Journal of Geomine (2025), 3(2), 126-139
DOI: 10.22077/jgm.2025.9727.1054

Artificial intelligence methods for estimation of formation porosity from well logging data

Reza Rooki[1], Mojtaba Rahimi[2,3]

Recived: 2025 Jul. 11, Revised: 2025 Sep. 02, Accepted: 2025 Oct. 05, Published: 2025 Oct. 13

ABSTRACT

Porosity is an essential rock property reflecting the capacity of the reservoir to store hydrocarbons, making it a critical parameter in the exploration and development of oil and gas resources. Well logging is a fundamental technique in the oil and gas industry for estimating the porosity of subsurface formations. Well logging methods do not directly measure porosity, but instead record physical parameters (e.g., bulk density, acoustic travel time, hydrogen index) that can be empirically or theoretically related to the porosity of the rock. The artificial intelligence (AI) methods of Backpropagation Neural Network (BPNN), General Regression Neural Network (GRNN), and Support Vector Machine (SVM) were adopted to estimate porosity from well-logging data points using MATLAB software. For this purpose, 70% of the data was used for training, and 30% of the data was used for testing these AI methods. The well logging dataset obtained from one of the Kaggle datasets belongs to the Peninsular Malaysia hydrocarbon field. The comparison of the porosity values predicted by the AI methods and the actual values indicated that the three AI methods predict porosity values with great accuracy (with *R* values all equal or very close to one). However, the BPNN method has a smaller error in estimating porosity compared to the other two AI methods, suggesting that BPNN outperforms the other two AI techniques in this study.

KEYWORDS

Formation porosity, well logging, artificial intelligence, machine learning, oil and gas resources

I. INTRODUCTION

Accurate prediction of porosity is a crucial task in petroleum engineering as it directly affects the estimation of reservoir quality and hydrocarbon potential. Porosity is an essential rock property reflecting the capacity of the reservoir to store hydrocarbons, making it a critical parameter in the exploration and development of oil and gas resources (Al-Khafaji et al., 2024). Well logging is a fundamental technique in the oil and gas industry for estimating the porosity of subsurface formations. Various well logs are routinely run in boreholes to measure properties related to porosity. These logs provide continuous, in-situ measurements that can be directly related to the pore space within rocks. They do not directly measure porosity, but instead record physical parameters (e.g., bulk density, acoustic travel time, hydrogen index) that can be empirically or theoretically related to the porosity of the rock. By integrating data from multiple logs and applying corrections for lithology, borehole conditions, and fluid type, engineers can derive reliable estimates for formation porosity across the entire logged interval (Wang et al., 2024). Conducting experimental tests on core samples and applying well testing methods are other mainstream approaches in estimating the formation porosity. However, these traditional techniques have several significant limitations. Core samples provide localized measurements, which usually cannot represent the entire formation due to heterogeneity (Denney, 2013). Acquiring and analyzing cores is expensive and takes significant time, limiting the number of samples. Moreover, core analysis can be affected by laboratory errors and differences in measurement techniques, leading to inconsistencies in porosity values (Hook, 1983; Bao et al., 2019). Besides, laboratory conditions may not replicate downhole pressures and temperatures, affecting the measured porosity compared to in situ porosity values. The application of most well logging and well testing techniques in an oil field is costly, challenging, and time-consuming (Gamal et al., 2021). Well test methods are often conducted under transient (unsteady state) conditions, which can complicate the analysis and interpretation of in situ test results.

Artificial intelligence (AI) techniques, such as Backpropagation Neural Network (BPNN), General

[1] Birjand University of Technology, Birjand, Iran, [2] Department of Petroleum Engineering, Kho.C., Islamic Azad University, Khomeinishahr, Iran
[3] Stone Research Center, Kho.C., Islamic Azad University, Khomeinishahr, Iran
✉ M. Rahimi: mrahimi@iau.ac.ir, rahimi2726@gmail.com

Regression Neural Network (GRNN), and Support Vector Machine (SVM), offer notable advantages over classical methods used for porosity estimation. AI models can capture complex, nonlinear relationships between well log data and porosity, leading to more accurate predictions, especially in geologically complex or heterogeneous formations (Elkatatny et al., 2018; Zhang et al., 2021). Unlike traditional methods that may be subject to human interpretation and bias, AI algorithms rely on data-driven approaches, leading to more objective results. AI methods can be continuously trained and improved as more data becomes available, allowing them to adapt to changing reservoir conditions and enhance predictive capabilities over time. Once trained, AI models can rapidly process large datasets, providing real-time or near-real-time estimations (Gamal et al., 2021; Iklassov et al., 2022).

Ahmadi and Chen (2019) discussed the comparison of various machine learning (ML) methods for estimating porosity and permeability in oil reservoirs using petrophysical logs. The study adopted several ML techniques, including artificial neural networks (ANNs), genetic algorithms (GAs), and hybrid approaches. The findings indicated that hybridized methods yield reliable estimations, achieving an average relative absolute deviation of less than 1% compared to actual data. The authors emphasized the importance of accurate porosity and permeability estimations for enhancing oil recovery and effectively managing reservoir evaluation. Ayantola and Amigun (2020) discussed the application of ANNs for the accurate prediction of reservoir porosity, particularly in heterogeneous reservoirs, where traditional estimation methods face challenges due to the varying nature of reservoir rocks. Utilizing well log data, such as sonic, resistivity, and density measurements, the study demonstrated how the ANN can effectively predict actual porosity by training on core porosity data from one selected well and validating results against other wells. The results indicated a strong correlation between ANN predictions and actual core porosity values. The study highlighted the cost-effectiveness and accuracy of using ANN for porosity estimation in the oil and gas industry. Chen et al. (2021) presented a study on predicting reservoir porosity from well logging data using a BPNN optimized by a GA (BPNNGA). This approach aims to provide a reliable and effective method for estimating reservoir physical parameters, which is crucial for reservoir characterization and management. The research was conducted in the Chang 8 oil group of the Yanchang Formation in the Ordos Basin, China, utilizing five well logging curves as inputs for the model. The results indicated that BPNNGA offers better accuracy in porosity predictions compared to conventional regression methods and standard BPNN models, achieving an average relative error of 10.77%. Ifrene et al. (2023) reported a study on fracture porosity estimations using a hybrid ML approach in the context of fractured tight reservoirs. They emphasized the importance of fracture porosity for storage and production efficiency. The research explored the effectiveness of combining ANNs and SVM to improve fracture porosity predictions using geophysical image logs from well data, exhibiting promising results with lower error rates compared to traditional methods. The findings suggested that integrating ML with well log analysis can enhance the reliability of fracture porosity estimations, especially in regions lacking advanced logging data. Ardebili et al. (2024) discussed an AI-based approach developed to enhance the estimation of reservoir parameters, such as porosity and shale volume. The study involved measuring the petrophysical properties of 27 samples and utilizing a geostatistical algorithm to augment laboratory data to a total of 686 porosity and 702 shale volume samples. Additionally, it incorporated 2263 well logging data. The optimal multilayer perceptron (MLP) ANN demonstrated significant improvements in estimating these parameters. Mirghaed et al. (2024) presented the enhanced evaluation of petrophysical properties through the integration of ML techniques and well logging data in an Iranian oil field. They highlighted the significance of using advanced logging technologies to identify key petrophysical parameters such as porosity and saturation. The study compared various ML methods to model petrophysical data. The findings showed that the AdaBoost model achieved the lowest error in estimating petrophysical properties, indicating its effectiveness. Ali et al. (2024) discussed the importance of porosity assessment in reservoir evaluation within the oil and gas industry. The research focused on utilizing ML techniques, specifically ANN and fuzzy logic (FL), to improve the accuracy of porosity curve predictions compared to conventional methods like multiple linear regression (MLR). The study adopted various geophysical logs incluading gamma ray, neutron porosity, density, and sonic logs. The results highlighted the superior predictive performance of ANN and FL over MLR. He et al. (2025) elaborated on the importance of accurately predicting porosity in tight reservoirs for the optimization of oil and gas exploration and production. They highlighted the limitations of traditional predictive models, which often struggle with high costs, inefficiency, and accuracy issues. To address these challenges, advanced ML algorithms were applied using well logging data. The study found that the particle swarm optimization-gradient boosting decision tree (PSO-GBDT) model significantly enhances predictive accuracy, achieving an R^2 greater than 0.99.

In summary, while traditional well logging, well testing, and core measurement-based methods provide a practical means and valuable insights into estimating formation porosity, they are limited by costs, challenges,

and reproducibility. The three AI-based techniques, namely BPNN, GRNN, and SVM, have not been used simultaneously in the previous investigations to predict porosity using well logging data. These three AI methods can address the aforementioned challenges by offering higher accuracy, efficiency, and adaptability in porosity prediction from existing field data. This makes AI approaches increasingly attractive for engineers in the oil and gas industry. In line with these motivations, we adopted the above-mentioned three AI methods to predict porosity from well logging data acquired from one of the Kaggle datasets (https://www.kaggle.com/). The dataset belongs to the Peninsular Malaysia hydrocarbon field. To predict formation porosity through these AI techniques, some computer codes were written based on the commands of the Artificial Neural Networks and Machine Learning Toolbox present in MATLAB software. For the sake of comparison, a multivariate linear regression (MLR) model was also developed. The performance of these approaches was evaluated based on the correlation coefficient (R) and root mean square error (RMSE).

II. Artificial Intelligence

Artificial intelligence (AI) is the science and engineering of creating intelligent machines, utilizing computers and modeling human or animal intelligence, with the ultimate goal of achieving AI mechanisms at the human intelligence level. AI techniques were initially developed to solve those types of problems that could not be easily solved by functional programming or mathematical methods. Branches of AI include expert systems, artificial neural networks (ANN), fuzzy logic, genetic algorithms, and many other techniques, each designed based on a specific mechanism from the natural world.

ANNs are designed to simulate the computational behavior of the human brain. Various types of computational models, collectively referred to as ANNs, including multilayer perceptrons (MLPs), radial networks, and support vector machines (SVMs), have been introduced. Each of these models is suitable for a specific category of applications and is inspired by a particular aspect of the capabilities and properties of the human brain (Specht, 1991; Haykin, 1998; Demuth and Beale, 2002). In this paper, the AI methods of MLP, general regression radial networks (GRRN), and SVM are employed to estimate total porosity from well logging data.

A. MLP-Based ANN

Neural networks can be divided into three structures: single-layer networks, multi-layer networks, and networks with competitive layers. The number of layers in the network is defined based on the connection weights in the neurons. A single-layer network includes only one layer of connection weights (Fig. 1), whereas multi-layer networks include more than one layer of connection weights (Fig. 2).

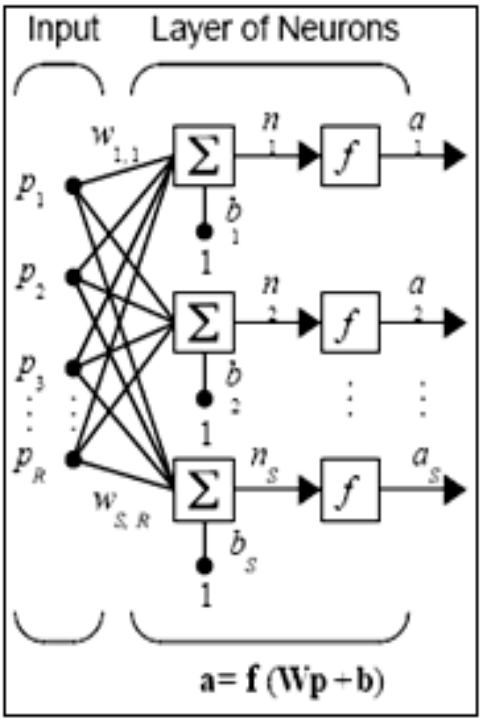


Fig. 1. Single-layer network with R inputs and S neurons in the hidden layer (Demuth and Beale, 2002)

A single-layer network in Fig. 1 has R input units P_1,..., P_R and S neurons. In this network, each input element from the vector p is connected to each neuron input through the weight matrix w. The i-th neuron has a summing unit that sums the weighted inputs and bias to produce a scalar output $n(i)$. The different $n(i)$ values form the net input vector n. Finally, the outputs of the neuron layer form the column vector a.

$$n_j = \sum_{i=1}^{R} (p_i w_{ij} + b_j) \quad , j = 1,2,...S \tag{1}$$

where w_{ij} are the connection weights between the input unit p_i and the output unit n_j, and b_j are the biases related to the j-th cell.

MLP networks can solve more complex problems compared to single-layer networks. Neural networks are divided into two groups based on the way nodes are connected: feedforward networks and feedback networks. In feedforward networks, there is only a one-way flow from the input layer toward the output layer. In contrast, in a feedback network, at least one output signal feeds back to itself or a previous neuron. The structure of different networks is defined by the connection pattern between neurons and layers. BPNNs are among the most widely used MLP networks (Demuth and Beale, 2002).

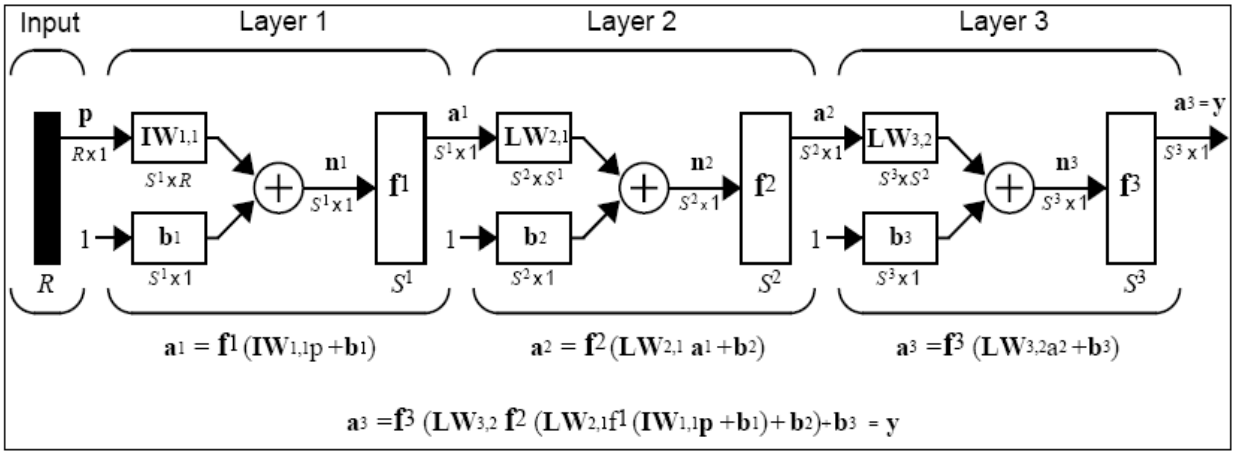


Fig. 2. Structure of a multilayer perceptron (Demuth and Beale, 2002)

B. Radial Basis Function (RBF) Neural Networks

Similar to the architecture of MLP neural networks, there is another type of neural network in which the processing units are focused on specific locations in the input space. This focus is modeled using RBFs. In terms of overall structure, RBF neural networks are not significantly different from MLP networks; the main distinction lies in the type of processing that neurons perform on their inputs. However, RBF networks typically have a faster learning and preparation process. In Fig. 3, you can see an RBF network with *R* inputs.

Here, the input to the network for the radial basis transfer function is the vector distance between the weight vector *w* and the input vector *p*, multiplied by the bias. The box labeled ||dist|| in this figure takes the input vector *p* and the single-row weight matrix *w* and produces their dot product. The transfer function of the radial basis neuron is as follows:

$$radbas(n) = e^{-n^2} \tag{2}$$

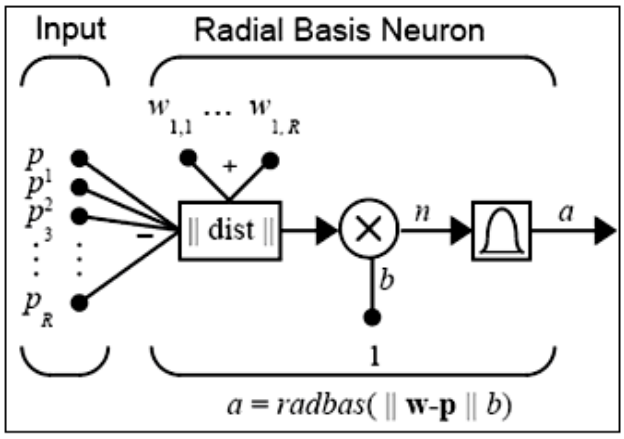


Fig. 3. Neuron model of RBF neural network (Demuth and Beale, 2002)

Fig. 4 shows the plot of the radial basis transfer function.

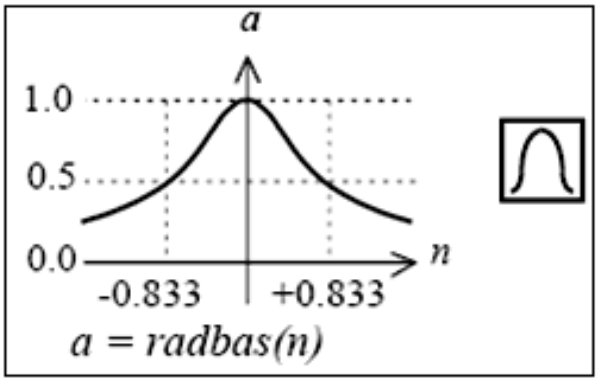


Fig. 4. Transfer function used in RBF neural network (Demuth and Beale, 2002)

As the distance between *w* and *p* decreases, the output value increases. Therefore, a radial neuron acts like a detector, producing a value of one when the input *p* exactly matches its weight. The bias value *b* adjusts the sensitivity of the radial neuron. Radial networks consist of two layers: a hidden radial layer with S_1 neurons and a linear output layer with S_2 neurons (Fig. 5), where R stands for the number of input vector elements, S_1 is the number of neurons in layer 1, S_2 is the number of neurons in layer 2, a_i^1 represents the i-th element of a^1, and $i^{IW_{1,1}}$ denotes the i-th row of the vector $IW_{1,1}$.

The ||dist|| box in this figure receives the input vector *p* and the input weight vector $IW_{1,1}$, producing a vector with S_1 elements. These elements represent the distances between the input vector and the elements of $i^{IW_{1,1}}$ formed by the rows of the input weight matrix. The bias vector b_1 and the output of ||dist|| are then combined using the ".*" operator, which performs element-wise multiplication (Demuth and Beale, 2002).

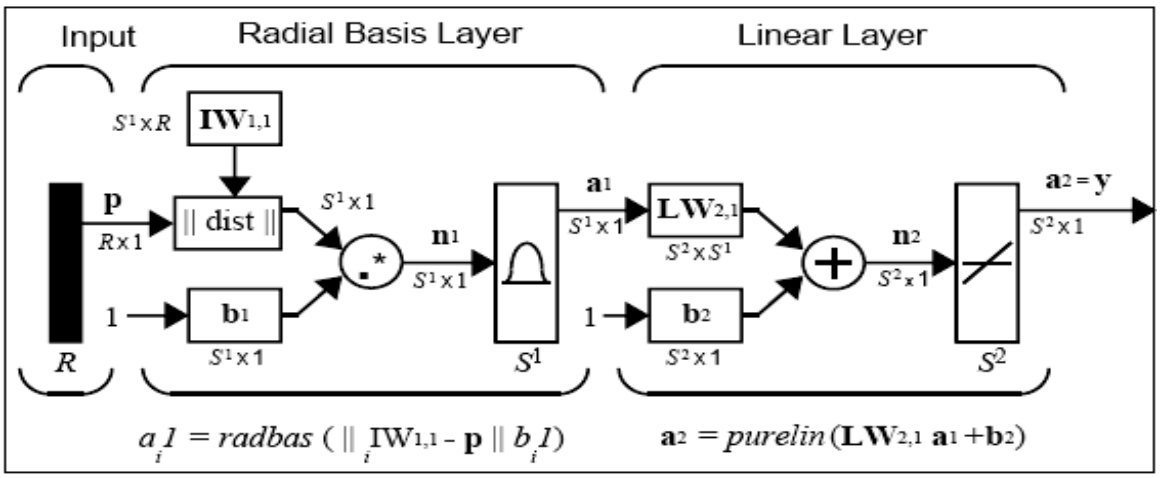


Fig. 5. Structure of RBF neural network (Demuth and Beale, 2002)

The GRNN can be considered as a normalized radial network that has one hidden neuron for each training unit. These networks are based on the probability density function (PDF) and have the advantages of fast training time and the ability to model nonlinear functions. The algorithm form of this network can be used for any regression problem where no assumptions about linearity exist. This network does not have the parameters of error backpropagation networks; instead, it has a smoothing factor whose optimal value is obtained through multiple runs based on the mean squared error. The structure of this network is similar to the general structure of a radial network, with only a slight difference in the second layer (Fig. 6) (Demuth and Beale, 2002).

C. Support Vector Machine

An SVM is a supervised learning tool used for classification and regression. The structure of an SVM network shares many similarities with an MLP neural network, with the main difference being the learning method. An SVM-based classification system uses nonlinear decision functions to transform data into a higher-dimensional space and then linearly partitions them. This algorithm can be expressed for regression purposes as Eq. (5), where $f(x)$ is a function of m variables to be estimated, w is the coefficient vector, b is the bias vector, and ϕ is the kernel function.

$$y = f(x) = \sum_{i=1}^{m} w_i \phi_i(x) + b_i = w^T \phi(x) + b \qquad (3)$$

To estimate the function $f(x)$, w, b, and the kernel function parameters must be calculated based on a set of training data in a way that produces the least possible error based on this model.

The regression error using SVM is expressed as Eq. (6) using Vapnik's ε-insensitive loss function, shown in Fig. 7.

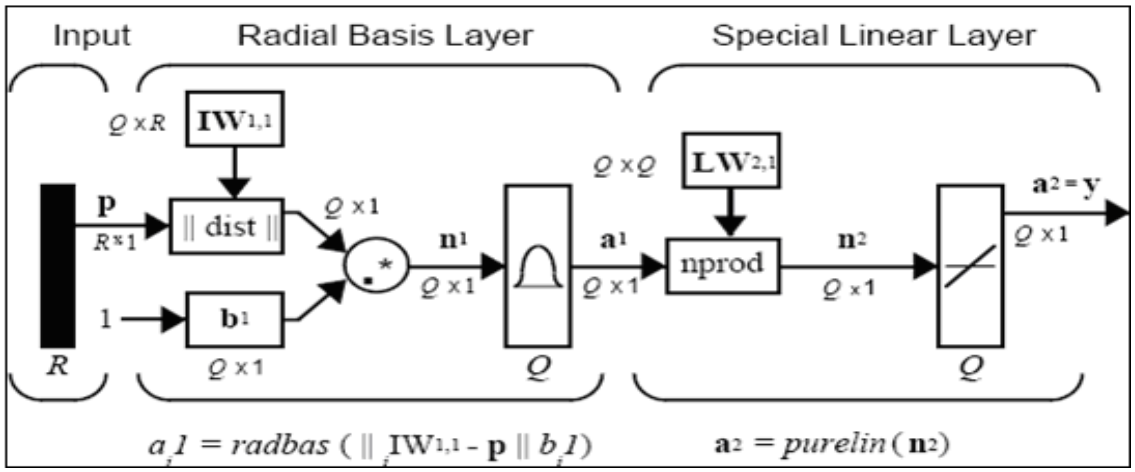


Fig. 6. Structure of GRNN (Demuth and Beale, 2002)

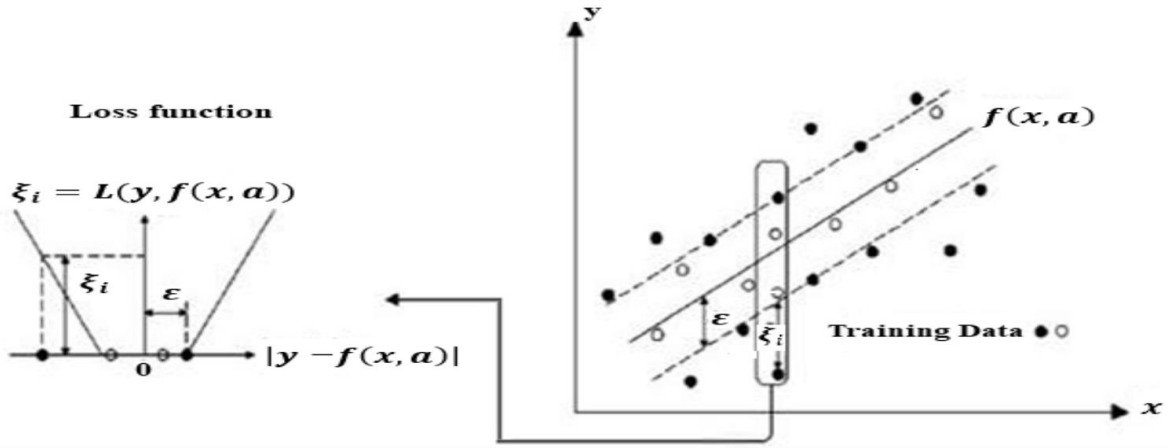


Fig. 7. The procedure for calculating the regression error in SVM, only samples outside the ± range will have non-zero error (ξ) (Liu et al. 2009)

$$\xi = |y - f(x)|_\varepsilon = max\{0, |y - f(x)| - \varepsilon\} \quad (4)$$

In the classical SVM method, vectors w and b are obtained by minimizing the risk function $R(W)$ for a set of N training data.

$$\mathrm{R(w)} = \frac{1}{2}\|w\|^2 + C \cdot \sum_{i=1}^{N} \xi_i \quad (5)$$

The following constraint is implemented while applying Eq. (7).

$$y_i \cdot (w^T x + b) \geq 1 - \xi_i \quad \text{i= 1, … , N} \quad (6)$$

Minimizing the function R in a classical SVM is performed through a second-order programming optimization method, which requires a large amount of computation. This drawback is overcome using the least squares method, which performs the optimization by solving a set of linear equations instead of a second-order program.

III. Data Gathering, and Geological, Lithological, and Petrophysical Properties

This study utilized 1474 data points gathered from one of the Kaggle datasets (https://www.kaggle.com/) to predict total porosity (Phi). The dataset belongs to the Peninsular Malaysia hydrocarbon field, with longitude 104° 28' 48" E and latitude 5° 25' 48" N. The geological, lithological, and petrophysical context of Peninsular Malaysia for oil and gas-bearing formations is influenced by its complex tectonic setting, sedimentary basins, and reservoir characteristics. This context can be summarized as follows:

a) Geological Context: Peninsular Malaysia's geological setting includes three main north-south trending tectonic belts: western, central, and eastern, shaped by Paleozoic to Cenozoic processes. The region has a diverse sedimentary terrain, with significant Paleozoic black shales (about 25% of the area) considered important hydrocarbon source rocks. Structural features like faults and fractures related to regional tectonics control fluid migration and reservoir formation. Offshore basins, such as the Malay Basin, are typical Tertiary sedimentary basins composed of half grabens filled by lacustrine shales and clastics overlain by deltaic systems that act as hydrocarbon source and reservoir intervals (Bishop, 2002; Rashid et al., 2022; Shlof et al., 2025).

b) Lithological Context: Lithologically, the Paleozoic to Mesozoic successions include clastic and carbonate reservoirs. These are interbedded with source rock intervals like Devonian black shales and high total organic carbon (TOC) shales. The offshore sedimentary basins are dominated by alternating sandstones, siltstones, and mudstones, with variations in grain size and consolidation influencing reservoir quality. Coarse-grained sandstones generally present the best reservoir quality due to high porosity and permeability, while fine-grained and bioturbated sandstones exhibit lower reservoir quality owing to clay concentration and cementation (Bishop, 2002; Rashid et al., 2022).

c) Petrophysical Context: Porosity and permeability tend to be highest in coarse-grained sandstones, with values up to about 25% porosity and permeability reaching nearly 1900 mD, supporting good hydrocarbon flow. Fine-grained sandstones and shales tend to have low porosity and permeability, critically impacting fluid flow and storage capacity. Microfractures and fault zones extensively influence reservoir permeability and fluid migration pathways, with fault rocks exhibiting variable permeability depending on mineralogy, pore fluid composition, and stress conditions (Shar, 2015; Ibad and Padmanabhan, 2022).

In summary, Peninsular Malaysia's oil and gas-bearing formations occur primarily within Paleozoic to Tertiary sedimentary successions characterized by rich source rocks like black shales and diverse reservoir lithologies ranging from coarse sandstones to carbonates.

IV. 4. Modeling Porosity Using AI Methods

Some sample data points are given in Table 1. To predict Phi, some computer codes were written using the commands of the Artificial Neural Networks and Machine Learning Toolbox in MATLAB software. Three AI methods were employed in this research: the error backpropagation multilayer ANN (BPNN), the generalized radial regression neural network (GRNN), and the support vector machine (SVM).

In Table 1, GR stands for gamma ray log, SP denotes spontaneous potential log, Rt represents (uninvaded/virgin) formation resistivity, Rxo signifies flushed zone resistivity, DEN symbolizes density log, SONIC indicates sonic log, CNL stands for compensated neutron log, and Phi denotes (total) porosity.

The designed ANNs have three layers (input layer, hidden layer, and output layer). The inputs of the AI methods used are GR, SP, Rt, Rxo, DEN, SONIC, and CNL. The number of neurons in the hidden layer of the BPNN is obtained by trial and error, and in the GRRN is equal to the number of training examples (Fig. 8). The Pearson correlation heatmap between porosity (Phi) and the other parameters is given in Table 2. CNL has the most signicant direct effect on porosity estimation, reflected by the highest positive correlation (0.39) between porosity and CNL. This is because the output of the CNL is a type of porosity obtained based on neutron bombardment of the formation, and the neutrons received by the receiver, depending on the hydrogen content of the formation rock. Moreover, the highest negative correlation (-0.93) exists between the output of the density log and the porosity. This is a thoroughly expected trend because the lower the porosity, the higher the density, and vice versa.

Table 1. Well logging data points and Phi

DEPTH (ft)	GR (API)	SP (mv)	Rt (Ohm-m)	Rxo (Ohm-m)	DEN (g/cm^3)	SONIC (us/ft)	CNL (%)	Phi (%)
5195	91.4	-22.7	2.2	4.3	2	103.8	36.2	45.71429
5195.5	96.1	-25	3.4	6.9	2	104.3	36.7	45.71429
5196	95	-25	2.7	5.4	2	101.5	38.8	45.71429
5196.5	95.2	-23.1	2.9	5.5	2	103.8	36.6	45.71429
5197	92.8	-26.2	2.7	5.7	2.1	109.5	36.8	40
5197.5	95.1	-25.3	3.3	6.8	1.9	100.9	39	51.42857
5198	95.1	-24.5	3.1	6.1	2	110.8	38.8	45.71429
5198.5	93.7	-24.7	2.9	5.8	1.9	100.7	37	51.42857

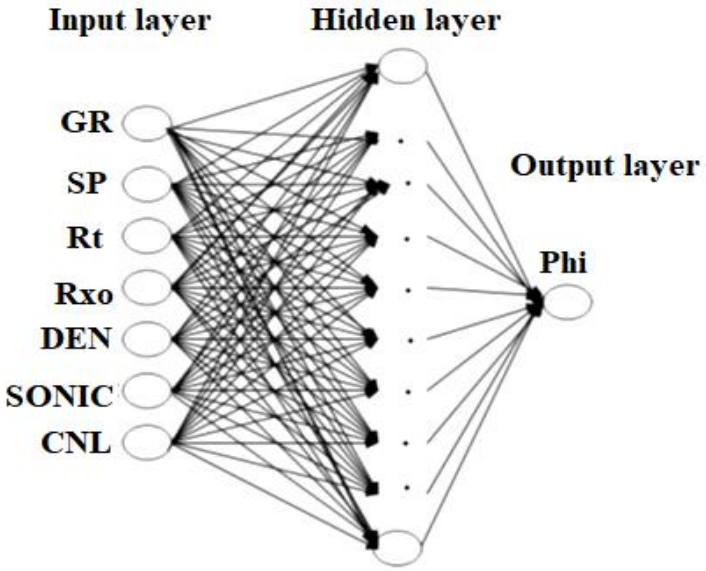
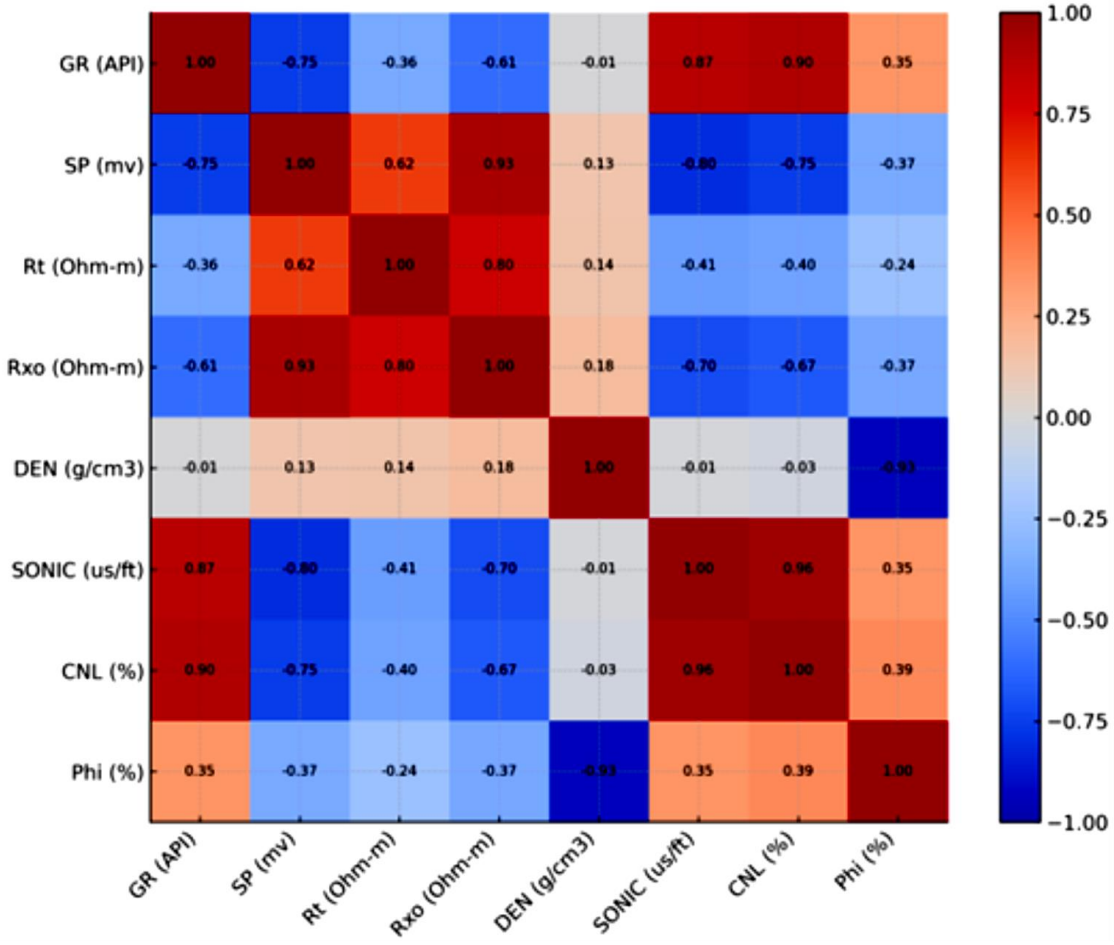


Fig. 8. Schematic diagram of AI methods

Table 2. Pearson correlation heatmap between the dependent variable (Phi) and independent variables



To obtain the optimal structure of the networks and SVM, normalization of the existing data (1474 data) was performed between the interval [-1, 1] using Eq. 9. Then, approximately 70% (1,050 data points) was used for training, and 30% (424 data points) for testing.

$$p_n = 2\frac{p - p_{\min}}{p_{\max} - p_{\min}} - 1 \qquad (7)$$

where p_n is the normalized value, p is the actual value, p_{min} is the minimum of the actual value, and p_{max} is the maximum of the actual value.

Then, three target AI methods were trained with different variable parameters to obtain the optimal parameters, considering the correlation coefficient and error between the actual value and the predicted value in the training and test data. Finally, the optimal BPNN that uses the Bayesian regularization algorithm for training was selected to avoid overfitting. The network was allocated an input layer (with eigth inputs), a hidden layer with 10 neurons with a log sigmoid transfer function (logsig), and an output layer with one neuron with the linear transfer function (purlin). The optimal GRRN was also designed with an input layer (8 inputs), a hidden layer with 1050 neurons with a smoothing factor of 0.08 with the radial basis transfer function (radbas), and an output layer with one neuron and the linear transfer function. The SVM was also considered with eigth inputs, the adjustment coefficient $C = 500$, $\sigma^2 = 0.01$, and the kernel of the RBF type.

The results of the optimal three artificial intelligence (AI) methods for porosity estimation from well logging data reveal varying levels of performance, as quantified by the correlation coefficient (R) and root mean square error (RMSE) in Table 3 and Figs. 9-14. For the BPNN model, Fig. 9 depicts the correlation between predicted and actual porosity values in the training data, where points align closely along the ideal 1:1 line, reflecting an R value of 1 and a low RMSE of 0.02, indicative of excellent fitting without significant deviations. In the test data (Fig. 10), the model maintains this high fidelity, with an R of 1 and RMSE of 0.04, showing minimal scatter and strong generalization to unseen data, which underscores its robustness against overfitting.

In contrast, the GRNN model exhibits near-perfect performance in training, as shown in Fig. 11, with points tightly clustered along the diagonal (R=0.999, RMSE=0.18), suggesting effective capture of underlying patterns in the training set. However, Fig. 12 illustrates a

noticeable degradation in the test data, where increased scatter results in a slightly reduced R of 0.998 and a substantially higher RMSE of 0.57, pointing to some overfitting and reduced predictive accuracy on new data.

The SVM model follows a similar trend, with Fig. 13 demonstrating strong alignment in the training phase (R=0.999, RMSE=0.18), as data points adhere well to the reference line. However, in the test data (Fig. 14), greater dispersion is evident, leading to an R of 0.986 and an elevated RMSE of 1.52, which highlights poorer generalization and higher prediction errors compared to the other models.

Overall, while all methods achieve high accuracy in training, the testing results emphasize BPNN's advantage in maintaining precision, making it the optimal choice for porosity prediction in this context. Further refinements, such as hyperparameter tuning or additional data, could potentially enhance GRNN and SVM performance.

To compare the intelligent methods utilized in this study with statistical modeling methods, the porosity value was also modeled using multivariate linear regression (MLR). In this method, the training data of the intelligent methods were used to create the multivariate regression relationship, and the test data of the intelligent methods were used to test the regression relationship. The multivariate regression relationship was obtained using SPSS statistical software as follows:

$$Phi\ (\%) = 154.145 + 0.028\ GR + 0.007\ SP - 0.00001524\ Rt + 0.032\ Rxo - 59.392\ DEN - 0.007\ SONIC + 0.238\ CNL \quad (8)$$

Table 3. R and RMSE in the training and testing data of the three AI methods and MLR adopted

Database	Method	RMSE	R
Training	BPNN	0.02	1
	GRNN	0.18	0.999
	SVM	0.18	0.999
	MLR	0.72	0.994
Testing	BPNN	0.04	1
	GRNN	0.57	0.998
	SVM	1.52	0.986
	MLR	0.67	0.994

Further, to compare the errors generated by the models while predicting porosity (Phi), a residual plot has been drawn and presented in Fig. 15. As can be observed, the difference between the actual and predicted values in the BPNN method in the test data is around zero. In the MLR, GRNN, and SVM methods, which follow the regression structure, this difference is high for several of the data points. As a result, the RMS values in the three methods in Table 3 are different from each other.

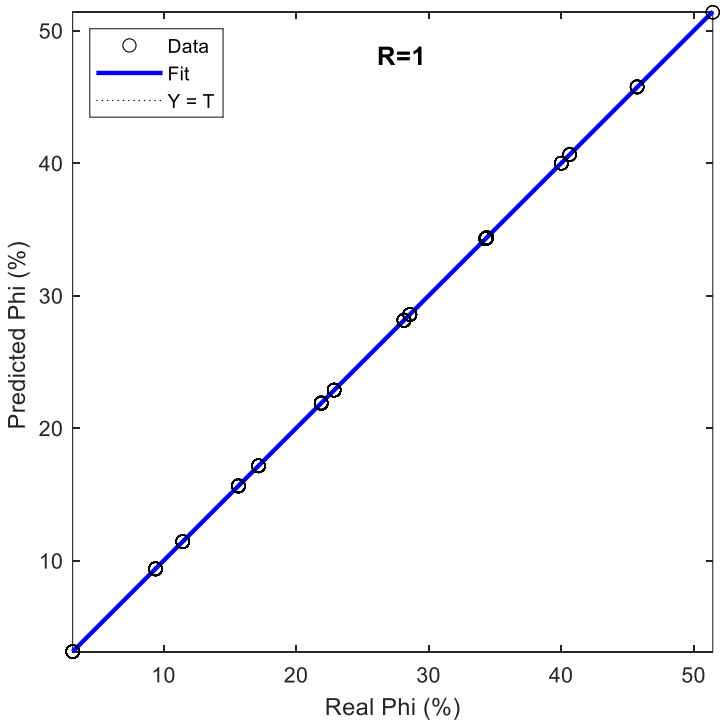


Fig. 9. Correlation between predicted porosity of BPNN vs. actual porosity (training data)

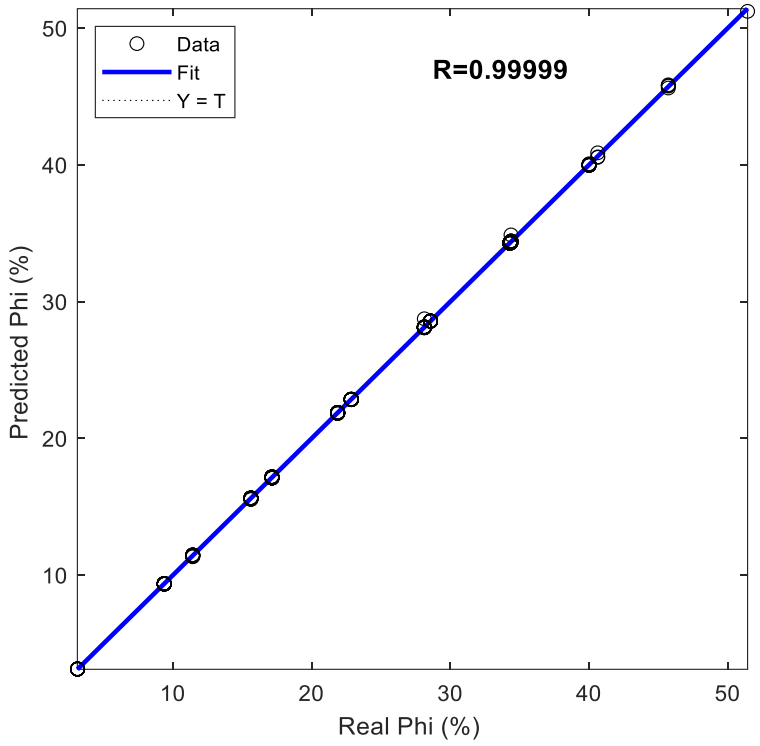


Fig. 10. Correlation between predicted porosity of BPNN vs. actual porosity (test data)

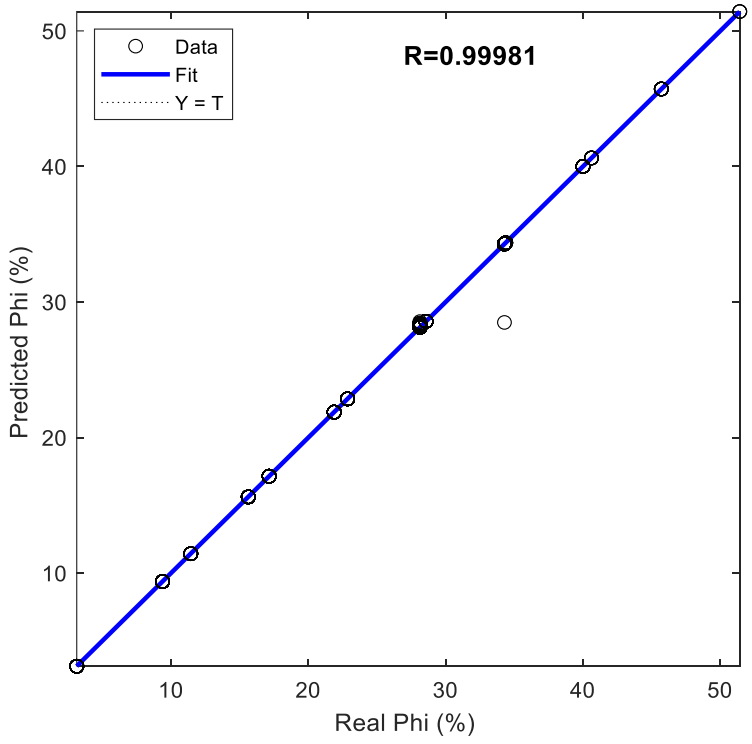


Fig. 11. Correlation between predicted porosity of GRNN vs. actual porosity (training data)

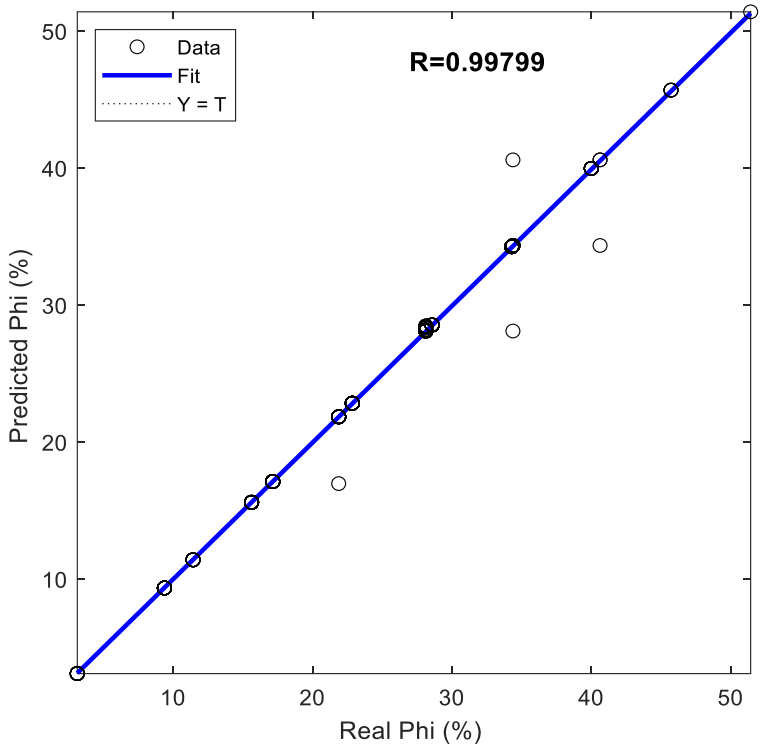


Fig. 12. Correlation between predicted porosity of GRNN vs. actual porosity (test data)

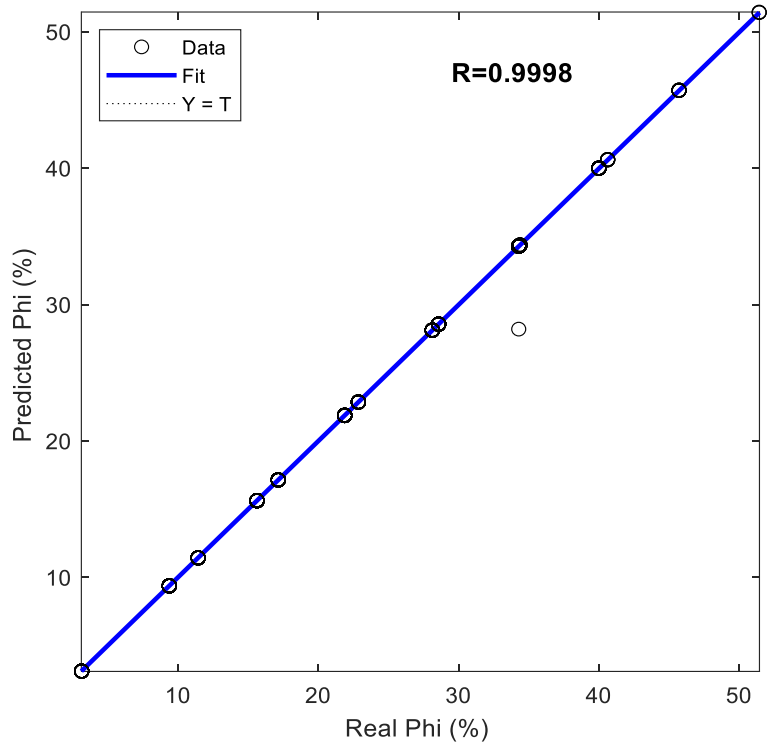


Fig. 13. Correlation between predicted porosity of SVM vs. actual porosity (training data)

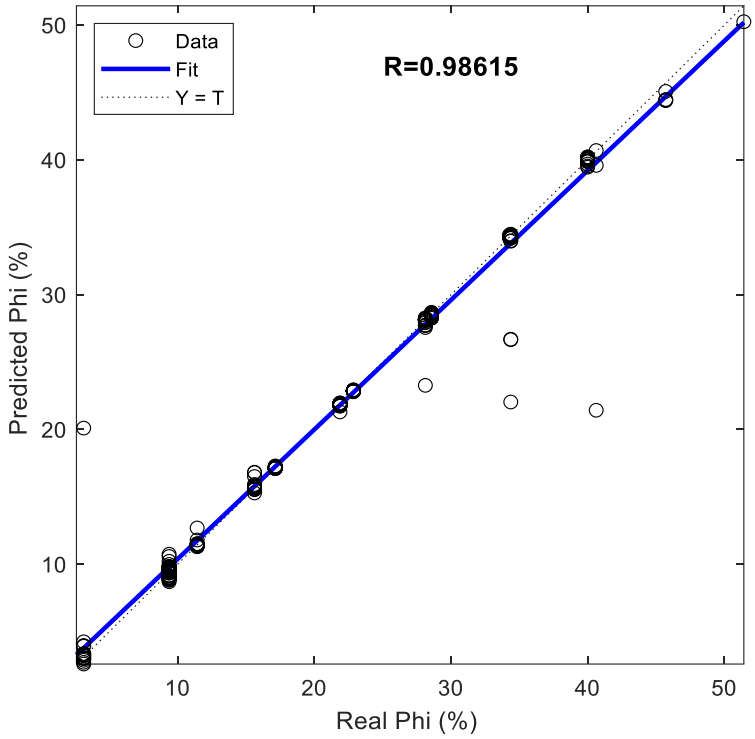


Fig. 14. Correlation between predicted porosity of SVM vs. actual porosity (test data)

In summary, the BPNN exhibited the most stable and accurate performance among all tested models. The GRNN also performed strongly in the training phase (R = 0.999, RMSE = 0.18), suggesting a perfect representation of the underlying data distribution. However, in the testing dataset, its accuracy dropped slightly, with R = 0.998 and RMSE increasing to 0.57. Although the correlation remains high, the increased error indicates reduced predictive accuracy and some degree of overfitting. Figs. 11-12 reveal greater scatter in the predicted versus actual porosity values, confirming this reduction in robustness. The SVM achieved high accuracy in the training dataset (R = 0.999, RMSE = 0.18). Nevertheless, its performance degraded considerably in the test dataset, with R decreasing to 0.986 and RMSE rising sharply to 1.52. This significant loss of predictive power, illustrated in Figs. 13-14, demonstrates that while SVM can model training data well, it is less capable of generalizing to unseen samples compared to BPNN and GRNN. The MLR model was developed using the same training and testing datasets for consistency. During training, the MLR achieved R = 0.994 and RMSE = 0.72, which is considerably weaker than the AI-based models. In the test dataset, the model retained a similar correlation (R = 0.994) with RMSE = 0.67. Although its performance is more stable across datasets than GRNN and SVM, the predictive accuracy remains lower, reflecting the limitations of linear models in capturing nonlinear relationships present in well logging data. The residual plots further emphasize the superiority of BPNN. For this model, the residuals (differences between predicted and actual/real values) are tightly clustered around zero, confirming its strong predictive reliability. In contrast, the residuals for MLR, GRNN, and SVM are more widely scattered, which explains their higher RMSE values.

V. CONCLUSION

Intelligent methods of BPNN, GRNN, and SVM were designed to estimate porosity from well logging data using MATLAB software. For this purpose, 70% of the data was used for training, and 30% of the data was used for testing the AI methods. The BPNN with a Bayesian regularization training algorithm is composed of input, hidden, and output layers. The input layer consists of seven types of well logging records, the hidden layer comprises 10 neurons with a log sigmoid transfer function, and the output layer is composed of one neuron, namely (total) porosity, with a linear transfer function. The GRNN is also similar to the BPNN, with three layers, with the difference that the number of neurons in the hidden layer of this network is the same as the number of training patterns (i.e., 1050 neurons with a radial transfer function and a smoothing factor of 0.08). The inputs and outputs of the SVM were also designed similarly to neural networks, and its optimal parameters were as follows: the adjustment coefficient C=500, σ^2=0.01, and the kernel of the RBF type. According to the comparison of the porosity prediction values of the AI methods and the actual values, we found that the three AI methods predicted porosity with high accuracy. Among the tested methods, the BPNN stands

out as the most effective and reliable approach for porosity estimation. It not only achieved near-perfect accuracy in training but also maintained excellent generalization to testing data, avoiding overfitting. Although GRNN and SVM exhibited promising results in training, their performance degraded in testing, limiting their practical application. The MLR model, while stable, lacked the accuracy and flexibility required for high-quality porosity predictions. These results demonstrate that AI-based nonlinear modeling—particularly BPNN—provides significant advantages over traditional statistical approaches in predicting porosity from well logging data. Future work may focus on optimizing hyperparameters and expanding the dataset to improve the generalization ability of GRNN and SVM models.

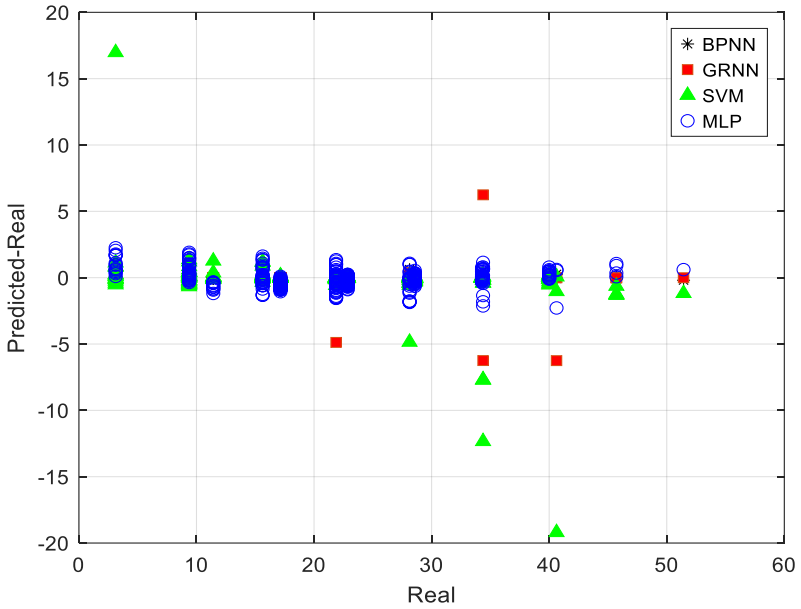


Fig. 15. Residual plot of the models adopted

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmadi, M. A., & Chen, Z. (2019). Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. Petroleum, 5(3), 271-284.

Ali, N., Fu, X., Chen, J., Hussain, J., Hussain, W., Rahman, N., ... & Altalbe, A. (2024). Advancing reservoir evaluation: machine learning approaches for predicting porosity curves. Energies, 17(15), 3768.

Al-Khafaji, H., Meng, Q., Yahya, W., Waleed, S., Hussain, W., AlHusseini, A. K., ... & Al-Khulaidi, G. (2024, September). Advanced Porosity Prediction in Heterogeneous Oil Reservoirs: Using Novel Machine Learning and Deep Learning Techniques. In International Field Exploration and Development Conference (pp. 518-545). Singapore: Springer Nature Singapore.

Ardebili, P. N., Jozanikohan, G., & Moradzadeh, A. (2024). Estimation of porosity and volume of shale using artificial intelligence, case study of Kashafrud Gas Reservoir, NE Iran. Journal of Petroleum Exploration and Production Technology, 14(2), 477-494.

Ayantola, S. O., & Amigun, J. O. (2020). Artificial neural network application for optimum prediction of porosity in heterogeneous reservoir using well logs. Academia. Edu, 8(1), 11-22.

Bao, Y., Li, Z., Yang, Z., QIAN, M., LIU, P., & TAO, G. (2019). Porosity measurement error and its control method [J]. Petroleum Geology & Experiment, 41(4), 593-597.

Bishop, M. G. (2002). Petroleum systems of the Malay Basin province. Malaysia: US Geological Survey Open-file Report.

Chen, L., Lin, W., Chen, P., Jiang, S., Liu, L., & Hu, H. (2021). Porosity prediction from well logs using back propagation neural network optimized by genetic algorithm in one heterogeneous oil reservoirs of Ordos Basin, China. Journal of Earth Science, 32, 828-838.

Demuth H., Beale M., (2002). Neural network toolbox for use with MATLAB. User's Guide Version 4. https://doi.org/10.1002/0470848944.hsa251

Denney, D. (2013). Core-Analysis Elephant in the Formation-Evaluation Room. Journal of Petroleum Technology, 65(08), 85-88.

Elkatatny, S., Tariq, Z., Mahmoud, M., & Abdulraheem, A. (2018). New insights into porosity determination using artificial intelligence techniques for carbonate reservoirs. Petroleum, 4(4), 408-418.

Gamal, H., Elkatatny, S., Alsaihati, A., & Abdulraheem, A. (2021). Intelligent prediction for rock porosity while drilling complex lithology in real time. Computational intelligence and neuroscience, 2021(1), 9960478.

Haykin S. (1998). Neural networks: a comprehensive foundation. Prentice Hall PTR.

https://doi.org/10.1142/s0129065794000372

He, Y., Zhang, H., Wu, Z., Zhang, H., Zhang, X., Zhuo, X., ... & Dang, W. (2025). Porosity prediction of tight reservoir rock using well logging data and machine learning. Scientific Reports, 15(1), 13124.

Hook, J. R. (1983, June). The precision of core analysis data and some implications for reservoir evaluation. In SPWLA Annual Logging Symposium (pp. SPWLA-1983). SPWLA.

Ibad, S. M., & Padmanabhan, E. (2022). Lithofacies, mineralogy, and pore types in Paleozoic gas shales from Western Peninsular Malaysia. Journal of Petroleum Science and Engineering, 212, 110239.

Ifrene, G., Irofti, D., Ni, R., Egenhoff, S., & Pothana, P. (2023). New insights into fracture porosity estimations using machine learning and advanced logging tools. Fuels, 4(3), 333-353.

Iklassov, Z., Medvedev, D., Nazarov, O., & Razzokov, S. (2022). AI for Porosity and Permeability Prediction from Geologic Core X-Ray Micro-Tomography. arXiv preprint arXiv:2205.13189.

Liu, H., Wen, S. M., Li, W. H., Xu, C. B., & Hu, C. Q. (2009). Study on identification of oil/gas and water zones in geological logging base on support-vector machine. In Fuzzy Information and Engineering Volume 2 (pp. 849-857). Berlin, Heidelberg: Springer Berlin Heidelberg.

Rashid, A., Siddiqui, N. A., Ahmed, N., Jamil, M., EL-Ghali, M. A., Ali, S. H., ... & Wahid, A. (2022). Field attributes and organic geochemical analysis of shales from early to middle Permian Dohol Formation, Peninsular Malaysia: Implications for hydrocarbon generation potential. Journal of King Saud University-Science, 34(8), 102287.

Rezaei Mirghaed, B., Dehghan Monfared, A., & Ranjbar, A. (2024). Enhanced petrophysical evaluation through machine learning and well logging data in an Iranian oil field. Scientific Reports, 14(1), 28941.

Shar, A. M. (2015). Petrophysical properties of fault rock-Implications for petroleum production (Doctoral dissertation, University of Leeds).

Shlof, A. M., Arifin, M. H., Zakaria, M. T., & Salufu, E. O. (2025). Exploration of geothermal resources in Peninsular Malaysia: A review of geological, geochemical, and geophysical techniques. Journal of Mining and Environment, 16(2), 405-438.

Specht, D. F. (1991). A general regression neural network. IEEE transactions on neural networks, 2(6), 568-576.

Wang, X., Fu, G., Fan, B., Wang, S., Feng, L., & Peng, C. (2024). Integration of conventional well logs and core samples to predict porosity of tight reservoir: a case study from Ordos Basin. Quarterly Journal of Engineering Geology and Hydrogeology, 57(3), qjegh2023-042.

Zhang, Z., Wang, Y., & Wang, P. (2021). On a deep learning method of estimating reservoir porosity. Mathematical Problems in Engineering, 2021(1), 6641678.