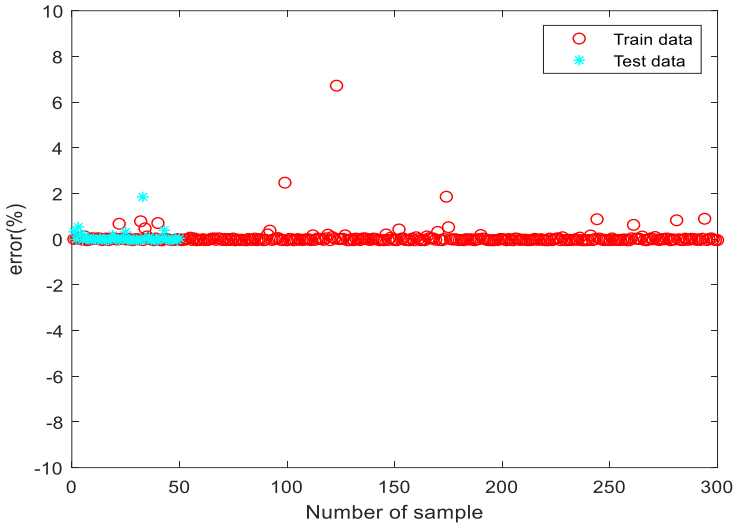

Fig. 14. Relative error for training and testing data of RBF

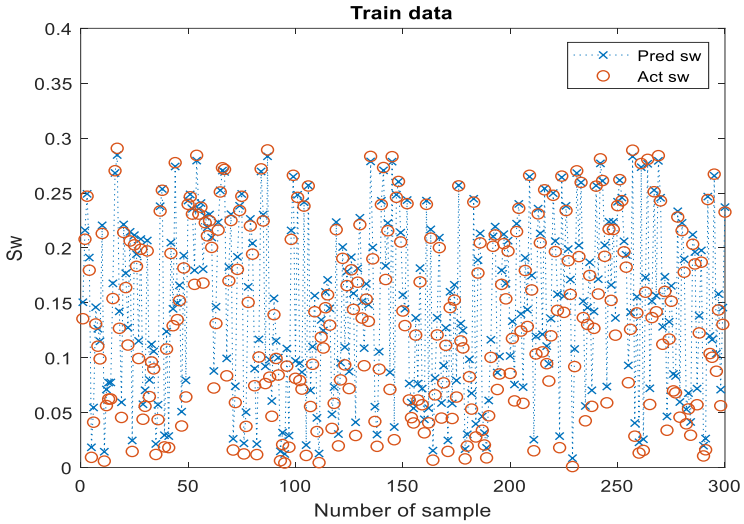

Fig. 15. Comparison of predicted values of water saturation by RBF with water saturation data obtained through well logs for training data

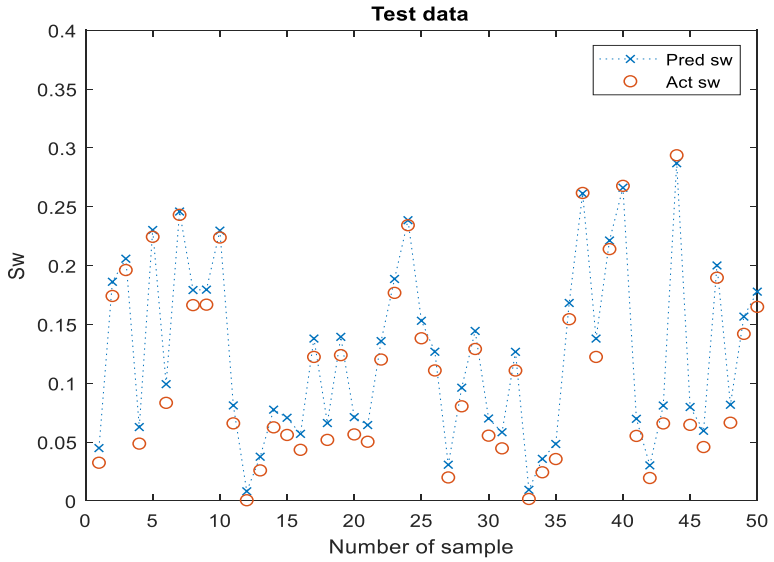

Fig. 16. Comparison of predicted values of water saturation by RBF with water saturation data obtained through well logs for testing data

The water saturation obtained from well logging data was successfully modeled using intelligent networks, yielding favorable results. The results from the intelligent methods employed in this research are presented in Table 3. As indicated, the RBF model has higher R values and lower AARD and RMSE values compared to the MLP model. This demonstrates that the RBF model is superior and has higher accuracy than the MLP model.

Table 3. Statistical parameters of different estimators of water saturation

Network Type	Data Type	*R* (fraction)	AARD (%)	RMSE (fraction)
MLP	Training Data	0.9766	30.21	0.0810
	Testing Data	0.9739	33.24	0.0824
RBF	Training Data	0.9986	8.12	0.0028
	Testing Data	0.9986	7.47	0.0024

VII. CONCLUSIONS

The following key results are obtained from this research:

1) AI models are capable of effectively identifying the relationships and patterns between input and output parameters that have completely nonlinear relationships, providing accurate predictions of output values. As demonstrated in the present study, although the relationship between the water saturation and well logging data is entirely nonlinear, the AI models MLP and RBF can effectively recognize this nonlinear relationship and yield predictions with high R values and low AARD and RMSE values.

2) The values of the statistical parameters for the R parameter, AARD, and RMSE for the MLP model are obtained as 0.9739, 33.24, and 0.0824, respectively. The values of these parameters for the RBF model are obtained as 0.9986, 7.47, and 0.0024, respectively.

3) Considering the statistical parameters, although both models demonstrate acceptable accuracy in predicting the water saturation data using well logging data, it can be concluded that the RBF model is superior to the MLP model due to its higher R value and lower AARD and RMSE values.

It is worth mentioning that the present study was conducted using well log data from a single oil carbonate reservoir. While the proposed ANN-based models demonstrate high predictive accuracy in this context, further validation using data from additional wells and reservoirs with varying lithological and petrophysical characteristics is recommended to fully assess the models' generalizability.

DATA AVAILABILITY

The field data used in this study are confidential and cannot be shared openly.

NOMENCLATURE

ANN	Artificial neural network
PSO	Particle swarm optimization
MLP	Multi-layer perceptron neural network
RBF	Radial basis function neural network
AI	Artificial intelligence
R	Correlation coefficient
AARD	Average absolute relative deviation
RMSE	Root mean square error
MSE	Mean squared error
SVM	Support vector machine
BPSO	Binary particle swarm optimization
GR	Gamma ray log
RT	Resistivity log
SP	Spontaneous potential log
CNL	Compensated neutron log
MSFL	Micro-spherically focused log

References

Abbasi, J., & Tabatabaei, H. (2019). Natural fracture analysis of reservoir rock via FMI image logs in Bibi-Hakimeh oil field, SW of Iran. Petroleum & Coal, 61(4).

Al-Bulushi, N., King, P. R., Blunt, M. J., & Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and engineering, 68(3-4), 197-208.

Azizi , Y., & Shad Manamanan, N. (2017). Permeability Prediction in one of the Iranian Carbonate Oil Reservoir using Artificial Neural Network and Support Vector Machine. Journal of Earth and space physics, 43(2), 259-281.

Barlow, H. B. (1989). Unsupervised learning. Neural computation, 1(3), 295-311.

Dayhoff, J. E. (1990). Neural network architectures: an introduction. Van Nostrand Reinhold Co..

Devijver, P.A., & Kittler. J. (1982). Pattern recognition: A statistical approach. Englewood Cliffs: Prentice/Hall International.

Du, K. L., & Swamy, M. N. (2006). Neural networks in a softcomputing framework (Vol. 501). London: Springer.

Faghih, A., Mehraban, E., Seraj, M., Zarei. S., Hosseini, S.M., & Soleimani. M. (2022). Detection of regional fractures effect in the Bibihakimeh oilfield based on the relative frequency of fractures method, Southern Dezful Embayment, Iran. Journal of Techtonics, 6(21), 1-13.

Fausett, L. V. (2006). Fundamentals of neural networks: architectures, algorithms and applications. Pearson Education India.

Gamal, H., & Elkatatny, S. (2022). Prediction model based on an artificial neural network for rock porosity. Arabian Journal for Science and Engineering, 47(9), 11211-11221.

García-Benítez, S. R., & Arana-Hernández, O. A. (2023). Neural model to estimate permeability from well logs and core data. Boletín de Geología, 45(1), 141-153.

Huang, S. H., & Zhang, H. C. (1994). Artificial neural networks in manufacturing: concepts, applications, and perspectives. IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A, 17(2), 212-228.

Kaydani, H., Mohebbi, A., & Eftekhari, M. (2014). Permeability estimation in heterogeneous oil reservoirs by multi-gene genetic programming algorithm. Journal of Petroleum Science and Engineering, 123, 201-206.

Okon, A. N., Adewole, S. E., & Uguma, E. M. (2021). Artificial neural network model for reservoir petrophysical properties: porosity, permeability and water saturation prediction. Modeling Earth Systems and Environment, 7(4), 2373-2390.

Riazi, M. R., Al-Haddad, A. A., & Mansoori, G. A. (2007). An update on the developments in petroleum production research in the Middle East. Journal of Petroleum Science and Engineering, 55(1), 1-5.

Shedid, S. A., & Saad, M. A. (2017). Comparison and sensitivity analysis of water saturation models in shaly sandstone reservoirs using well logging data. Journal of Petroleum Science and Engineering, 156, 536-545.

Soleimani, F., Hosseini, E., & Hajivand, F. (2020). Estimation of reservoir porosity using analysis of seismic attributes in an Iranian oil field. Journal of Petroleum Exploration and Production Technology, 10(4), 1289-1316.