Natural Resources Research (© 2023)
https://doi.org/10.1007/s11053-023-10207-2

Original Paper

Application of GMDH to Predict Pore Pressure from Well Logs Data: A Case Study from Southeast Sichuan Basin, China

Melckzedeck M. Mgimba,[1,2,3] Shu Jiang,[1,3] Edwin E. Nyakilla,[1] and Grant Charles Mwakipunda[1]

Received 10 May 2022; accepted 27 April 2023

Pore pressure prediction is significant in the petroleum industry because, compared to direct measurement, it is cost-effective and it generates an extensive range of data. Mathematical correlations fail to predict pore pressure due to their failure to include lateral transfer in the reservoir, high temperature and mixed lithology and other mechanisms like aqua-thermal expansion, dehydration of clay and mineral alterations. Also, several machine learning techniques provide unsatisfactory results when predicting pore pressures due to poor selection of input data, over-fitting, slow convergence of results, and manual adjustment of model parameters like hidden layers and weights. To counteract these challenges, we employed, for the first time, group method of data handling (GMDH) technique to predict formation pore pressures from well logs data in the Nanye 1 well, southeast of the Sichuan Basin. Then, the performance of the GMDH technique was compared to other machine learning techniques, including polynomial classifier (POL) and artificial neural networks (ANNs). The GMDH technique provided results with the highest accuracy compared with the other two techniques, giving the lowest root-mean-square error (RMSE) of 0.0308 MPa. In addition, the GMDH technique provided a high coefficient of determination of 0.998. The ANN and POL gave RMSEs 0.0322 and 0.5873 MPa, respectively. Apart from the good results, the GMDH technique was able to identify data structure, direct approximate the results, automatically select the model running parameters and select the relevant input data for predicting the pore pressure, which were the challenges for other techniques. Therefore, the GMDH can be applied to predict pore pressure from the well logs data.

KEY WORDS: Machine learning, GMDH, *K*-means clustering, Pore pressure, Well logs.

INTRODUCTION

Pore pressure is the pressure of water, gas, oil, or both, in rock's space (pores and fractures) (Veeken, 2006). This pressure reflects a reservoir's energy and it is the fluid's power pushing in the formation. The pore pressure is part of the overburdened stress covered by liquids and gases in the pore space. The remaining part is covered by the rock, making the in situ rock stress. The overburden stress is formed

[1]Key Laboratory of Tectonics and Petroleum Resources of Ministry of Education, China University of Geosciences, Wuhan 430074, China.
[2]Department of Geosciences and Mining Technology, Mbeya University of Science and Technology, P.O Box 131 Mbeya, Tanzania.
[3]To whom correspondence should be addressed; e-mail: melckmgimba1@gmail.com, jiangsu@cug.edu.cn

Published online: 10 May 2023

by the total weight of a rock's mass comprising the lithostatic column. Thus, the difference between the overburden and vertical rock stresses is an estimate of pore pressure. When the value of pore pressure resembles the value of hydrostatic pressure, the formation is normally pressured. When the pore pressure is greater or less than the hydrostatic pressure, the formation is over-pressured or under-pressured, respectively (Lin et al., 2020).

Pore pressure plays a significant role in oil and gas exploration and exploitation. Prediction of pore pressure is cheaper and it gives an extensive range of data compared to direct measurement of pore pressure (Abbey et al., 2021). Moreover, pore pressure prediction helps drilling engineers, geophysicists, petrophysicists, and geologists plan drilling and production operations. Precise pore pressure prediction can help minimize drilling risk, including pressure kicks and well blowouts, stabilize wellbore, improve casing seat selection, and reduce drilling mud loss (Azadpour et al., 2015; de Souza et al., 2021; Mutumba et al., 2021). Also, precise pore pressure prediction helps in planning daily oil or gas production and cumulative production of hydrocarbon reservoirs (Zou, 2017). Precise pore pressure prediction can also help in planning enhanced oil recoveries techniques such as water and gas flooding (Farsi et al., 2021).

Several methods can be used to predict pore pressure, including Eaton, Bowers, and compressibility (Azadpour et al., 2015). Eaton's method involves determining pore pressure using well logs (sonic transient time and formation resistivity). This technique assumes that overburden pressure is a combination of pore pressure and effective vertical stress, and the effective vertical stress is straightly related to porosity. Therefore, the effective vertical stress is approximated from porosity, and pore pressure can be evaluated from overburden pressure and effective vertical stress as expressed by Terzaghi's and Biot's effective stress law (Biot, 1941; Azadpour et al., 2015), thus:

$$P_p = \frac{\sigma_{ov} - \sigma_{eff}}{\alpha} \tag{1}$$

where P_p is pore pressure, σ_{ov} is overburdened stress, σ_{eff} is effective stress and $\times$ is the Biot effective stress coefficient obtained from volume changes. Predicting pore pressure from the sonic transit time, Eaton (1975) modified Eq. 1 and developed Eq. 2:

$$P_{pg} = \sigma_{ov} - (\sigma_{ov} - P_{hd}) * \left(\frac{\Delta t_n}{\Delta t}\right)^x \tag{2}$$

where P_{pg} is formation pressure gradient, P_{hd} is hydrostatic pressure gradient, Δt is sonic time measured in shale by well logging, \Delta*t*n is sonic transit time, which is measured in shale at the normal pressure condition, and x is a constant (Contreras et al., 2011).

Bowers' method is the one that predicts pore pressure from effective stress, and the method covers the effects of disequilibrium compaction and unloading. Based on Bowers (1995), the relationship between effective stress and sonic velocity can be shown as:

$$V_d = V_0 + C\left[\sigma_{max}\left(\frac{\sigma_v}{\sigma_{max}}\right)^{\frac{1}{U}}\right]^D \tag{3}$$

$$\sigma_{max} = \left(\frac{V_{max} - 5000}{C}\right)^{\frac{1}{D}} \tag{4}$$

where V_d is velocity at a certain depth, V_0 is velocity at the surface, σ_v is effective stress in the vertical, C and D are constants obtained after adjusting regional offset velocity vs. effective stress data, and U is the parameter that represents unloading. When U is equal to 1, it indicates temporary deformation; when U is equal to infinity, it represents permanent deformation. The σ_{max} is the greatest effective stress and Vmax is the maximum speed when unloading first begins (Bowers, 1995).

Rock compressibility is used to determine pore pressure in the compressibility method. Pore pressure is a function of changes in pore volume. These changes depend on the compressibility of rock and fluid. Pore volume diminishes when a formation experiences high compression, and it leads to the emergence of over-pressure in the formation. Therefore, compressibility is applied as a parameter to evaluate pore pressure. According to Zimmerman (1990) and Atashbari and Tingay (2012), pore pressure can be predicted depending on compressibility, thus:

$$P_p = \left(\frac{1(1-\emptyset)C_b\sigma_{eff}}{(1-\emptyset)C_b - \emptyset C_p}\right)^{\gamma} \tag{5}$$

where $\emptyset$ represents porosity, C_b represents bulk compressibility in psi^{-1}, C_p represents pore compressibility in psi^{-1}, σ_{eff} defines effective vertical

pressure in psi,[1] and γ represents a constant within the range from 0.9 to 1.0.

Prior studies documented the failure of above-discussed methods to work in some formations because variables such as lateral transfer in a reservoir, high temperature, and mixed lithology were not considered (Swarbrick, 2001, 2012). In addition, the above-discussed methods do not consider other over-pressure mechanisms like expansion due to aquathermal, clay dehydration, and mineral and osmosis alterations, which are not associated with disequilibrium and non-mechanical compaction. Therefore, the current utilization of machine learning (ML) techniques to predict pore pressure seems to eliminate the aforementioned challenges. These techniques involve learning and adopting the condition of formation and then predicting a specific output (pore pressure).

ML techniques are applied to predict pore pressure to counteract the mentioned challenges. These techniques involve learning and adopting the condition of formation and then predicting a specific output (pore pressure). One of the works that used artificial intelligence in predicting pore pressure was by Hutomo et al. (2019), who predicted pore pressure by combining the Eaton method and artificial neural network (ANN). In their work, seismic data, including shear impedance, acoustic impedance, seismic amplitude, and seismic frequency, were trained to predict pore pressure. Ahmed et al. (2019a, 2019b) used drilling parameters and well logs data to predict pore pressure using ANN. Hu et al. (2013) used a backpropagation (BP) neural network to estimate fluid pressure in a formation from gamma ray, formation density, depth, and interval transit time; in their work, pore pressure was predicted with an error of 4.61%. Keshavarzi and Jahanbakhshi (2013) predicted pore pressure gradient in the Asmari oil field in Iran by applying a backpropagation artificial neural network (BPANN) and the Eaton method. The BPANN showed better performance compared to the Eaton method.

According to the literature on ML techniques, the performance of support vector machine (SVM) is slightly better than ANNs (Ahmed et al., 2019a, 2019b). However, Yu et al. (2020) evaluated the performance of optimization of proposed ML techniques, namely a multilayer perceptron neural network (MLPNN), SVM, random forest (RF), and gradient boosting. Their findings showed that the RF model surpassed the competition regarding prediction correctness, generalizability, and good fitness. Further, in investigating a better model, Farsi et al. (2021) applied three ML techniques, namely MLPNN, least square support vector machine (LSSVM), and multiple-hidden-layer-extreme learning machine (MELM), in predicting pore pressure from well logs data. All these techniques were optimized by particle swam optimization (PSO). As a result, the work found the best predictions (with root mean square error (RMSE) of 11.551 psi) using MELM compared to the other techniques of predicting pore pressure. Lastly, pore pressure predictions by other ML techniques, namely RF, support vector regression (SVR), ANN, and decision tree (DT), were conducted by Zhang et al. (2022), who found that DT produced the best predictions with R^2 and RMSE of 0.9985 and 14.460 psi, respectively.

All assessments on ML techniques revealed a better performance made by the models. However, the techniques encounter the problem of automatic selection of relevant input data, over-fitting and slow convergence because of the manual adjustment of model parameters, including several hidden layers, weights, and biases (Asante-Okyere et al., 2020; Mulashani et al., 2021). To address these issues, some prior authors suggest utilization of a new ML technique called the Group Method of Data Handling (GMDH) (Ivakhnenko and Ivakhnenko, 1995; Srinivasan, 2008; Shaghaghi et al., 2017). Based on the findings, GMDH has the ability to overcome the problem of the single output in many input data (Najafzadeh et al., 2015) and can be used for nonlinear input–output relations (Menad et al., 2019). Additionally, this technique showed better performance in predicting other oil and gas parameters like total organic carbon (TOC) (Mulashani et al., 2021), hydrate formation temperature (Mesbah et al., 2022), standpipe pressure (Youcefi et al., 2022), methane adsorption capacity (Nait Amar et al., 2022) and reservoir permeability (Mulashani et al., 2022).

Therefore, for the first time, we applied the GMDH techniques to predict the formation pore pressure from the well logs data. Then, this technique's performance was compared with other ML techniques, the polynomial classifier (POL) and artificial neural network (ANN). Reservoir engineers can acquire new knowledge of predicting pore pressure from well logs data through this study using this suggested technique (GMDH). But also, experts

[1] * 1 psi = 6894.76 pascals (Pa).

will increase their awareness of the influences of different formations and fluid properties on predicting pore pressures.

GEOLOGICAL SETTING

The data used in this study were obtained from Nanye 1 well located in the southeastern Sichuan Basin in an area called Nanchang, as depicted in Figure 1. Sichuan Basin is located in the southwestern China and northwestern Yangtze Platform. It is a giant and stable intra-cratonic basin in the South China Block. The basin covers an area of 23×10^4 km^2, and in 2014, the basin was approximated to have a gas reserve of around 3.22×10^{12} m^3. The sediments deposited in a region for a long time, and the basin formed from the late Proterozoic to recent years (Korsch et al., 1991; Liu et al., 2017). Intermediate and intermediate basite magmatic rocks with extreme metamorphism are found in the basement in the basin's center (Zhili, 1998). Mountains on all sides bound Sichuan Basin: in the north, south, west, and east, the basin is bounded by Micang and Daba mountains, Daliang mountain, Longmen mountain, and Dalou Mountains, respectively, as shown in Figure 1 (Xu et al., 2018).

Deposition of the Sichuan Basin passed several tectonic evolution histories. A craton basin stage and a foreland basin stage are some periods of tectonic evolution in the Sichuan Basin. Marine carbonate rocks were formed in the craton stage, and terrestrial clastic rocks were formed in the foreland stage. Some formations are missing from some basin areas due to tectonic uplift and erosion. Also, the tectonic uplift caused some beneath-surface formations to be exposed on the surface (Yi-Feng et al., 2015).

The region is also comprised of successions of tectonic uplift and erosion. Figure 1: Location of Nanye 1 (NY 1) well in the southeast of the Sichuan Basin (modified from Nie et al. (2017)); the data from this well were used in pore pressure prediction.

Figure 2 shows that this region's tectonic uplift and erosion started in the Devonian and Carboniferous periods. Whereas in the Permian and Early Triassic, significant erosion appeared. In the late Triassic, another tectonic uplift occurred; during this time, North China and South China Plates collided (Qi et al., 2015; Wang et al., 2015). According to Wang et al. (2015), the region was compressed during the Early to Middle Jurassic, creating a number of thrust faults. During the Cretaceous, the stress behavior of the thrust faults transitioned from extrusion to extension. However, the stress in the Neogene changed once more from extension to extrusion as a result of the deposition of the India–Australian intrusive plate into the Eurasian plate (Wo et al., 2007). Tectonic uplift and erosion influenced a reduction of the formations' pore pressure. According to Liu et al. (2014), the tectonic uplift and erosion reduce the temperature, which results in reducing the pore pressure. Also, tectonic uplift and erosion cause a reduction in pressure by enlarging a trap space and pore volume resilience (Xu et al., 2018).

Based on Figure 1: Location of Nanye 1 (NY 1) well in the southeast of the Sichuan Basin (modified from Nie et al. (2017)); the data from this well were used in pore pressure prediction.

Figure 2, the primary source rocks in this area are Longmaxi, Wufeng, Longtan, and Maukou. The Early Silurian era saw the deposition of the Longmaxi shale in the sedimentary contexts of deep-sea shelves. These shales underwent profound burial during the Yanshan epoch. However, the hydrocarbons migrated out of the formation as a result of tectonic uplift and erosion (Hailong et al., 2012). Due to foreland uplifts in the orogenic belts during the Late Ordovician period in the foreland basins, the Wufeng shale was created (Jing et al., 2016).

METHODOLOGY

This study involves several steps to reach its goal, including clustering the data into a group whose data are related using the *K*-means technique and predicting the pore pressure using three different techniques: POL, ANN, and GMDH. After obtaining the results from all techniques, each technique's error in the predicted results based on the measured pore pressure was found by calculating the RMSE. Therefore, the techniques were compared based on the results and the RMSE. The well-log data that were used in this work were acoustic log (AC), caliper log (CAL), neutron porosity log (CNL), gamma ray log (GR), density log (DEN), laterolog deep (LLD), laterolog shallow (LLS), resistivity log (RS), shallow formation resistivity (RXO), true formation resistivity (RT), spontaneous potential (SP) and micro-spherical focused log (MSFL). The well logs varied with depth, and the

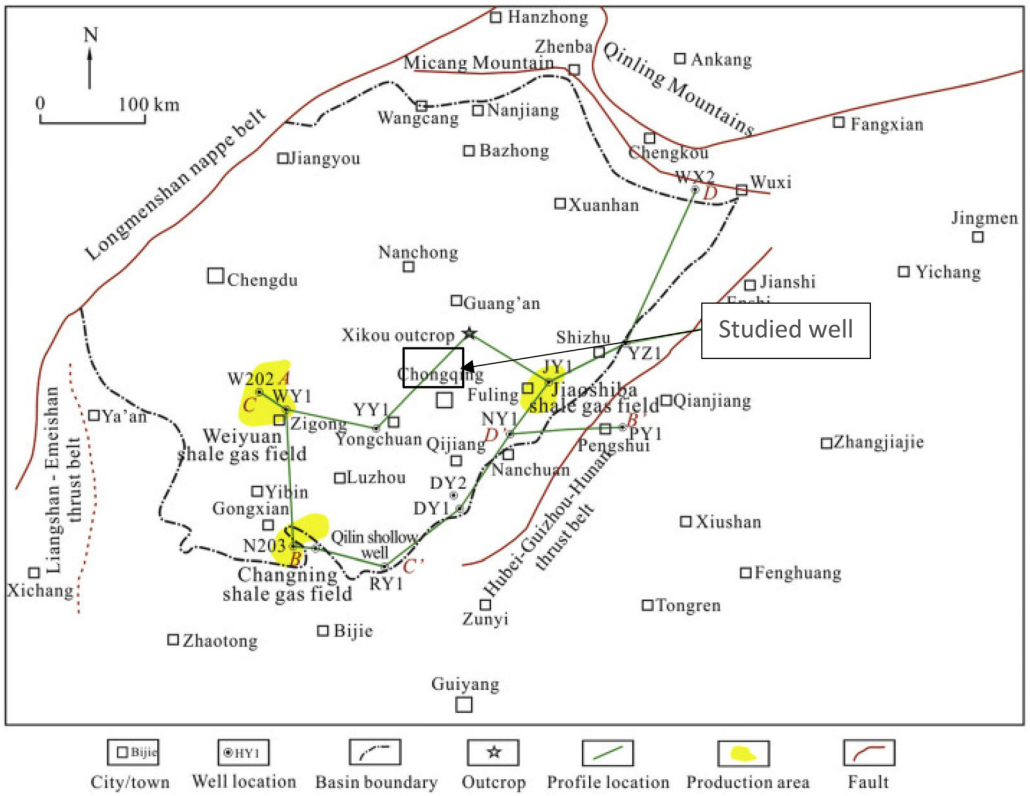


Figure 1. Location of Nanye 1 (NY 1) well in the southeast of the Sichuan Basin (modified from Nie et al. (2017)); the data from this well were used in pore pressure prediction.

depth had an influence on the overburden stress; thus, depth was added to the training data.

Clustering Data by Using the *K*-Means Method

The data were recorded from 1544 to 4410.754 m depth, but data from 2870.154 to 4071 m depth were unavailable. Due to the unavailability of data for that depth range, well logs data varied significantly between the top and bottom depths. This variation created difficulty in training the well logs data to predict pore pressure. Due to this difficulty, the *K*-means method was used to select the groups of related data to train and predict pore pressure.

The *K*-means method was first used in signal processing. It was used to divide the data into several groups in which each datum belonged to the group with the nearest centroid while controlling the overlapping of data in the divided groups. These divided groups are referred to as clusters (Al-Mohair et al., 2015). The method involves organizing the data, finding the centroids, dividing the data depending on the closest centroid, and then deciding the number of clusters. The clustering of the data followed the procedure shown in Figure 3.

A cluster's centroid was determined as:

$$DI = \sqrt{(x-a)^2 + (x-b)^2 + (x-n)} \qquad (6)$$

where DI is the Euclidian distance of the selected data and x, a, b and n are data. The sum of squared error recognized the centroids of a cluster by minimizing the objective function (F) as (Nyakilla et al., 2022):

$$F = \sum_{i=1}^{k} \sum dist\left(G_p, N(i)\right)^2 \qquad (7)$$

where F is the sum of square error of all pressure data selected, G_p is pore pressure data, and $N(i)$ represents the centroid. Then, the standard deviation and weighted summation were determined, respectively, as:

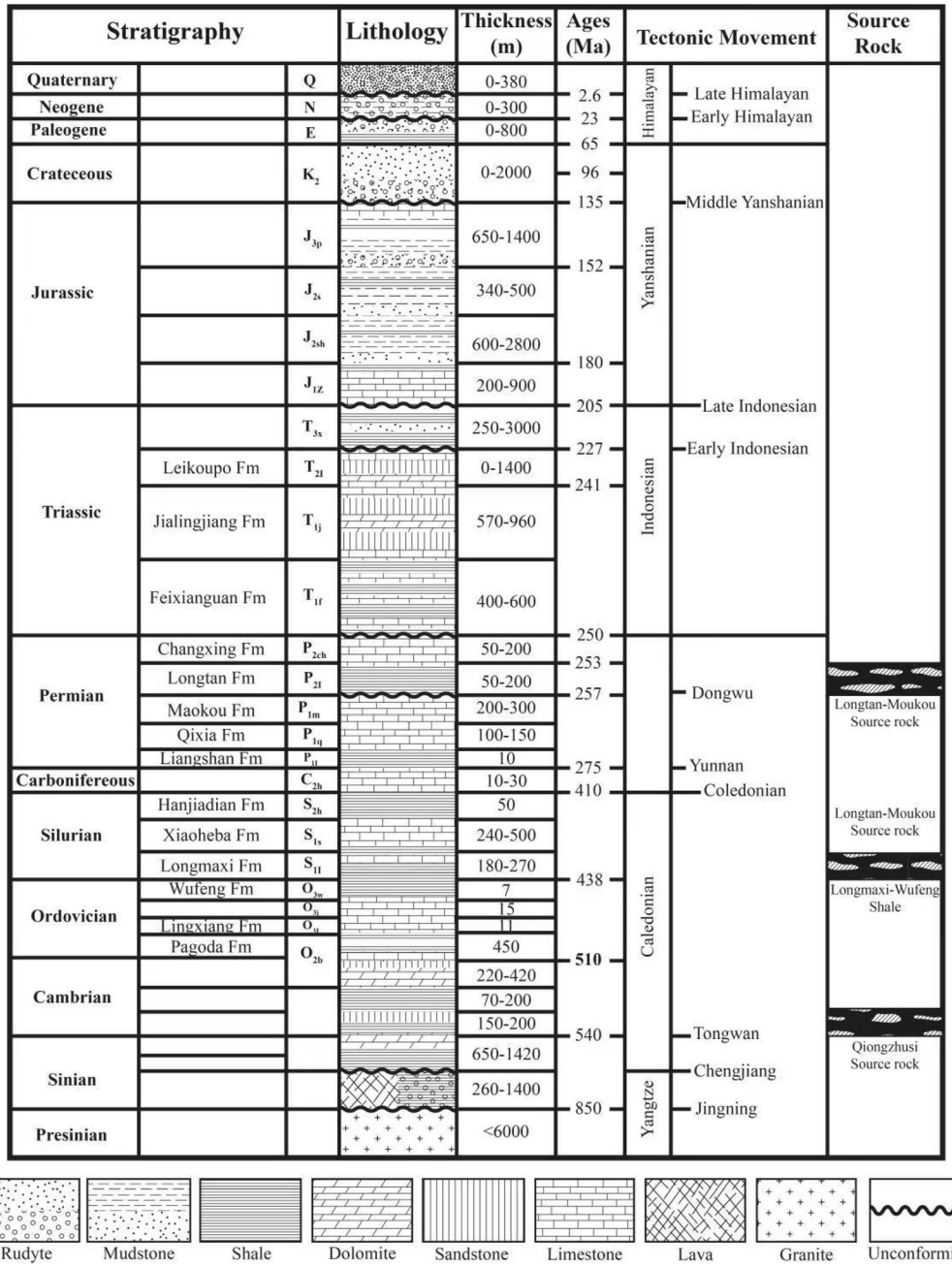


Figure 2. Stratigraphy of southeastern Sichuan Basin showing the potential hydrocarbon regions (modified from (Cao et al., 2020)).

$$S = \frac{\sqrt{\left(\sum (x_i - \overline{x})^2\right)}}{mv - 1} \tag{8}$$

$$A = \frac{1}{mv} \sum_{i=1}^{mv} a(i) \tag{9}$$

where A represents the weight, $a(i)$ represents the sub-variables, mv is the number of variables, S is the

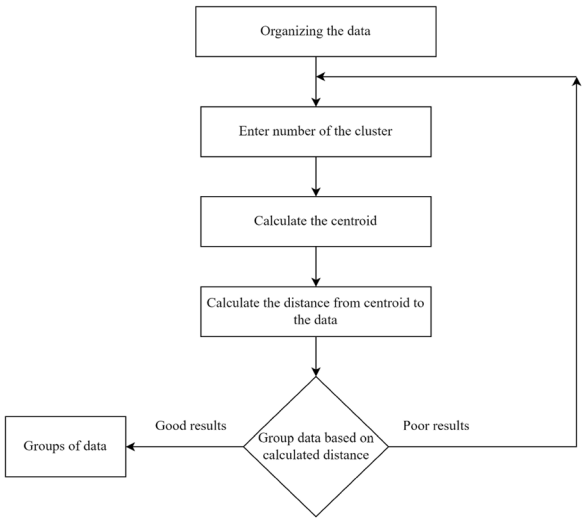


Figure 3. Procedure for clustering of the data.

standard deviation, x_i is the value of variables in the pressure data and $\bar{x}$ is the average of values of variables in the pressure data. The parameters S and A were found to ensure that the selected data in the cluster are close to each other, with no significant deviation. The algorithm for K-means in this work was run in MATLAB (R2021a).

Polynomial Classifier Methods (POL)

The polynomial classifier was the method that was applied in nonlinear output relationships. The application of this method showed success in the analysis of medical images and the prediction of football match results (Do Nascimento et al., 2013). The accuracy of this method increases as the degree of the polynomial function increases. The polynomial function used in the prediction of output results was:

$$g = w_o + \sum_{i=1}^{t} w_i x_i + \sum_{i1=1}^{t}\sum_{i2=1}^{t} w_{i1} x_{i1} x_{i2} + \sum_{i1=1}^{t} \dot{s} \sum_{\text{in}=i1-1}^{t} w_{i1\dot{s}iL} x_{i1\dot{s}} x_{iL} \quad (10)$$

where w_i represents the weight and x_{iL} represents the input data. The offset and the first sum define a first polynomial order, and the second sum defines the second order of polynomials, such as x_{12}, x_1x_2. The last sum defines the lth polynomial order term such as x_{11}, $x_{11\text{-}1}x_2$. The polynomial classifier was solved by first organizing the input and output of the training data, respectively, as:

$$\text{Input} = \begin{bmatrix} \text{x}_{11} & \text{x}_{21} & \text{x}_{31} & . & \text{x}_{\text{m}1} \\ \text{x}_{12} & \text{x}_{22} & \text{x}_{32} & . & \text{x}_{\text{m}2} \\ . & . & . & . & . \\ . & . & . & . & . \\ \text{x}_{\text{i}N} & \text{x}_{2N} & \text{x}_{3N} & & \text{x}_{\text{m}N} \end{bmatrix}_{\text{Nxm}} \quad (11)$$

$$Output(y) = \begin{bmatrix} P_1 \\ P_2 \\ . \\ . \\ P_N \end{bmatrix}_{Nx1} \quad (12)$$

where N represents the number of samples and m represents the number of features. Then, the matrix was expanded to the higher-order polynomial matrix (B), thus:

$$B = \begin{bmatrix} 1 & x_{11} & x_{21} & . & x_{m1} & x_{11}x_{21} & . & x_{11}x_{(m-1)1} & x_{11}^2 & x_{m1}^2 \\ 1 & x_{12} & x_{22} & . & x_{m2} & x_{12}x_{22} & . & x_{12}x_{(m-1)2} & x_{12}^2 & x_{m2}^2 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 1 & x_{1N} & x_{2N} & . & x_{mN} & x_{1N}x_{2N} & . & x_{1m}x_{(m-1)N} & x_{1N}^2 & x_{mN}^2 \end{bmatrix}_{Nx(2m+1+\frac{m(m-1)}{2})} \quad (13)$$

The weight (w) was then calculated from the training data, thus:

$$w = \left(B^T B\right)^{-1} B^T y \quad (14)$$

Finally, the weight and testing data were used to find the output (predict the pore pressure) (Abu-Kheil, 2009; Martins et al., 2017), thus:

$$Y_m = w * B \quad (15)$$

The algorithm of POL was run in MATLAB (R2021a) to approximate pore pressure.

Artificial Neural Network (ANN)

Another method that was used for pore pressure prediction was ANN. This method is very popular for nonlinear system identification (Korbicz and Mrugalski, 2008). The method originated from neurobiology, and the architecture of the human brain has influenced it. Therefore, it has the intelligence and nonlinear characteristics of the brain. ANNs employ Eq. 16 to predict an output, where x is the input, w is the weight allocated to each input, p is the feature bias and $\text{h}_{\text{w},p}(x)$ is the output:

$$h_{w,p}(x) = f\left(w^T x + p\right) = \frac{1}{1 + exp(-(w^T x + p))} \quad (16)$$

ANNs consist of various elements interconnected and organized together with weighted networks, which are referred to as neurons (Liu, 2001). The neuron in a system functions by capturing the total of its weighted inputs and multiplying it with the inputs in a nonlinear activation function to obtain the output, as shown in the ANN architecture in Figure 4.

Neural Network Architectures

Neural networks have various categories that vary in their architectures, practices, and uses. The most used neural networks are feedforward and recurrent neural networks (Murphy, 2002). A feedforward network contains input and output layers, and between them there are hidden layers. These hidden layers connect the input layers to the output layer (Rahim et al., 2006). A multilayer feedforward network assembly can be depicted in Figure 5a. The recurrent network is the same as the feedforward network, except that it contains a loop that carries information from the output layer to the input layer or the inverse. Due to that, this neural network is called dynamic network (Fig. 5b).

Selecting the Quality Data and Running the Algorithm

The important step in this method is selecting high-quality data and then running the algorithm. The quality of this method depends on the quality of data obtained from experiments or simulations. Thus, the quality and valid data were selected, whereas invalid data (data like − 999.25 and − 9999.25, which had no physical meaning) were removed before the training and testing began. After selecting the data, the data were divided into three groups: training, validation and testing. In the training group, the weight and threshold values were calculated. The validation data were used to check the results of the training process and to improve the model (weights). Then, the evaluated weights were used to predict pore pressure in both training and testing groups (Fig. 6). These procedures were followed, and the algorithm was run in MATLAB (R2021a) to predict pore pressure.

Group Method of Data Handling (GMDH)

GMDH is another nonlinear trained machine method that contains multiple layers. The method uses a polynomial function to solve problems, and it can have a higher order of polynomials without mathematical complications (Korbicz and Mrugalski, 2008). It is designed to capture the relationship between one output and more inputs. Also, the method can identify which inputs influence the output. The network contains layers, and each layer has neurons with two inputs. The output of each neuron uses the quadratic equation from the two inputs in the neurons. Linear regression analysis is used to obtain the quadratic functions, and before proceeding to the next layer, the earlier layer is trained. Each training neuron is unique in the training process, with the best performance chosen. After selecting the neuron with the best performance, the coming layer proceeds, and the training is repeated. The addition of layers continues until the stopping criteria are achieved (Abu-Kheil, 2009).

The output of each neuron is determined using the Ivakhnenko polynomial, thus:

$$\begin{aligned} y &= f(x_1, x_2) \\ &= w_o + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 x_1^2 + w_5 x_2^2 \end{aligned} \quad (17)$$

First, the complete GMDH network combines the two input variables in each layer. Then, the least square fitting method evaluates the polynomial coefficients for each group and its associated output. Finally, when the polynomial coefficients are evaluated, the external criteria of accuracy are used to evaluate and test the results. The criteria to analyze the model's accuracy is:

$$RE^2 = \frac{\sum_{i=1}^{N} (P_i - y_i)^2)}{\sum_{i=1}^{N} p_i^2} \quad (18)$$

where RE represents the regularity criteria measure, N is the sample number, P is the required results, and y is the prediction by the GMDH neuron. This equation calculates and tests the output of each neuron. The results of this equation can indicate which input is more relevant to the network. Also, it reflects the ability of the neuron polynomial to predict the required output. A minor value of the answer indicates better fit, whereas a larger value indicates worse fit.

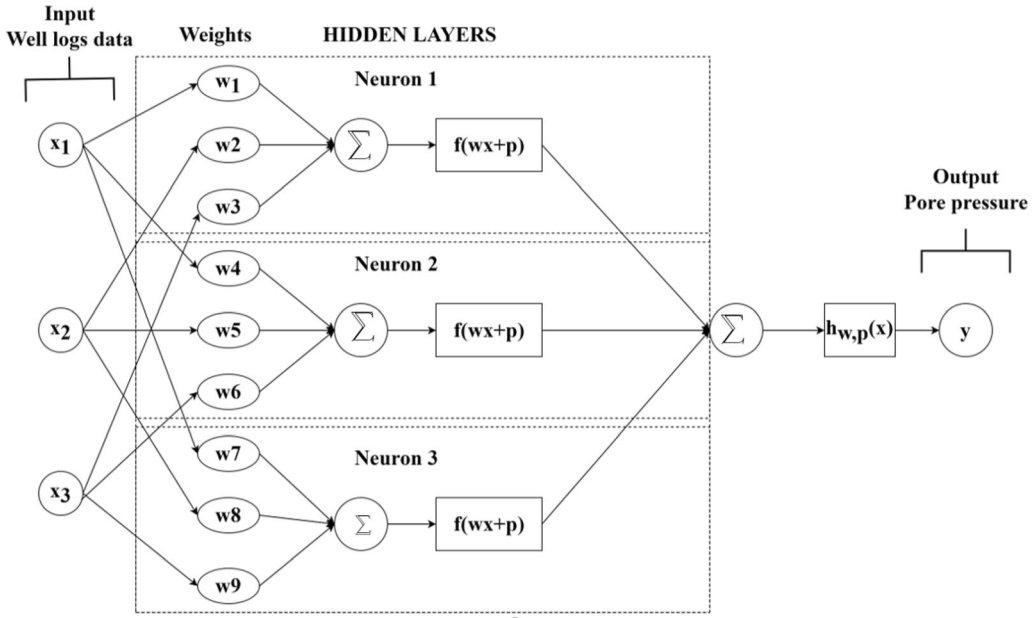


Figure 4. Architecture of ANN.

Equation 18 is used to quantify the influence of each input data on the output and it helps in the selection of proper input data to be used in predicting the output. Also, the number of neurons and layers are chosen automatically depending on the value of RE^2 obtained from Eq. 18. A neuron will be selected when RE^2 is less than a predefined threshold value, whereas neurons with larger values will be eliminated (Fig. 7). Also, the layer with the smallest RE^2 is saved. If the value of RE^2 in the coming layer is larger than RE^2 in the earlier layer, then the new layer will stop. The final results of the GMDH network will be the output of the neuron with the minimum RE^2 in the last layer (Mulashani et al., 2021).

The implementation of this method followed several steps. The first step involved isolating the data into training and testing data. The training data help to evaluate the weight of GMDH neurons, whereas the testing data were used to predict pore pressure. The data division was done randomly, whereby 70% of the data were training data and 30% were testing data. The second step involved finding the combination of the two inputs among all inputs using:

$$Combinations = \frac{mi(mi-1)}{2} \tag{19}$$

$$Z = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{11}x_{21} & x_{11}^2 & x_{21}^2 \\ 1 & x_{12} & x_{22} & x_{12}x_{22} & x_{12}^2 & x_{22}^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{1N} & x_{2N} & x_{1N}x_{2N} & x_{1N}^2 & x_{2N}^2 \end{bmatrix} \tag{20}$$

where m_i is the number of input variables. Then, each combination was expanded to a quadratic polynomial (Z). Then, the polynomials' weights were found using: $w = \left(Z^T Z\right)^{-1} Z^T y$ (21)

The third step was to determine and test the results of each polynomial using the selected data in the testing group. Finally, the pore pressure prediction was evaluated as:

$$Y_m = w * Z \tag{22}$$

The GMDH technique was written and run in MATLAB (R2021a) to predict pore pressure. The performance of each model was evaluated by using (Nyakilla et al., 2022):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2} \tag{24}$$

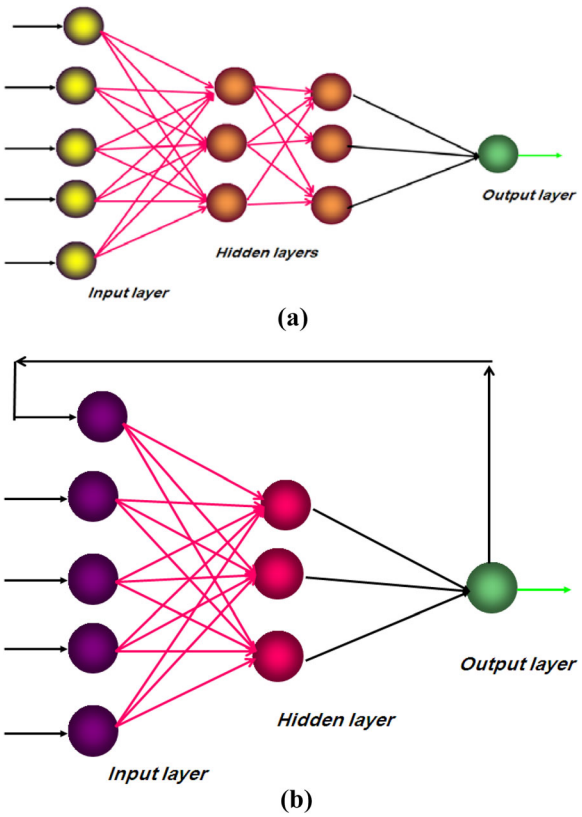


Figure 5. Architectures of neural networks: (**a**) feedforward network; (**b**) recurrent neural network.

where MSE is mean square error, RMSE (root mean square error) is the square root of MSE, p_i represents the measured value of pore pressure, y_i represents the pore pressure predicted by the model, and n is the number of sample data. The RMSEs were obtained for the training and testing data, and then the results were compared for all models.

RESULTS AND DISCUSSION

K-Means Results

Before predicting pore pressure from the well logs, the data were grouped based on their relations (minimum distances from point to point) using the *K*-means method. Two data groups were created, namely Group 1 (*K*-means 1) and Group 2 (*K*-means 2). Group 1 pertained to 1544–2570.154 m depth, and Group 2 to 4071–4410.754 m depth. The correctness of this classification was depicted by the high Silhouette value (Fig. 8). Most Silhouette values in this work ranged from 0.5 to 0.9 with average of 0.75. The Silhouette value of 0.75 indicates the least distance within points in a cluster, which means good clustering. Therefore, that value showed that the clustering of data was good. The negative Silhouette values (at depths between 2570.154 and 2870.154) showed the region where the data had poor correlation; thus, the data in that region were discarded.

After having the two groups of data, the illogical data were removed. These are data that were not related to other close data, and so they were not correct. The incorrectness of these data was caused by the failure of equipment to record due to poor conditions of the formation or equipment. After removing the illogical data, Group 1 remained with data for 1544–2405.754 m depth, and Group 2 remained with data for 4071.854–4410.754 m depth. Then, all well logs data from the selected depth were used in predicting pore pressure in Group 1 and Group 2.

Prediction of Pore Pressure by Polynomial Classifiers (POL)

Before applying the POL, the data were divided into training and testing data. Thus, in Group 1, 70% of the data for 1544–2154.354 m depth were used as the training data. The remaining 30%, for 2154.454–2405.754 m depth, were used as the testing data. Also, in Group 2, 70% of the data for 4071.854–4365.254 m depth were used as the training data; the remaining 30%, for 4365.254–4410.754 m depth, were used as the testing data.

Pore Pressure Prediction in Group 1 Data by Using POL

The POL method was applied to Group 1 data, whereby 852 data were used in training the model and 366 data were used in testing the model. The pore pressures predicted by the POL compared to those measured by the drill stem test (DST) are shown in Figure 9. The figure shows that at 2150–2250 m depth, the predicted pore pressures were almost equal to the measured pore pressures; but, at a depth of 2320 m, the deviation of pore pressure predicted by POL became significant. The last portion, at 2350–2400 m depth, also showed promising

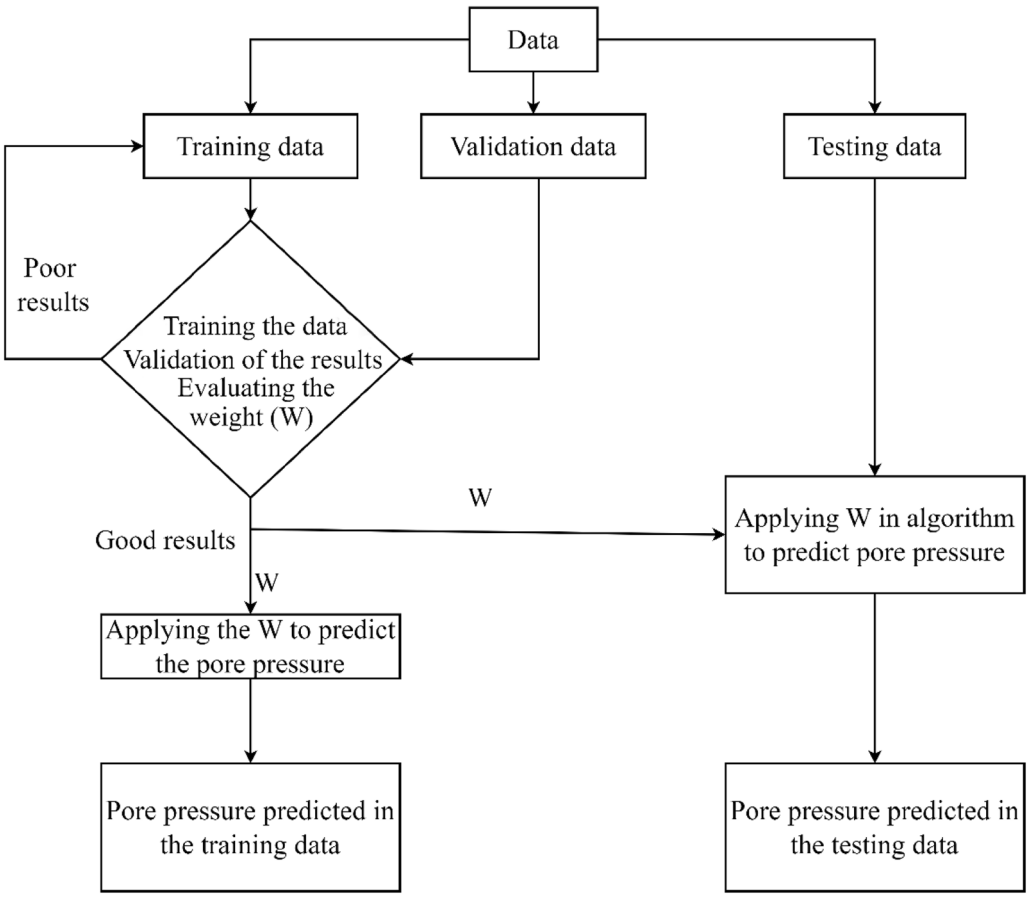


Figure 6. Steps followed when applying ANN.

results. However, generally, Figure 9 shows the closeness between measured and predicted pore pressures. Also, the RMSEs for both trained data and testing data showed good pore pressure prediction using POL. For example, the RMSE of the trained model was 0.1038 MPa, and the RMSE for the testing data was 0.5873 MPa.

Pore Pressure Prediction in Group 2 Data by Using POL

The POL method was applied to Group 2 data, where 1065 data were used for training the model and 456 data were used for testing the model and predicting pore pressure. The differences between pore pressures measured from DST and pore pressures predicted by the POL are shown in Figure 10. Variations in pore pressure increase with depth; at 4365 m, the variation was small but, as the depth increased, the variation in pore pressure also increased. The RMSE for the training data of Group 2 was smaller than that of the training data in Group 1. However, the RMSE of the Group 2 testing results was higher than the one obtained in Group 1. The trained model had an RMSE of 0.0736 MPa, and the testing and prediction results had an RMSE of 0.7981 MPa. These results show that the model was trained well, but the prediction deviated slightly from the measured values. Thus, the RMSE shown in Figure 10 indicates higher deviation between predicted and measured pore pressures compared to that shown in Figure 9.

Prediction of Pore Pressure by Artificial Neural Network (ANN)

The ANN was applied to the Group 1 and Group 2 data similarly as the POL. As shown in

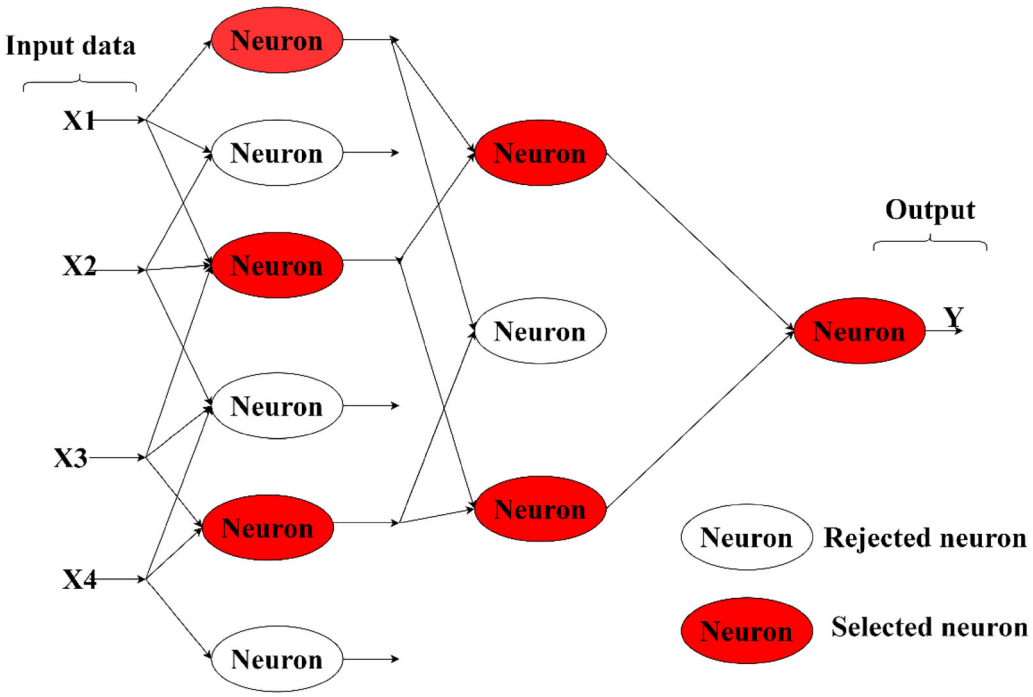


Figure 7. Selection of neurons that gives accurate results in the GMDH technique.

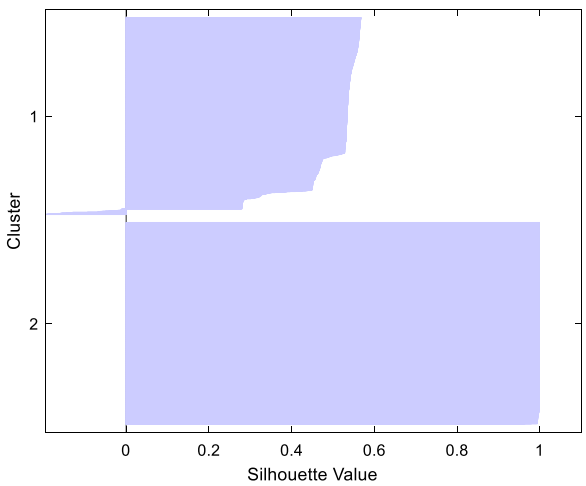


Figure 8. Silhouette values for the clusters in the pore pressure data from the Nanye 1 well.

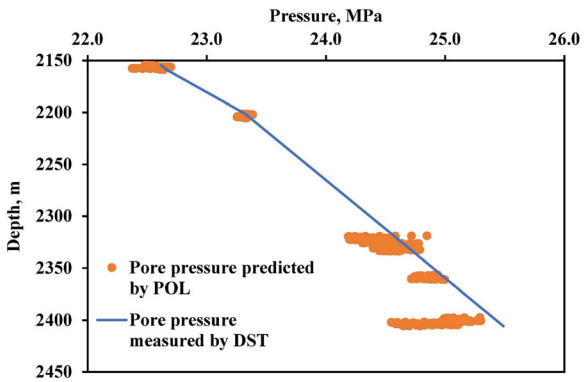


Figure 9. Pore pressures predicted from the well logs data by using a polynomial classifier (POL) compared to the pore pressures measured by DST (from Group 1 data).

Figures 11, 10 neurons in the hidden layer and 1 neuron in the output layer were applied in the ANN method to predict pore pressure. The data were divided into training, validation and testing data, in the proportions of 40%, 30%, and 30%, respectively.

Pore Pressure Prediction in Group 1 Data by Using ANN

The ANN obtained results after 247 epochs, which ran for 2 min and 25 s. Other parameters, including the performance and gradient parameters used in the ANN method are shown in Table 1. The pore pressure predicted by ANN compared to the ones measured by DST are shown in Figure 12. The results indicate that pore pressures predicted by

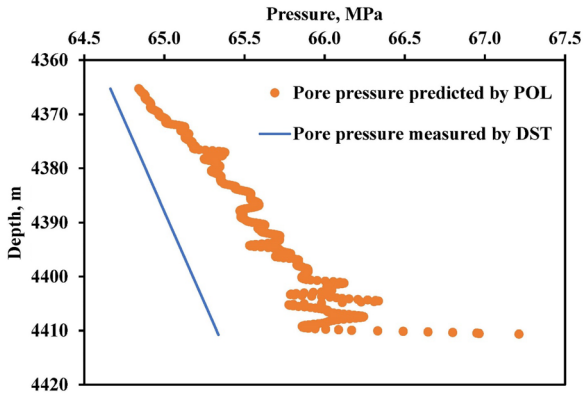


Figure 10. Pore pressures predicted from the well logs data using a polynomial classifier compared to the pore pressures measured by DST (from Group 2 data).

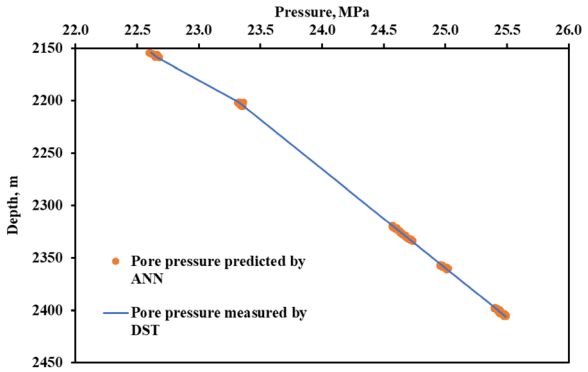


Figure 12. Pore pressures predicted from the well logs data using ANN with 10 neurons compared to pore pressures measured by DST (from Group 1 data).

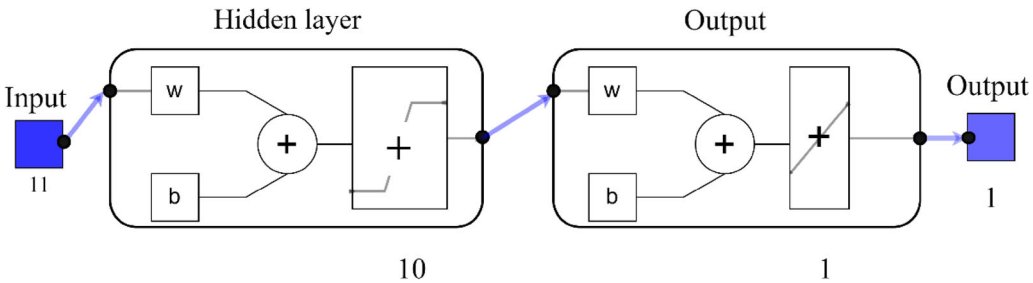


Figure 11. Number of neurons applied in the hidden and output layers during the ANN training process.

Table 1. Parameters used during the ANN prediction of pore pressure in Group 1 data

Unit	Initial value	Stopped value	Target value
Epoch	0	247	1000
Elapsed time	–	00:02:25	–
Performance	6.75	0.000791	0
Gradient	23.1	0.0492	0.0000001
Mu	0.001	1×10^{-5}	1×10^{10}
Validation checks	0	6	6

ANN were close to the those measured by DST. The ANN method predicted pore pressures almost equal to the measured pore pressures. The RMSE for the trained data was 0.0830 MPa, whereas the RMSE for the testing data was 0.0883 MPa. The RMSEs for both trained and testing data showed good pore pressure prediction. Compared to POL, the ANN had RMSEs, which imply better prediction of pore pressure.

Pore Pressure Prediction in Group 2 Data by Using ANN

The ANN used 137 epochs in 1 min and 1 s to obtain the results. The performance of the simulation was 0.0078, and the gradient was 0.0495 (Table 2). The pore pressures predicted by ANN in Group 2 are shown in Figure. 13. At depths of 4365–4375 m, there was good closeness between the predicted and measured pore pressures. At depths of 4380 m, the deviation of the predicted pore pressure from the measured pore pressure increased slightly. Then, the deviations decreased and increased again to a depth of 4410 m.

Generally, there were small differences between pore pressures measured by DST and pore pressures predicted by the ANN. In detail, there was a zigzag curve of increasing and decreasing differences between predicted and measured pore pressures. The trained data had an RMSE of 0.0328 MPa, and the testing and prediction results had an RMSE of 0.0322 MPa. These results show that the model was trained well, and the prediction

Table 2. Parameters assigned and used in the ANN training process for the Group 2 data

Unit	Initial value	Stopped value	Target value
Epoch	0	137	1000
Elapsed time	–	00:01:01	–
Performance	4.18	0.078	0
Gradient	15.5	0.0495	0.0000001
Mu	0.001	1×10^{-6}	1×10^{10}
Validation checks	0	6	6

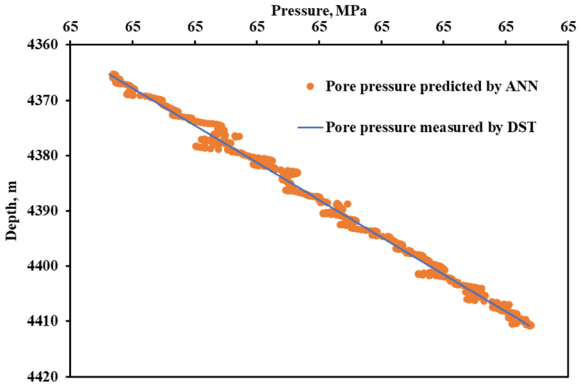
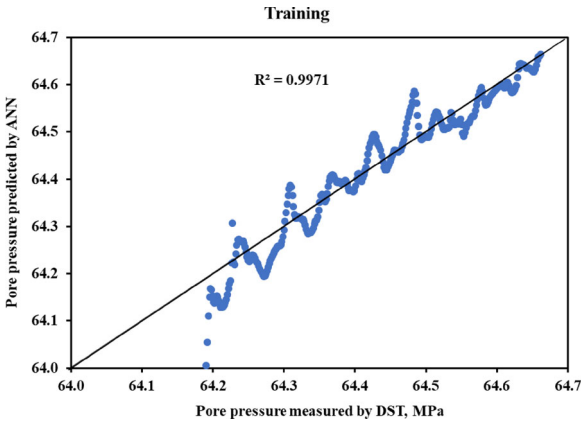


Figure 13. Pore pressures predicted from the well logs data using an ANN compared to pore pressures measured by DST (from Group 2 data).

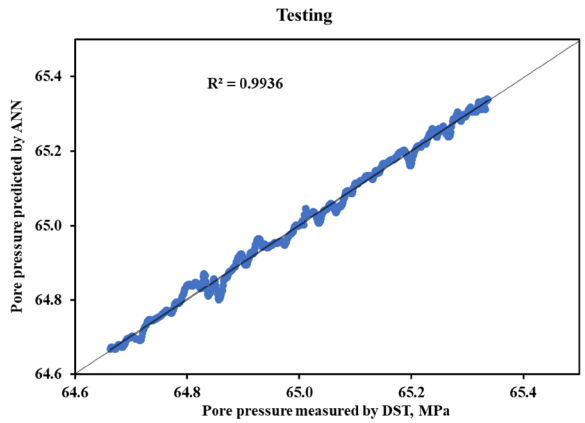
of pore pressure was good. The predictions of ANN were also described by a curve of predicted results (output) against measured pore pressure values (target), as shown in Figure 14. The figure shows that the coefficient of determination (R^2) for the training data was 0.9971 and for the testing results was 0.9936. These R^2 values imply good closeness between the measured (target) and predicted (output) pore pressures.

Prediction of Pore Pressure by Group Method of Data Handling (GMDH)

The GMDH technique was applied to predict pore pressure for Group 1 and Group 2 data. The initial parameters used in this network were 50 layers, 20 neurons, and 0.7 for classifying the training and testing data. There were 13 types of well logs used in this work, which led to a set of 72 pairs of two variables. These variables required 72 objective functions to be used in this work. With the aid of MATLAB (R2021a), all these functions were solved to predict pore pressure from the well logs.



(a) Training data



(b) Testing data

Figure 14. Comparison of pore pressures predicted by ANN (output) and measured pore pressure (target) at depths of 4360–4410.754 m (Group 2 testing data).

Pore Pressure Prediction in Group 1 Data by Using GMDH

The GMDH technique quantified the influence of each well log dataset on pore pressure for Group 1 data. Then, four well logs, namely depth, AC, CNL and RT, were selected to predict pore pressure. After choosing the appropriate well logs to predict the pore pressure, the GMDH selected the appropriate neurons to predict pore pressure. Then, the data were trained, and the weights were obtained through trained data. Finally, the GMDH generated several equations using the evaluated weights in various layers, which were used to provide the final results. Thus, the technique's final results comprised equations for predicting pore pressure and the output. The equations obtained depended on the selected neurons in each layer, and these equations were used to predict pore pressure in the training

and testing datasets. The final results were obtained in the last layer; thus, the equation in the last layer (Eq. 25) was the one that provided the output.

The following equations were generated by the GMDH technique to predict pore pressure:

$$Y_1 = -0.0169181 - 5.5121Z_1 + 6.51256Z_2 + 7.7816Z_1^2 + 7.49561Z_2^2 - 15.2772Z_1Z_2 \quad (25)$$

$$Z_4 = 18.6434 - 0.0970024N_2 + 8.69416N_3 + 0.000132174N_2^2 + 0.853451N_3^2 - 0.0212144N_2N_3 \quad (26)$$

$$Z_1 = -0.0602115 + 2.43497N_1 - 1.42748N_2 - 1.54742N_1^2 - 1.45886N_2^2 + 3.00608N_1N_2 \quad (27)$$

$$N_3 = -0.0469896 + 1.55873 \times 10^{-5}X_{11} + 1.00196X_4 - 2.2362 \times 10^{-11}X_{11}^2 + 1.23 \times 10^{-10}X_4^2 - 6.83819 \times 10^{-7}X_{11}X_4 \quad (28)$$

$$N_2 = -2.91643 + 0.0136783X_1 + 4.92853 \times 10^{-6}X_1X_2 - 4.25118 \times 10^{-7}X_1^2 - 0.0264323X_2 + 2.55632 \times 10^{-5}X_2^2 \quad (29)$$

$$N_1 = -6.16672 + 0.0141647X_1 - 0.064584X_4 - 3.64524 \times 10^{-7}X_1^2 + 0.000554602X_4^2 + 1.3687 \times 10^{-5}X_1X_4$$

where X_1 is depth, X_2 is AC, X_{11} is RT, and X_4 is CNL

The pore pressures predicted by the GMDH in Group 1 data had an RMSE of 0.0724 MPa in the trained data and an RMSE of 0.0712 MPa in the testing data. The predicted pore pressures were close to the measured pore pressures (Fig. 15) as indicated by R^2 of 0.998, which indicates the good performance of the GMDH technique.

After obtaining the results, sensitivity analysis was done to evaluate the influence of the selected input data (depth, AC, CNL and RT) on predicting pore pressure. Equation 30 was used to evaluate the significant variable (SV), thus

$$SV = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \text{output}}{\Delta \text{input}} \times 100 \quad (30)$$

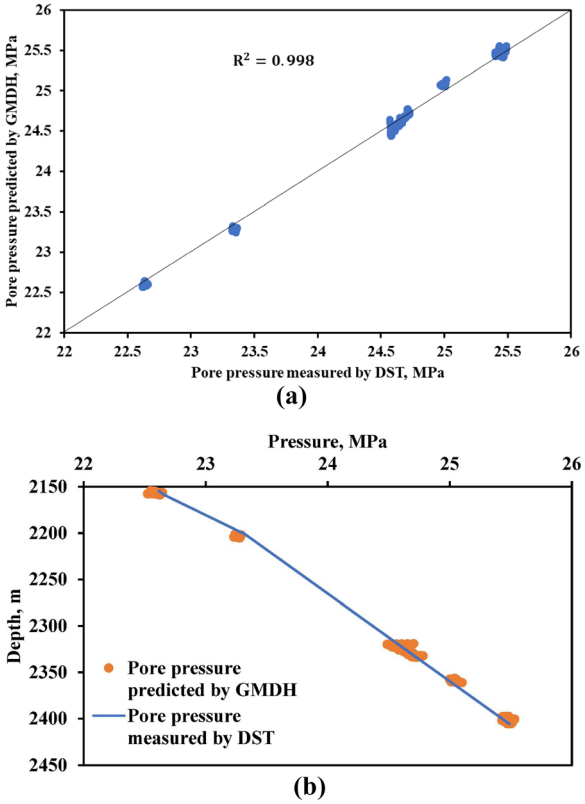


Figure 15. Relationship between pore pressure predicted by GMDH and measured by DST at depths of 2150–2405.754 m: (**a**) target pressure (measured) vs. predicted pore pressure (output); (**b**) variation of measured and predicted pressures at various depth.

where Δ output is the change of output data and Δ input is the change in input data. The value of SV shows the influence of each input dataset on predicting pore pressure; a higher SV value means stronger influence, and a lower value means weaker influence. The sensitivity analysis was done by varying the input data by percentages of 10, 20, 30, − 10, − 20 and − 30. The results show that depth had the strongest influence, followed by AC; meanwhile, CNL and RT had weaker impact on predicting pore pressure (Fig. 16).

Pore Pressure Prediction in Group 2 Data by Using GMDH

Using the GMDH technique, the influence of each well log dataset on pore pressure was evaluated; as a result, six well logs were selected to predict pore pressure in Group 2. The selected well logs were depth, SP, LLS, RS, DENS, and MSFL. Also,

the appropriate neurons were selected to predict pore pressure (Fig. 17). The GMDH technique attained the results after five layers and generated 15 functions by training the data. These functions were used to predict pore pressure in the testing data. The generated functions were represented by:

$$L_{\text{out}} = a + bL_{\text{in1}} + cL_{\text{in2}} + dL_{\text{in1}}^2 + eL_{\text{in2}}^2 + fL_{\text{in1}}L_{\text{in2}} \tag{31}$$

where L_{out} is the output in a layer, L_{in1} is the input 1, L_{in2} is the input 2, and a, b, c, d, e, and f are the coefficients. Each function can be expressed by inserting the coefficients shown in Table 3. The pore pressures predicted by GMDH were obtained using the equation of the last layer (Layer 5).

The results showed an RMSE of 0.0299 MPa in the training data and an RMSE of 0.0308 MPa in the testing data. The pore pressures predicted by GDMH fit well with the pore pressures measured by DST (Fig. 18), with R^2 of 0.9969. These results indicate a good prediction of pore pressure by using the GMDH technique.

The sensitivity analysis was conducted to evaluate the influence of the selected parameters in

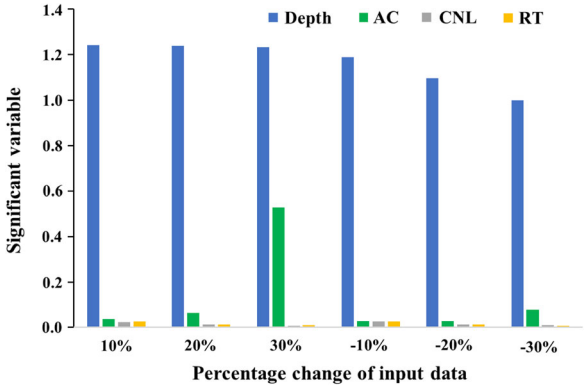


Figure 16. Sensitivity analysis of influence of depth, AC, CNL and RT on pore pressure.

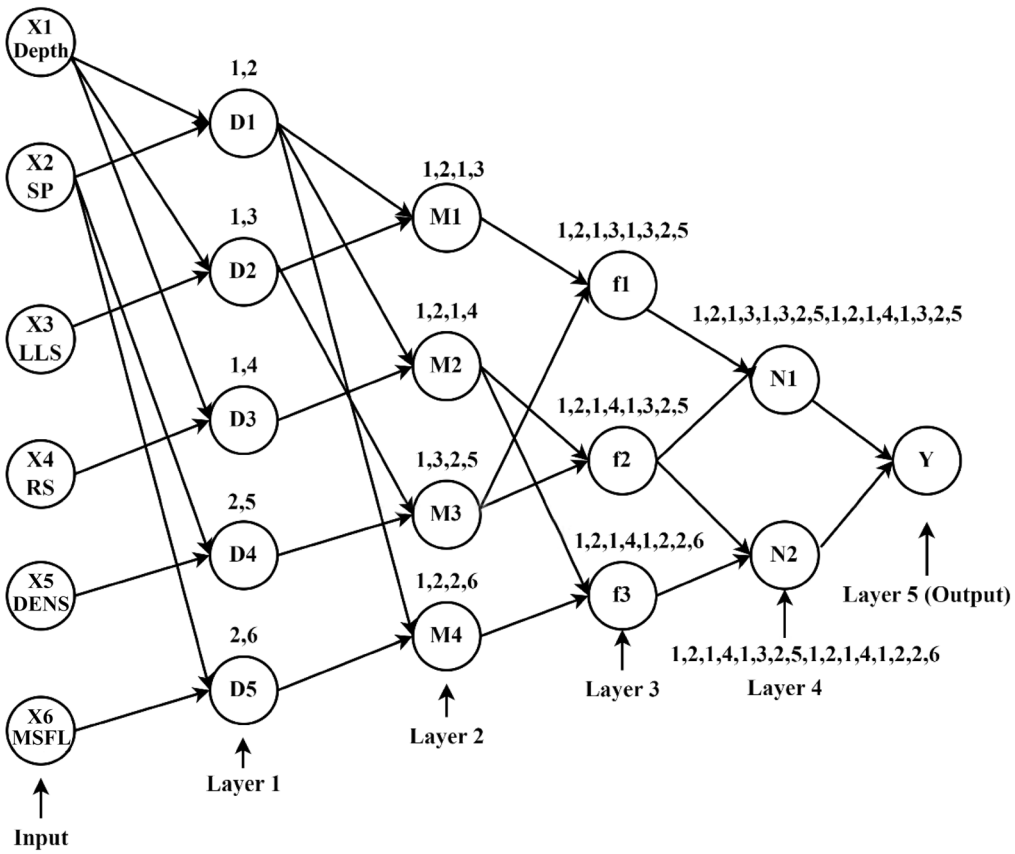


Figure 17. Selected neurons in the GMDH network connections structure for Group 2 data.

Table 3. Coefficients of the 15 functions generated by the GMDH technique to predict pore pressure in Group 2 data

Layers	Input		Output	Coefficients					
				a	b	c	d	e	f
Layer 1	X1	X2	D1	1786.971	− 0.807	− 1.34	9.44×10^{-5}	7.37×10^{4}	0.00030
	X1	X3	D2	871.589	− 0.391	0.0079	4.73×10^{-5}	-2.09×10^{8}	-1.858×10^{-6}
	X1	X4	D3	871.589	− 0.391	0.007947	4.73×10^{-5}	-2.09×10^{8}	-1.858×10^{-6}
	X2	X5	D4	61.814	0.441	0.142	-8.65×10^{2}	-2.19×10^{3}	0.0065
	X2	X6	D5	61.681	− 2900	0.185	1.52×10^{-5}	-2.45×10^{3}	9.3891×10^{-5}
Layer 2	D1	D2	M1	− 17.446	− 6.092	7.6352	0.957	0.843707	− 1.805
	D1	D3	M2	0.4524	11.945	− 10.963	0.325	0.497	− 0.823
	D2	D4	M3	− 615.493	67.038	− 46.836	0.1165	0.987796	− 1.254
	D1	D5	M4	50.71825	16.951	− 17.53	0.4658	0.727401	− 1.181
Layer 3	M1	M3	f1	0.223	31.954	− 30.964	− 0.279	0.214539	0.0645
	M2	M3	f2	1.393	31.683	− 30.73	− 0.3187	0.170704	0.1484
	M2	M4	f3	44.966	− 37.843	37.436	5.3448	4.756498	− 10.09
Layer 4	f1	f2	N1	15.963	− 839.21	839.72	− 898.395	− 911.428	1809.827
	f2	f3	N2	− 4.8179	− 26.032	27.182	− 3.3634	− 3.78054	7.143
Layer 5	N1	N2	Y	1.7212	27.366	− 26.42	3.3455	3.770905	− 7.116

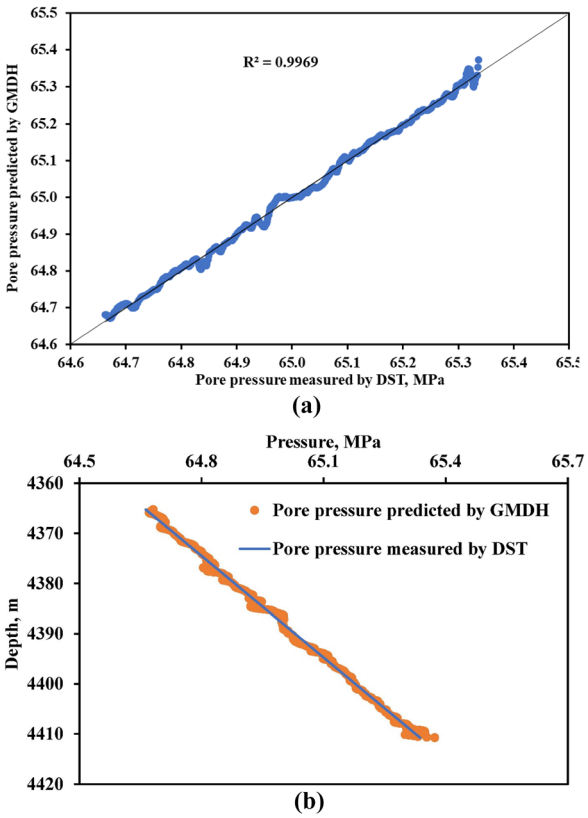


Figure 18. Relationships between pore pressure predicted by GMDH and measured by DST at depths of 4360–4410.754 m: **a** target pore pressure (measured by DST) and predicted pore pressure (output); **b** variations of measured and predicted pressures with depth.

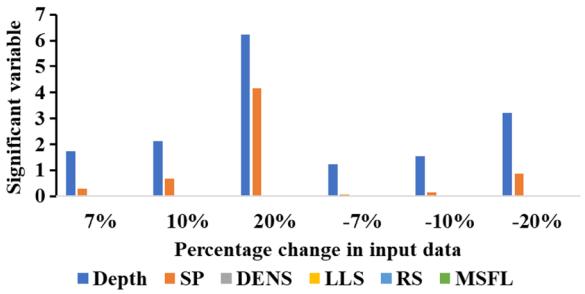


Figure 19. Influence of depth, SP, DENS, LLS, RS and MSFL on predicting pore pressure.

predicting pore pressures. Equation 30 was employed to evaluate the SV. The analysis was conducted by changing the input data by percentages of 7, 10, 20, − 7, − 10 and − 20. The results indicate that depth had the most significant influence, followed by SP logs. Other well logs had less influence in predicting pore pressure (Fig. 19).

DISCUSSION

According to Azadpour et al. (2015), the variation of porosity with pore pressure causes changes in petrophysical properties like compaction and fluid motion. These properties are well shown by the well logs data, and so based on this fact the well logs data can be used to predict pore pressure. Azadpour et al. (2015) showed the partial use of well logs data in

predicting pore pressure using mathematical models, including the Eaton, Bowers, and compressibility methods. All these models show good predictions of pore pressure from well logs.

All the three techniques used in approximating pore pressure in this work showed better results with small errors (Table 4). The GMDH technique obtained the best testing results (prediction) in Group 1 with RMSE of 0.0712 MPa, followed by ANN with RMSE of 0.0883 MPa. The polynomial classifier showed the worst results compared to the two techniques, whereby the RMSE was 0.5873 MPa, which was the highest compared to the other two methods. Also, in Group 2 data, the GMDH technique showed the best performance in predicting pore pressure, whereby the RMSE was 0.0308 MPa (the smallest among the three methods used), and the ANN technique was second with RMSE of 0.0322 MPa. The polynomial classifier performed the worst in predicting pore pressure compared to the two techniques (RMSE was 0.7981).

Also, the GMDH tested the influence of each well log dataset on pore pressure. In the first layer, all well logs data were arranged in a group of two data called a neuron. The output of each neuron was found using the two input data. Then, the results of each neuron were analyzed based on the regularity criteria measure. The neuron with a larger value of regularity criteria measure (large error between actual pore pressure measured by DST and neuron output) compared to the threshold specified value was discarded. Only the neurons with smaller regularity criteria measures were selected for the next stage. This mechanism shows the advantage of GMDH to quantity the influence of each well log data on the prediction of pore pressure, which is not done by the other ML techniques used in this work. Also, the sensitivity analysis showed that depth had the strongest influence on predicting pore pressures, followed by SP and AC. Other well logs, including DENS, RS, LLS, MSFL, CNL and RT, had a lower influence on predicting pore pressures.

The results are consistent with previous predictions using GMDH. The study of Mulashani et al. (2021) on the prediction of TOC from well logs using GMDH was the best compared to other methods like $\Delta logR$ technique and ANN. Mathew Nkurlu et al. (2020) showed the prediction of permeability by GMDH was best when compared to other methods, including back propagation neural network (BPNN) and radial basis function neural network (RBFNN). The GMDH showed higher performance (RMSE of 0.0308 MPa or 4.47 psi) when compared to other techniques like the DT technique (RMSE was 14.46 psi) (Zhang et al., 2022), MELM technique (RMSE was 11.551 psi) (Farsi et al., 2021) in predicting formation pore pressure from well logs. However, these techniques were applied in different fields. Apart from its performance, the GMDH has other advantages, including identifying the structure of the data or system, direct approximation of results, and it is fast (it took less than a minute in both Groups 1 and 2) (Abu-Kheil, 2009).

The validation of the results was based on the assumption that pore pressure measured by DST is the true formation pore pressure. Thus, the small errors (RMSEs) of the pore pressures predicted by POL, ANN, and GNDH against the measured by DST indicate the validity of the results (Table 4). In addition, the results showed good relationships between well logs and pore pressures, such that the well logs can be used to predict pore pressure. Also, the R^2 values in Figures 16a and 18a are greater than or equal to 0.9936, showing the closeness between the predicted and measured pore pressures, indicating the validity of using well logs to predict pore pressure. Therefore, all methods showed valid results, but GMDH was selected because it showed the best results among all the methods used in this work.

The relationships between the pore pressures measured by DST, predicted by GMDH, and well logs are shown in Figure 20. The figure indicates that both logs impact pore pressure changes; for instance, the rapid increase in pore pressure at depths starting from 4200 m is also related to changes in well logs (AC, CAL, CNL, DENS, and GR). Also, the low pressures in depths below 2000 m are reflected by the increase in AC, CAL, CNL, and RS well logs and the decrease in DEN and SP well logs. Moreover, the pressure predicted by the GMDH looked similar to the pressure measured by DST. This fact shows the good ability of GMDH to predict pore pressure. The success of the GMDH in predicting pore pressure can reduce the cost of measuring pore pressure by using expensive methods like DST.

CONCLUSIONS

Applying a ML technique to predict pore pressure shows outstanding achievement for POL, ANN, and GMDH. These techniques had small

Table 4. MSE and RMSE of training and testing results of the three different methods applied to the two groups of data

Methods	Group 1				Group 2			
	MSE $(MPa)^2$		RMSE (MPa)		MSE $(MPa)^2$		RMSE (MPa)	
	Train	Test	Train	Test	Train	Test	Train	Test
Polynomial classifier	0.0108	0.3450	0.1038	0.5873	0.0054	0.6370	0.0736	0.7981
ANN	0.0069	0.0078	0.0830	0.0883	0.00107	0.00103	0.0328	0.0322
GMDH	0.0052	0.0051	0.0724	0.0712	0.00089	0.00095	0.0299	0.0308

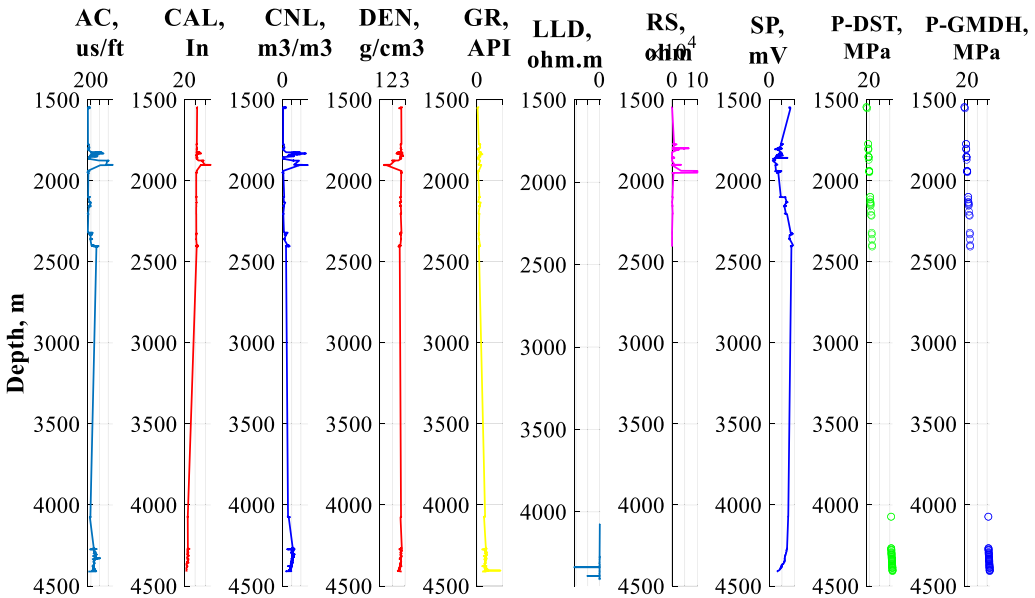


Figure 20. Relationships between measured pressure, predicted pressure, and well-log data.

RMSEs (the lowest was 0.0308 MPa) in predicting pore pressure. Also, the R^2 values of linear regressions in the fitness curve were up to 0.998. These results show good relationships between the well logs data and the pore pressure as well as the ML methods' ability to predict pore pressure. Thus, pore pressure prediction from the well logs data is valid. Depth showed the greatest influence on predicting pore pressure, followed by SP and AC well logs. However, DENS, RS, LLS, MSFL, CNL and RT also showed some though weaker influences.

Among all the techniques used, the GMDH shows more advantages compared to the other two techniques. These advantages include automatic selection of relevant input data and its ability to identify data structure and direct approximation. Also, the GMDH shows the best results compared to the other methods because it had the smallest RMSE of 0.0308 MPa in predicting pore pressure from both Group 1 and Group 2 data. Therefore, the GMDH is proposed as the best method for predicting pore pressure from well logs.

ACKNOWLEDGMENTS

The authors would like to admit the financial support of the National Natural Science Foundation of China (No. 42130803). Also, special thanks are given to the China Scholarship Council (CSC No.:2019GBJ002427) for funding the first author to study and conduct research at the China University of Geosciences at Wuhan

REFERENCES

Abbey, C. P., chukwudi Meludu, O., & Oniku, A. S. (2021). Investigation of abnormal pore pressure variations by the application of seismic inversion in Norne Field, Mid-Norwegian margin Norway. *Petroleum Research*.

Abu-Kheil, Y. M. Z. (2009). *System Identification using group method of data handling (GMDH).*.

Ahmed, A., Elkatatny, S., Ali, A., & Abdulraheem, A. (2019a). Comparative analysis of artificial intelligence techniques for formation pressure prediction while drilling. *Arabian Journal of Geosciences, 12*(18), 1–13.

Ahmed, A., Elkatatny, S., Ali, A., Mahmoud, M., & Abdulraheem, A. (2019b). New model for pore pressure prediction while drilling using artificial neural networks. *Arabian Journal for Science and Engineering, 44*(6), 6079–6088.

Al-Mohair, H. K., Saleh, J. M., & Suandi, S. A. (2015). Hybrid human skin detection using neural network and K-means clustering technique. *Applied Soft Computing, 33*, 337–347.

Asante-Okyere, S., Shen, C., Ziggah, Y. Y., Rulegeya, M. M., & Zhu, X. (2020). A novel hybrid technique of integrating gradient-boosted machine and clustering algorithms for lithology classification. *Natural Resources Research, 29*(4), 2257–2273.

Atashbari, V., & Tingay, M. (2012). *Pore pressure prediction in carbonate reservoirs.* Paper presented at the SPE Latin America and Caribbean petroleum engineering conference.

Azadpour, M., Manaman, N. S., Kadkhodaie-Ilkhchi, A., & Sedghipour, M.-R. (2015). Pore pressure prediction and modeling using well-logging data in one of the gas fields in south of Iran. *Journal of Petroleum Science and Engineering, 128*, 15–23.

Biot, M. A. (1941). General theory of three-dimensional consolidation. *Journal of applied physics, 12*(2), 155–164.

Bowers, G. L. (1995). Pore pressure estimation from velocity data: Accounting for overpressure mechanisms besides undercompaction. *SPE Drilling & Completion, 10*(02), 89–95.

Cao, C., Li, L., Liu, Y., Du, L., Li, Z., & He, J. (2020). Factors affecting shale gas chemistry and stable isotope and noble gas isotope composition and distribution: A case study of lower Silurian Longmaxi Shale Gas. *Sichuan Basin. Energies, 13*(22), 5981.

Contreras, O., Tutuncu, A., Aguilera, R., & Hareland, G. (2011). *A case study for pore pressure prediction in an abnormally sub-pressured western Canada sedimentary basin.* Paper presented at the 45th US Rock Mechanics/Geomechanics Symposium.

de Souza, J. A., Martínez, G. C., de Leon, M. F. C. P., Azadpour, M., & Atashbari, V. (2021). Pore pressure and wellbore instability. In *Applied Techniques to Integrated Oil and Gas Reservoir Characterization* (pp. 355–394): Elsevier.

Do Nascimento, M. Z., Martins, A. S., Neves, L. A., Ramos, R. P., Flores, E. L., & Carrijo, G. A. (2013). Classification of masses in mammographic image using wavelet domain features and polynomial classifier. *Expert Systems with Applications, 40*(15), 6213–6221.

Farsi, M., Mohamadian, N., Ghorbani, H., Wood, D. A., Davoodi, S., Moghadasi, J., & Ahmadi Alvar, M. (2021). Predicting formation pore-pressure from well-log data with hybrid machine-learning optimization algorithms. *Natural Resources Research, 30*(5), 3455–3481.

Hailong, X., Guoqi, W., Chengzao, J., Wei, Y., Tianwei, Z., Wuren, X., Li, C., & Beiwei, L. (2012). Tectonic evolution of the Leshan-Longnüsi paleo-uplift and its control on gas accumulation in the Sinian strata. *Petroleum Exploration and development, 39*(4), 436–446.

Hu, L., Deng, J., Zhu, H., Lin, H., Chen, Z., Deng, F., & Yan, C. (2013). A new pore pressure prediction method-back propagation artificial neural network. *Electronic Journal of Geotechnical Engineering, 18*, 4093–4107.

Hutomo, P., Rosid, M., & Haidar, M. (2019). *Pore pressure prediction using eaton and neural network method in carbonate field "X" based on seismic data.* Paper presented at the IOP conference series: Materials science and engineering.

Ivakhnenko, A., & Ivakhnenko, G. (1995). The review of problems solvable by algorithms of the group method of data handling (GMDH). *Pattern recognition and image analysis c/c of raspoznavaniye obrazov i analiz izobrazhenii, 5*, 527–535.

Jing, T., Zhang, J., Xu, S., Liu, Z., & Han, S. (2016). Critical geological characteristics and gas-bearing controlling factors in Longmaxi shales in southeastern Chongqing. *China. Energy Exploration & Exploitation, 34*(1), 42–60.

Keshavarzi, R., & Jahanbakhshi, R. (2013). Real-time prediction of pore pressure gradient through an artificial intelligence approach: A case study from one of middle east oil fields. *European journal of environmental and civil engineering, 17*(8), 675–686.

Korbicz, J., & Mrugalski, M. (2008). Confidence estimation of GMDH neural networks and its application in fault detection systems. *International Journal of Systems Science, 39*(8), 783–800.

Korsch, R., Huazhao, M., Zhaocai, S., & Gorter, J. (1991). The Sichuan basin, southwest China: A late proterozoic (Sinian) petroleum province. *Precambrian Research, 54*(1), 45–63.

Lin, X., Zeng, J., Wang, J., & Huang, M. (2020). Natural Gas reservoir characteristics and non-Darcy flow in low-permeability sandstone reservoir of Sulige gas field, Ordos Basin. *Energies, 13*(7), 1774.

Liu, G. P. (2001). *Nonlinear identification and control: a neural network approach.* Springer.

Liu, S. F., Ma, Y. S., & Wang, G. Z. (2014). *Formation Process and mechanism of the Sinian-Silurian natural Reservoirs in the Sichuan Basin*. Science Press. (in Chinese).

Liu, Y., Qiu, N., Yao, Q., & Zhu, C. (2017). The impact of temperature on overpressure unloading in the central Sichuan Basin, southwest China. *Journal of Petroleum Science and Engineering, 156*, 142–151.

Martins, R. G., Martins, A. S., Neves, L. A., Lima, L. V., Flores, E. L., & do Nascimento, M. Z. (2017). Exploring polynomial classifier to predict match results in football championships. *Expert Systems with Applications, 83*, 79–93.

Mathew Nkurlu, B., Shen, C., Asante-Okyere, S., Mulashani, A. K., Chungu, J., & Wang, L. (2020). Prediction of permeability using group method of data handling (GMDH) neural network from well log data. *Energies, 13*(3), 551.

Menad, N. A., Noureddine, Z., Hemmati-Sarapardeh, A., Shamshirband, S., Mosavi, A., & Chau, K.-W. (2019). Modeling temperature dependency of oil - water relative permeability in thermal enhanced oil recovery processes using group method of data handling and gene expression programming. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 724–743. https://doi.org/10.1080/19942060.2019.1639549.

Mesbah, M., Habibnia, S., Ahmadi, S., Saeedi Dehaghani, A. H., & Bayat, S. (2022). Developing a robust correlation for prediction of sweet and sour gas hydrate formation temperature. *Petroleum, 8*(2), 204–209.

Mulashani, A. K., Shen, C., Asante-Okyere, S., Kerttu, P. N., & Abelly, E. N. (2021). Group method of data handling (GMDH) neural network for estimating total organic carbon (TOC) and hydrocarbon potential distribution (S1, S2) using well logs. *Natural Resources Research, 30*(5), 3605–3622.

Mulashani, A. K., Shen, C., Nkurlu, B. M., Mkono, C. N., & Kawamala, M. (2022). Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data. *Energy, 239*, 121915.

Murphy, K. P. (2002). *Dynamic bayesian networks: Representation, inference and learning*. University of California.

Mutumba, G., Echegu, S., Adaramola, S., & M. (2021). Prospects and challenges of geothermal energy in Uganda. *Journal of Energy Research and Reviews*. https://doi.org/10.9734/jenrr/2021/v9i230230.

Nait Amar, M., Larestani, A., Lv, Q., Zhou, T., & Hemmati-Sarapardeh, A. (2022). Modeling of methane adsorption capacity in shale gas formations using white-box supervised machine learning techniques. *Journal of Petroleum Science and Engineering, 208*, 109226.

Najafzadeh, M., Barani, G.-A., & Hessami-Kermani, M.-R. (2015). Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds. *Ocean Engineering, 104*, 387–396.

Nie, H., Jin, Z., Ma, X., Liu, Z., Lin, T., & Yang, Z. (2017). Dispositional characteristics of Ordovician Wufeng formation and Silurian Longmaxi formation in Sichuan Basin and its adjacent areas. *Petroleum Research, 2*(3), 233–246.

Nyakilla, E. E., Silingi, S. N., Shen, C., Jun, G., Mulashani, A. K., & Chibura, P. E. (2022). Evaluation of source rock potentiality and prediction of total organic carbon using well log data and integrated methods of multivariate analysis, machine learning, and geochemical analysis. *Natural Resources Research*, 1–23.

Qi, X., Hu, Q., Yi, X., & Zhang, S. (2015). Shale gas exploration prospect of Lower Cambrian Wangyinpu formation in Xiuwu Basin. *China Mining Magazine, 24*(10), 102–107.

Rahim, N. A., Taib, M., Adom, A., & Mashor, M. (2006). *The NARMAX model for a dc motor using mlp neural network.* Paper presented at the Proceeding of the First International Conference 0n MAN-MACHINE SYSTEMS (ICoMMS).

Shaghaghi, S., Bonakdari, H., Gholami, A., Ebtehaj, I., & Zeinolabedini, M. (2017). Comparative analysis of GMDH neural network based on genetic algorithm and particle swarm optimization in stable channel design. *Applied Mathematics and Computation, 313*, 271–286.

Srinivasan, D. (2008). Energy demand prediction using GMDH networks. *Neurocomputing, 72*(1–3), 625–629.

Swarbrick, R. E. (2001). *Challenges of porosity based pore pressure prediction.* Paper presented at the 63rd EAGE conference & exhibition.

Swarbrick, R. (2012). Review of pore-pressure prediction challenges in high-temperature areas. *The Leading Edge, 31*(11), 1288–1294.

Veeken, P. P. (2006). *Seismic stratigraphy, basin analysis and reservoir characterisation*: Elsevier.

Wang, G., Ju, Y., & Han, K. (2015). Early Paleozoic shale properties and gas potential evaluation in Xiuwu Basin, western Lower Yangtze Platform. *Journal of Natural Gas Science and Engineering, 22*, 489–497.

Wo, Y.-J., Zhou, Y., & Xiao, K.-H. (2007). The burial history and models for hydrocarbon generation and evolution in the marine strata in southern China. *Sedimentary Geology and Tethyan Geology, 27*(3), 100.

Xu, Q., Qiu, N., Liu, W., Shen, A., Wang, X., & Zhang, G. (2018). Characteristics of the temperature–pressure field evolution of Middle Permian system in the northwest of Sichuan Basin. *Energy Exploration & Exploitation, 36*(4), 705–726.

Yi-Feng, L., Lun-Ju, Z., Nan-Sheng, Q., Jing-Kun, J., & Qing, C. (2015). The effect of temperature on the overpressure distribution and formation in the central paleo-uplift of the Sichuan Basin. *Chinese Journal of Geophysics, 58*(4), 340–351.

Youcefi, M. R., Hadjadj, A., & Boukredera, F. S. (2022). New model for standpipe pressure prediction while drilling using group method of data handling. *Petroleum, 8*(2), 210–218.

Yu, H., Chen, G., & Gu, H. (2020). A machine learning methodology for multivariate pore-pressure prediction. *Computers & Geosciences, 143*, 104548.

Zhang, G., Davoodi, S., Band, S. S., Ghorbani, H., Mosavi, A., & Moslehpour, M. (2022). A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques. *Energy Reports, 8*, 2233–2247.

Zhili, L. (1998). New recognition of basement in Sichuan Basin. *Journal of Chengdu University of Technology, 25*(2), 191–200.

Zimmerman, R. W. (1990). Compressibility of sandstones.

Zou, C. (2017). *Unconventional petroleum geology*. Elsevier.