



# Estimation of porosity and volume of shale using artificial intelligence, case study of Kashafrud Gas Reservoir, NE Iran

Pooya Naghizadeh Ardebili<sup>1</sup> · Golnaz Jozanikohan<sup>1</sup> · Ali Moradzadeh<sup>1</sup>

Received: 12 April 2023 / Accepted: 18 November 2023 / Published online: 16 December 2023  
© The Author(s) 2023

## Abstract

The key problem in oil exploration and engineering is the lack of accurate and reliable data about the reservoir parameters of a field. Having a precise assessment of petrophysical properties can provide the ability to make decisions with a high degree of confidence about planning for production, exploitation, and further field development scenario. In this research, an artificial intelligence (AI)-based approach was developed to improve the estimation of reservoir parameters including porosity and volume of shale, which has a significant role in different stages of hydrocarbon exploration, in the Kashafrud Gas Reservoir in the northeast of Iran. For this purpose, we measured the petrophysical properties of 27 samples of the Kashafrud Formation. To increase the amount of data for employing a multilayer perceptron (MLP) artificial neural network (ANN), a geostatistical algorithm was used to increase the amount of laboratory measured data of porosity and volume of shale to 686 and 702, respectively. In addition, 2263 well-logging data from the same well were provided. The optimal MLP network with the topology of 6-7-1, and 6-8-1 was selected to estimate the porosity and shale volume with mean squared error (MSE) of  $2.78731E^{-4}$ , and  $1.28701E^{-9}$ , respectively. The training process was performed using two different sets of input data. In the first approach, all available well-logging data were used as input, ending up in high MSE. In the second approach, some selected well logs were used based on the results of sensitivity analysis which clearly improved the estimations. The ability of MLP networks made great improvements in the estimation of the both parameters up to 99.9%. The presence of valuable core data in this study significantly improved the process of comparison and conclusion. The final results prove that AI is a trusted method, also the potential of the ANN method for the reservoir characterization and evaluation associated problems should be taken into consideration. Due to the unavailability of core data along the whole wells, the application of intelligent methods, such as machine learning (ML) can be used to estimate the parameters in other oil or gas fields and wells.

**Keywords** Formation evaluation · Reservoir parameters · Clay minerals · Neural network · Data fusion

## List of symbols

CALI	Caliper, IN	$GR_{max}$	Average ray response in dirty (clay rich) zone, API or wt%
CGR	Corrected gamma ray, API	$GR_{min}$	Average ray response in clean (clay free) zone, API or wt%
DT	Sonic, US/F	$I_{GR}$	Natural gamma ray index
GR	Natural gamma ray, API	$I_{SGR}$	Standardized natural gamma ray index
$GR_{log}$	Total ray reading in the zone of interest, API or wt%	K	Permeability, mD
		LLS	Shallow resistance radius, Ohmm
		MSFL	Micro spherically, Ohmm
		NPHI	Neutron, V/V
		PEF	Photoelectric, B/E
		RHOB	Density, $g/cm^3$
		SCR	SUM gamma ray, API
		SP	Spontaneous, mV
		$S_{we}$	Effective water saturation
		$V_{sh}$	Shale volume, %
		$\rho_{b_{sh}}$	Bulk density in dirty (clay rich) zone, $g/cm^3$

✉ Golnaz Jozanikohan  
gjkohan@ut.ac.ir

Pooya Naghizadeh Ardebili  
pooya.naghizadeh@ut.ac.ir

Ali Moradzadeh  
a\_moradzadeh@ut.ac.ir

<sup>1</sup> School of Mining Engineering, College of Engineering, University of Tehran, Tehran, Iran

$\rho_b$	Density log data (RHOB), g/cm <sup>3</sup>
$\emptyset_{ND}$	Neutron–density log porosity, V/V
$\emptyset_N$	Neutron log porosity, V/V
$\emptyset$	Porosity, V/V
AI	Artificial intelligence
ANN	Artificial neural network
FTIR	Fourier transform infrared spectroscopy
KNN	K-nearest neighbor
ML	Machine learning
MLP	Multilayer perceptron
MSE	Mean squared error
$R$	Correlation coefficient
RF	Random forest
SVM	Support vector machine
XGBoost	Extreme gradient boost
XRD	X-ray diffraction

## Introduction

The reservoir properties including porosity and volume of shale play important roles in the reservoir production (Cheng and Pan 2020). Engineers design the best plans for the further reservoir development stages, also optimize the hydrocarbon recovery with the help of precise knowledge of these petrophysical parameters (Solanki et al. 2021). The porosity ( $\emptyset$ ) is greatly affected by the amount of clay minerals inside the reservoir known as volume of shale ( $V_{sh}$ ) (Iqbal and Rezaee 2020). One of the most crucial stages in the characterization of reservoirs is the estimation of the shale volume (Balaky et al. 2023). It can be very challenging to determine the amount of shale in many areas accurately (Hussain et al. 2023). Its overestimation causes the effective water saturation ( $S_{we}$ ) to be estimated very low, and this eventually leads to the wrong assumption of productivity. On the other hand, in the case of underestimation of shale volume, the water saturation will be estimated more than its actual value and this makes a productive zone to be overlooked. Moreover, the underestimation or overestimation of this parameter can cause some miscalculation in the estimation of effective porosity which is used for the determination of net pay (Iltaf et al. 2023).

ANNs can efficiently solve nonlinear problems. The ANN combines connected units including the artificial nodes (neurons), input, output, and processing layers. Each node is able to receive or transmit a pulse from or to the other nodes. A weight is assigned to each neuron and updated during the learning process. Typically, there are one or several hidden (processing) layers of nodes in an ANN model. Each of these layers has input and output values, gives and receives data to or from the next and previous layers, respectively. Eventually, data are weighted and mixed together to make

a new input for the upcoming layer (Gong et al. 2019). The ANN is considered as algorithms with two crucial functions, i.e., classification and regression. The outputs or responses generated by regression are normally continuous values, and the regression application in the oil industry is to estimate the porosity ( $\emptyset$ ), permeability ( $k$ ), volume of shale ( $V_{sh}$ ), and water saturation ( $S_w$ ) (Gong et al. 2019). The ANNs have played a significant role in the accurate estimation of reservoir parameters including porosity, permeability, water saturation (Okon et al. 2021), and volume of shale (Taheri et al. 2021; Hong and Tien 2022). The implementation of the intelligence models enables the reservoir engineers to tackle the challenging and time-consuming tasks more successfully. The ANNs help to fuse different data and acquire complete and accurate information (Saikia et al. 2020).

The development of shale formations has had a transformative effect, especially in the USA, leading to notable improvements in the industry (Alessa et al. 2022). Over the last decade, shale reservoirs have been the major focus of extensive discussion and research on hydrocarbon exploration and exploitation globally. The development of shale formations has been a turning point, particularly in the USA, resulting in significant improvements. Simultaneously, ML and artificial intelligence (AI) have been instrumental in driving rapid development across all industries by automating routine operations (Syed et al. 2022). Shale gas reservoirs have been explored at various depths, ranging from as shallow as 1000 ft to as deep as 12,000 ft, with a variable range of total organic content (TOC) from 1 to 12%. The quality of the shale formation is characterized based on a variety of factors, including petrophysical properties like TOC, thermal maturity, saturation, and geo-mechanical properties such as the percentage of quartz or carbonate in mineralogy, differential stress, and friability, (Sondergeld et al. 2010). Since the early 1980s, petroleum engineers have been using computer-aided petrophysical and geo-mechanical studies such as log analysis, interpretation, and integration (Doveton 1986). However, ML-based and AI geo-mechanical and petrophysical analyses have become more prominent over the past decade, resulting in faster and more successful development than ever before in the history of the oil industry. As an unconventional reservoir, shale formations' geo-mechanical properties are affected by diagenetic changes resulting from the depositional environment, temperature, and pressure. These changes cause mineralogical alterations, leading to changes in rock composition that directly impact sediment compaction and lithification. This makes it challenging to predict the geo-mechanical properties of shale (Syed et al. 2022).

In a study focused on the gas permeability of shale, Sakhaee-Pour & Bryant (2012) demonstrated that the narrowness of pore throats in shale is predominantly below 10nm and is situated within organic material where CH<sub>4</sub> is

absorbed. Consequently, the gas permeability of these rock formations is notably influenced by the gas that is absorbed, along with the movement of gas sliding along the pore walls. This effect is particularly pronounced at higher pressures (such as the initial pressures) found in typical shale gas reservoirs. In such conditions, the impact of the absorbed gas layer takes precedence over the influence of gas slipping through the pore spaces. Besides, it was projected that the permeability of the reservoir matrix can experience a substantial increase over the well operation, potentially growing by a factor of 4.5 as production continues and pressure decreases. According to Sakhaee-Pour and Steven (2015), non-random spatial distributions of throat sizes in acyclic void models offer more accurate portrayals of the void space within samples. These models are particularly suited for cases where drainage experiments demonstrate a capillary pressure versus saturation trend that deviates from the plateau-like pattern. Furthermore, they made successful permeability predictions that aligned well with laboratory measurements. The models they developed may find utility in other porous media where drainage data do not display a plateau-like variation.

The examination carried out by Sakhaee-Pour and Li (2016), with potential far-reaching effects on comprehending hydrocarbon movement in shale formations, scrutinized drainage experiments conducted on core samples to elucidate the interconnected pathway topology within the pore space on a core-scale level. Their investigation across various shale varieties revealed that the path traversed within the pore space by the nonwetting phase, measured as the length of the pore space, adheres to a fractal pattern. This is in contrast to the pore volume, which does not inherently exhibit fractal characteristics. While the assessment of matrix permeability through mercury injection capillary pressure measurements is a customary procedure in the petrophysical analysis of rock formations, it remains unachievable for shale formations due to the absence of a practical and reliable model. In 2018, Tran et al. (2018) introduce a straightforward correlation, rooted in the acyclic pore model, to approximate shale permeability. This uncomplicated relation was subjected to testing using seven samples drawn from three distinct formations.

Tran and Sakhaee-Pour (2018a, b) asserted that their research holds significant implications for the characterization of reservoirs using conventional petrophysical measurements. The results of numerical simulations revealed that the gas flow's effective pore-throat size is influenced by pore pressure. Furthermore, the measured permeability in the presence of liquid surpassed the nominal permeability, commonly known as the Hagen–Poiseuille model, without accounting for slippage effects. Tran and Sakhaee-Pour (2018a, b) utilized the acyclic pore model to incorporate the effective interconnections among shale samples on the core

scale. Their research, focused on exploring the core-scale critical properties ( $T_c, P_c$ ) of shale gas, holds significant potential for advancing a practical reservoir model tailored to shale formations. The findings indicated substantial alterations in displacement-critical properties, while modifications were unnecessary for storage-critical properties.

Yu et al. (2018) established a study about pore size of shale based on acyclic pore model. Their investigation into diverse shale types revealed that the average size of pore bodies typically exceeds 20nm. As a result, there is no necessity to consider pore proximity or confinement. In contrast, the pore-throat size distributions across different shales generally lie below 20nm, necessitating adjustments to a transport property that pertains to the formation's resistance against fluid flow. Another study conducted by Alessa et al. (2021), they investigated the comprehensive characterization of pore sizes within Midra shale. The research established that pore-throat and pore-body sizes exhibit both narrow and wide distributions, with average measurements approximately 22nm and 18nm, respectively. As a result, modifications are needed for transport properties influenced by pore-throat sizes in order to accurately represent subsurface conditions. Notably, properties like density, which are tied to the volume of pores in the matrix, can be reasonably estimated based on gas composition within broader channels. These findings hold relevance for advancing unconventional gas development, which is regarded as one of the cleaner fossil fuel alternatives.

In 2022, Alipour et al. (2022) introduced an empirical correlation designed to account for the nonplateau-like pattern and the estimated capillary pressure observed in shale formations. By a dataset of mercury capillary pressure measurements from 30 samples extracted from various US shale formations, their proposed model holds potential for analyzing two-phase displacement phenomena within shale environments. Alessa et al. (2022) introduced a simple formula to precisely ascertain the entry pressure. This relationship, offering a novel method of determination, was employed on real measurements from seven shale samples. Enhancements to its effectiveness were achieved through the integration of k-nearest neighbors (KNN), locally selective combination in parallel outlier ensembles (LSCP), and Savitzky–Golay (SG) filters. The optimal outcome emerged from the sequential amalgamation of the basic formula with unsupervised machine learning and noise-filtering techniques.

The conventional well-logging data which are used to estimate the volume of shale, namely, are gamma ray (GR) and its spectral components, SP log, density (RHOB) log, resistivity logs (LLD, LLS, ILD, MSFL), neutron (NPHI), and sonic (DT) log. Moreover, a combination of gamma ray-density, neutron-density, and sonic-density logs can be used in formulas to estimate the  $V_{sh}$  (Ehsan et al. 2019; Tali and Farman, 2021; Mohavvel and Jozanikohan 2022). The

decrease in porosity due to the presence of clay minerals will lead in a poor reservoir quality (Zhou et al. 2022). The presence of shale has an effect on the petrophysical properties and logging tool responses, thereby it causes a significant reduction in the effectiveness of the reservoir porosity (Radwan et al. 2020; El-Gendy, 2022; Ismail et al., 2023; Saleh et al. 2023). Using the factor index, a linear relationship between a special factor and shale content can be obtained with the natural gamma ray index (Szabó 2011). To have the best prediction of the hydrocarbon accumulations, one needs to know the reservoir quality distribution factors such as ( $\emptyset$ ) and  $V_{sh}$  (Mohammed 2020). Gamal and Elkatatny (2021) implemented a new approach developed by the machine learning techniques (ANN) to predict the porosity of the reservoir rock. Their approach overcomes all of the conventional problems in the domain of porosity estimation using empirical correlations, measurements of the core samples, and logging tools.

Taheri et al. (2021) conducted a study by seismic data from the Hendijan oil field to establish a correlation between seismic properties and shale volume values. The researchers employed three distinct methods, namely, the sparse spike inversion, model-based inversion, and band-limited inversion methods to select the seismic line between the wells. The results indicated that the model-based method yielded the most favorable outcomes. Besides, they utilized ANNs in conjunction with seismic properties to estimate the shale volume. In another research, Ali (2021) employed traditional petrophysical techniques, such as linear gamma ray, nonlinear gamma ray, and spontaneous potential with the aim of creating a dataset for training ML algorithms, including random forest (RF), extreme gradient boost (XGBoost), and k-nearest neighbor (KNN). Ultimately, the nonlinear gamma ray method was identified as the most effective among the classical approaches, while the XGBoost algorithm demonstrated superior performance, achieving a mean squared error (MSE) of 0.078 (Ali 2021).

The findings obtained from the study of Jozanikohan and Abarghoeei (2022) offer valuable advantages for geoscientists in the upstream petroleum sector. They proved that by the conducted method, samples can be assessed before resorting to intricate and time-consuming chemical and mineralogical analyses, as the Fourier transform infrared spectroscopy (FTIR) method efficiently accomplishes both tasks with greater ease and reduced expenses. This technique proves especially beneficial for evaluating clastic reservoirs, shale oil, and shale gas targets, enabling a rapid evaluation of their potential. The study demonstrates the practicality of the approach using a set of Shurijeh core samples as an illustrative example. In recent years, there has been a noticeable trend in utilizing machine learning (ML) algorithms for shale volume estimation, marking a novel area of interest in the petrophysical evaluation stage. This development is evident from studies conducted in the past decade, such as

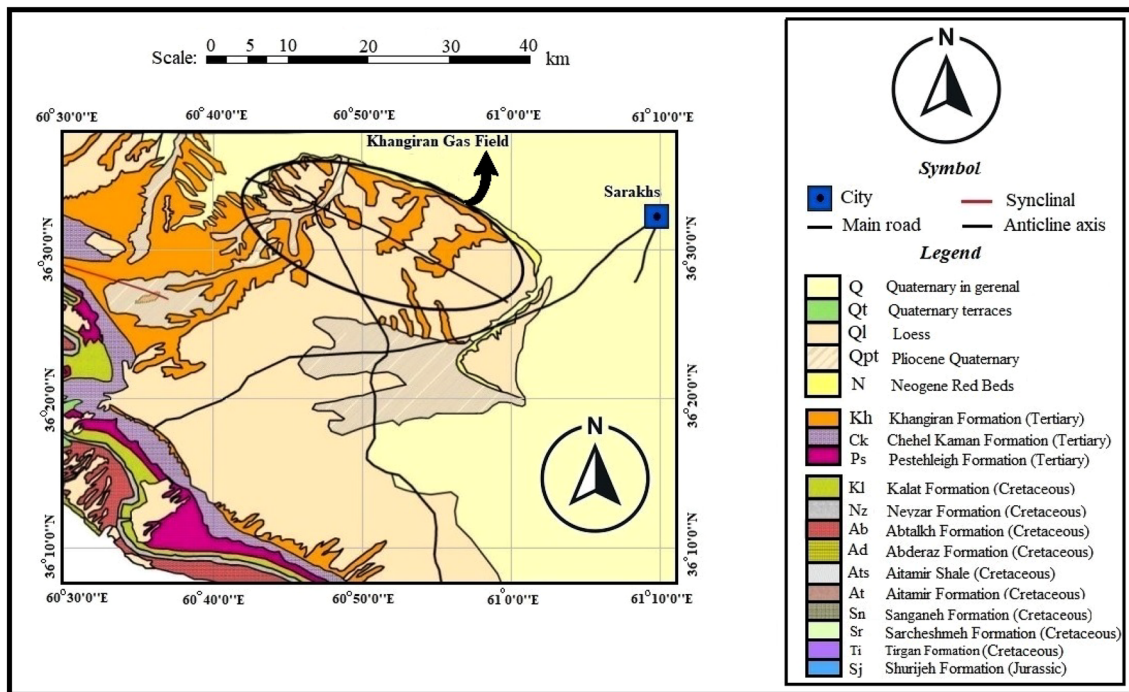
the research by Syed et al. in the year 2022, wherein they observed an increasing application of ML in various shale-related investigations. In a separate study conducted by Mohammadinia et al. (2023), the aim was to propose simplified techniques for shale volume estimation in a reservoir located in southern Iran. Furthermore, they sought to compare the performance of various ML methods in estimating shale volume. The conventional methods employed for comparison included gamma ray (GR), density-neutron (DN), and density-sonic (DS), while the ML methods consisted of ANN, support vector machine (SVM), and RF. The authors deduced that ANN, SVM, and RF methods estimated the shale volume with much better performance.

Since there are no detailed published data of the Kashafrud reservoir studies, the current research has been performed to investigate and evaluate its reservoir parameters including the shale content ( $V_{sh}$ ), and porosity ( $\emptyset$ ) by the laboratory, and petrophysical methods, as well as the intelligent methods (such as ANN). The aim of this paper was to shed light on the possible role of machine learning to estimate two critical parameters of reservoir quality assessment, porosity and volume of shale, in the Kashafrud Formation. During this process, the high accuracy of estimated parameters by artificial intelligence was carefully evaluated. The performance of the artificial neural network was measured using a criterion of comparing the results of calculations obtained for both results obtained from conventional petrophysical methods and artificial neural network methods. The presence of valuable core data in this study significantly improved the process of comparison and conclusion.

## Geological setting

The Kashafrud Formation is in the northeastern of Iran in a sedimentary basin of Kopet-Dagh (Fig. 1). The Kashafrud Formation was characterized as a reservoir by sedimentological and geochemical studies (Ershadnia et al. 2023). This sandstone formation, aging Aalenian-Bathonian (Middle Jurassic) mostly consists of the sedimentary rocks such as shale, sandstone, and conglomerate. A large area in the Kopet-Dagh basin, across the northeastern of Iran has been widespread by the Kashafrud Formation (Poursoltani & Gibling 2011).

The Khangiran anticline is approximately located at 180 km northeast of Mashhad and 25 km west of Sarakhs city. Based on the geophysical information, the general trend of the structure is northwest-southeast and it is asymmetric. Also, the northern edge has a steeper slope than the southern edge. In the mentioned anticline, the existence of three separate gas reservoirs including two sweet ones in the Shurijeh Formation, and one sour gas reservoir in the Mozdorán Formation has been confirmed (Mashayekhi et al. 2022). The Khangiran



**Fig. 1** The geographic location map of the studied area and well (Revised from Miri et al. 2018)

Formation, aging Lower–Middle Eocene with a thickness of 500 m is mainly consisted of succession of olive green, silty, calcareous and clay shales of gray, green-gray, silty, sticky, and calcareous rocks (Ghorbanpour et al. 2023). The stratigraphic column of the studied well (Well A) in the Khangiran gas field has been drawn by Strater software ver. 5 (Fig. 2).

The sandstones of Kashafrud Formation are mostly from the arkosic and lithic arenite types, rich of the fragments from the volcanic and sedimentary sources (Poursoltani and Gibling 2011). The thickness of drilled Kashafrud Formation is 433 m. The drilled thickness in the upper parts includes succession of light gray, light brown, gray, medium to coarse sandstones, hard to slightly porous bituminous, calcareous and gray, green-gray, silty and slightly pyrite. The lower drilled parts mainly consist of light gray, brownish gray, light brown, silty, sandy, calcareous, soft and thin layers of light gray sandstone, medium grain, semi-hard to hard (Ghorbanpour 2023).

## Materials and methods

### Core data

In the present research, the dataset was collected from one well, i.e., well A (Fig. 1) in the Khangiran gas field, NE Iran. This well was drilled to investigate the hydrocarbon status of the bottom formations under Mozdoran (especially Kashafrud

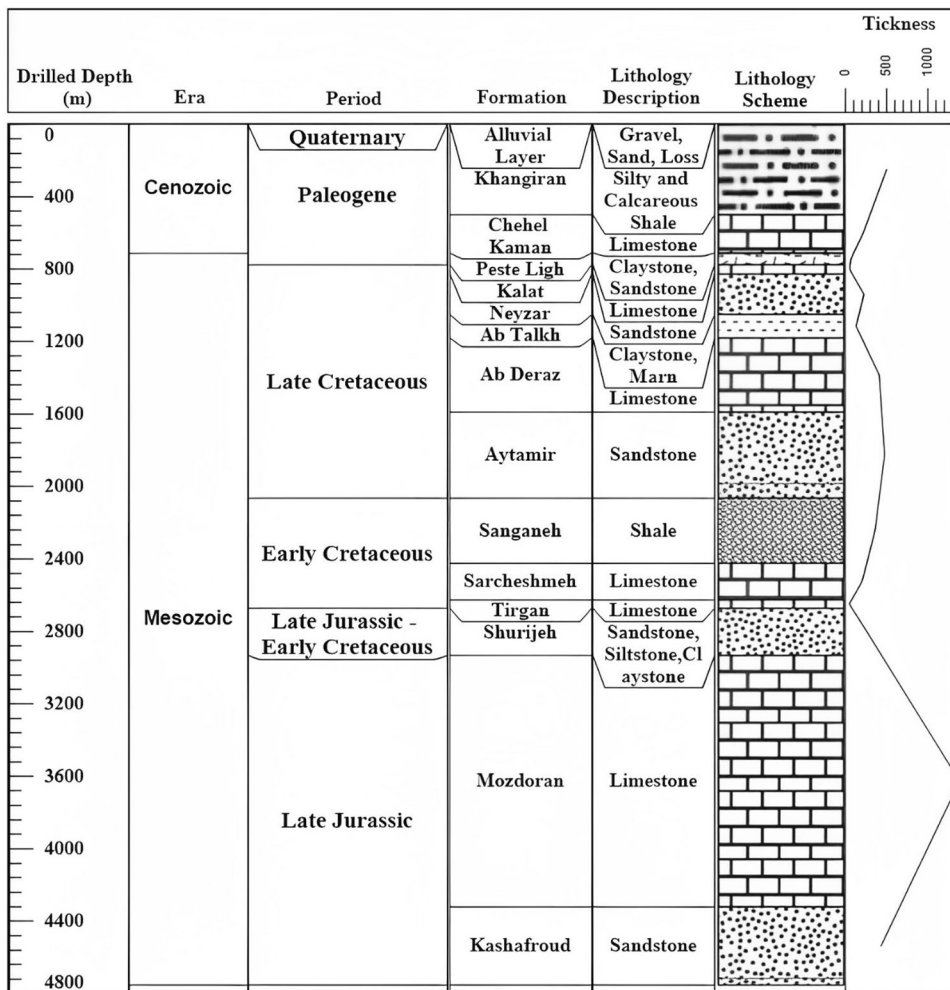
Formation) as well as the hydrocarbon production from the Mozdoran, and Kashafrud Formation. This well is drilled up to Kashafrud Formation (with a drilled thickness of 433 m).

Additionally, nine intervals were cored between depths of 3080.5 and 4397.5 m. During the drilling operation, nine core boxes of 0.9 m length were obtained. 10 core samples were then carefully selected and cut from the core #9 of well A. The laboratory measurements of the porosity (mercury porosimetry) and volume of shale (XRD test and densitometry) were performed on these 10 core samples.

### Wireline logging data

In this study, 2263 petrophysical data were provided with a depth interval of 0.061 m. One set of well-logging data from an eastern Kopet-Dagh field's gas producing well, including natural gamma ray (GR), sonic (DT), photoelectric (PEF), density (RHOB), neutron (NPHI), caliper (CALI), spontaneous potential (SP), and shallow & deep laterolog (LLS & LLD) were available from wireline logging process. Since there was a discrepancy between the depths of core samples and well logs, the depth matching was conducted by averaging between the upper and lower depths for each depth whose well-logging data was absent.

**Fig. 2** The stratigraphic column of the studied well in the Khangiran gas field, plotted by Strater software ver. 5

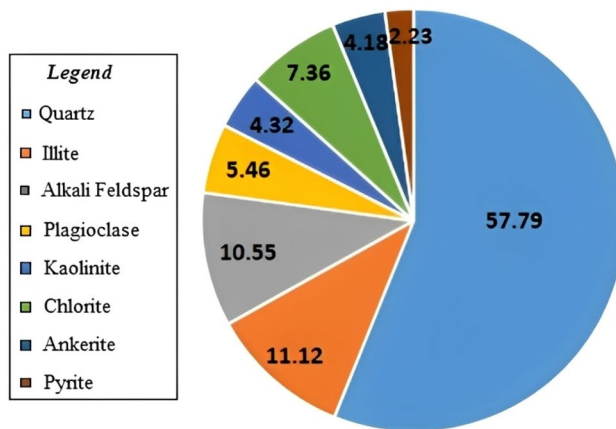


**Methods**

**Core analyses**

To analyze 10 core samples, the X-ray diffraction (XRD) was used to determine how much clay content existed in the Kashafroud Formation. The analysis was performed to fully identify the type of clay minerals and to calculate the laboratory shale weight percent. The results indicated that the constituent minerals in order of abundance in the studied samples were quartz, clay minerals, alkali feldspars, plagioclase, ankerite, and pyrite, respectively. In Fig. 3, the average weight percentage of each mineral in all samples is plotted separately.

The result of the XRD experiments is generally based on the weight percent and since it is necessary to make a comparison with the petrophysical data based on the volume percentage, one needs to have the density of each sample to convert the weight percent to the volume percent of clay minerals. The densitometry of the samples was performed by a 25-cc standard pycnometer by means of an



**Fig. 3** The average weight percentage of samples' minerals, obtained from the XRD analysis

organic fluid such as acetone. Therefore, the densitometry tests the samples were performed and each total weight

percent of clay minerals were converted to the volume percent. The relevant information is listed in Table 1.

The mercury porosimetry can detect the nanopores and macropores up to the size of 400  $\mu\text{m}$ . The mercury porosimetry remained the preferred method for analyzing the microporous materials (Schlumberger & Thommes 2021). Using the mercury porosimetry method, the porosity of 10 samples was precisely measured in the laboratory.

## The conventional petrophysical methods for porosity and volume of shale estimation

The most well-logging data can detect the clay minerals. Therefore, the estimation of the shale volume is possible from any logs. The definition of volume of shale in the literature is the ratio of the volume of fine grain particles such as silt and clay to the total volume of the rock (Shah et al. 2021). It has been proven that the gamma ray log and its spectral components (potassium, thorium, and uranium) are the best logs for the volume of shale estimation (Al Al-Azazi and Albaroot 2022; Khamees et al. 2022).

To determine the quantity of the clay minerals, the estimated volume of shale needs to be corrected. Below, Eqs. (1)–(6) illustrate the conventional petrophysical relationships to estimate the shale volume from the natural gamma ray log including Bhuyan and Passey (1994), Stieber (1973), Clavier (1971), Larionov-1 (according to the age of the Kashafud Formation) (1969), and combination of gamma density logs. The symbols, values, and parameters used in the formulas are listed in List of symbols section.

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (1)$$

$$V_{cl} = 0.6I_{GR} \quad (2)$$

$$V_{cl} = \frac{0.5 \times I_{GR}}{1.5 - I_{GR}} \quad (3)$$

$$V_{cl} = 1.7 - \sqrt{3.38 - (I_{GR} + 0.7)^2} \quad (4)$$

$$V_{cl} = 0.33 \left[ 2^{(2 \times I_{GR})} - 1.0 \right] \quad (5)$$

$$V_{cl} = I_{GR} \times \left( \frac{\rho_b}{\rho_{b_{sh}}} \right)^3 \quad (6)$$

After calculating the volume of shale using petrophysical and laboratory relationships, to measure the accuracy of the data, the values obtained from these two methods were compared. The results obtained from the petrophysical methods with error were calculated. Through this method, it is possible to match the volume percentages achieved in the laboratory with the values obtained through the experimental relationships and validation. According to the curve of the average percentage of errors (Fig. 4), the natural gamma ray (GR) was the criterion for further petrophysical studies and analytical methods such as the neural network.

After the calculations, it was observed that the average error rate was considerably high due to the laboratory validations (89.46%). Thus, the intelligent methods (ANN) became the basis of the next step.

In the porosity calculation segment, to compare the performance and results of both conventional petrophysical and laboratory methods, the average errors percentages obtained from these two methods were calculated (Table 2). It was observed that the average error rate was high, standing at 58.3%. Thereby to reduce the error rate, the ANN was chosen to accurately calculate the porosity at different depths.

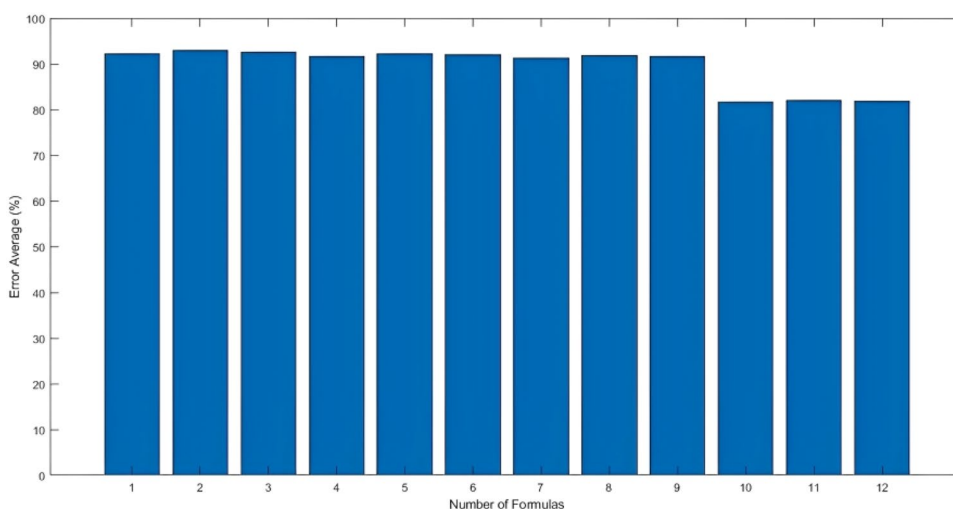
**Table 1** The density, volume percentage, and total volume percentage of the clay minerals

#Samples	$\rho_{lab} \left( \frac{g}{cm^3} \right)$	$V_{Kao} \%$	$V_{ill} \%$	$V_{Chl} \%$	$V_{Total} \%$
1	2.712	1.9	3.4	2.5	7.9
2	2.896	0.8	2.7	3.2	6.7
3	2.624	1.8	4.0	2.0	7.9
4	2.666	1.0	4.7	2.0	7.7
5	2.628	1.5	3.9	3.5	8.9
6	2.726	2.2	2.9	1.9	7.0
7	2.623	2.0	5.6	4.1	11.7
8	2.574	1.7	3.2	2.9	7.8
9	2.737	1.4	4.4	1.1	7.0
10	2.620	1.7	6.8	4.2	12.8
Average	2.681	1.6	4.2	2.8	8.5

## The estimation of the $\phi$ and $V_{sh}$ , using the multilayer perceptron (MLP) artificial neural network (ANN)

The neural network is a simulation of the human brain in the form of an artificial system that consists of a myriad of processor organs which are known as neurons with a special order that is similar to the human mind. A neural network consists of an input layer to the apply features of problem, a hidden layer to process, and an output layer to provide the answer(s). All of the training algorithms aim to minimize the mean squared error (MSE) between the outputs of the

**Fig. 4** The curve of average errors percentages for volume of shale. (1) Stieber/ $I_{GR}$ , (2) Stieber/ $I_{SGR}$ , (3) Stieber/ $I_{CGR}$ , (4) Clavier/ $I_{GR}$ , (5) Clavier/ $I_{SGR}$ , (6) Clavier/ $I_{CGR}$ , (7) Larionov/ $I_{GR}$ , (8) Larionov/ $I_{SGR}$ , (9) Larionov/ $I_{CGR}$ , (10) Gamma-Density logs/ $I_{GR}$ , (11) Gamma-Density logs/ $I_{SGR}$ , (12) Gamma-Density logs/ $I_{CGR}$



**Table 2** The corrective relationships and average error for porosity

Number	The name of method	The average error (%)
1	Neutron Log	63.95
2	Neutron–Density Log	52.65

predicted model and the observed outputs with respect to the training dataset (Adegbite et al. 2021). The methods based on the artificial intelligence proved their effectiveness and ability to provide robustness modeling on the basis of their high correlation coefficient between the actual and estimated volume of shale.

**The application of MLP network for the porosity estimation**

In the well under study, the core laboratory measured porosity contains 10 data. Since the artificial intelligence-based methods needs a large number of data to well train the network, the number of data has been increased to 686 based on the geostatistical algorithms. The selection of input data was performed in two ways. In the first approach, all the available logs were inserted to the MATLAB software ver. 2021 (Rajabi et al. 2021, 2023; Radwan et al. 2022; Abdelghany et al. 2023). In the other approach, some selected well-logging information chosen from the sensitivity analysis were inserted to the mentioned software as input data. In the both approaches, the input data was standardized to avoid one variable dominates the model.

In general, 70%, 15%, and 15% of the data were assigned for the training, validation, and testing, respectively. The Levenberg–Marquardt algorithm was used to train the MLP neural network. During several trainings,

the main criteria for evaluation of the most appropriate network, was chosen to be the mean squared error (MSE) and the correlation coefficient (R). The structure of the network consists of nine and six neurons (for the both approaches) in the input layer, one and two hidden layers, and one output layer. The input, output, parameters of the network, and their symbols are summarized in Table 3.

An outline of the optimal MLP network model for estimating porosity in the Kashafrud Formation can be seen in Fig. 5. All the possible mathematical functions for generation of the output were tested, and the results are listed in Table 4. There are several characteristics distinguishing each of neurons in the network, including the input weights, and the activation functions. Compared to the rest of the functions, the Tan-Sigmoid transfer function showed a better performance.

Since the input data ranged between 0 and 1, having a function that computes the output between zero and one was a logical reason for choosing the Tan-Sigmoid function. Therefore, the Tan-Sigmoid (tansig) was assumed as one of the most commonly used activation functions in the MLP networks. Equation (7) describes the Tan-Sigmoid function in which  $\beta$  indicates the slope parameter:

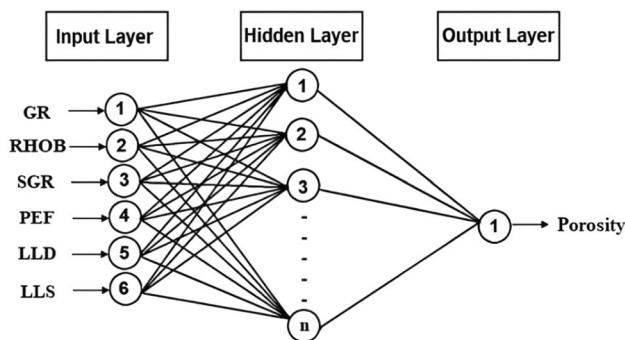
$$f(x) = \frac{1}{1 + e^{-\beta}} \tag{7}$$

First, all available well-logs data were entered into the ANN (Table 5). As the proposed MLP network had a low R-value and a high MSE, it did not capture the laboratory data successfully. To improve the results, it was decided to limit the input data to the most relevant parameters. To find the most correlated parameters with the porosity, a sensitivity analysis in the form of a Pearson matrix (Table 6) was performed.



**Table 3** The input and output data of the MLP network for estimating porosity

Data Type	Parameter	Symbol	Unit	Range	Mean	Standard Deviation	
Input	Sonic	DT	US/F	48.00–63.55	54.92	3.57	
	Density	RHOB	g/cm3	2.52–3.05	2.75	0.07	
	Neutron	NPHI	V/V	0.81–19.27	5.20	2.96	
	Micro Spherically	MSFL	Ohmm	0.20–161.58	26.53	27.84	
	Deep Resistance Radius	LLD	Ohmm	0.25–1949.91	184.96	211.22	
	Shallow Resistance Radius	LLS	Ohmm	0.38–1212.56	171.10	180.07	
	Spontaneous	SP	mV	–11.7904–28.0451	–32.59	3.61	
	Natural Gamma Ray	GR	API	28.92–140.35	84.12	32.01	
	SUM Gamma Ray	SCR	API	7.72–155.23	82.58	33.61	
	Standard Gamma Ray	CGR	API	-15.86–154.90	74.68	37.11	
	Caliper	CALI	IN	5.48–6.85	5.99	0.23	
	Photoelectric	PEF	B/E	–6.36–10.03	–2.45	3.70	
	Output	Porosity	∅	V/V	9.35–14.56	9.21	3.06



**Fig. 5** The outline of the optimal MLP network model for estimating porosity in the Kashafrud Formation

To calculate the correlation coefficient (R), and the MSE (Eq. (8)), the resulting estimates were compared to the actual measurements of the porosity in the laboratory.

$$MSE = \frac{\sum (x_{i_{meas}} - x_{i_{est}})^2}{N} \tag{8}$$

In which  $x_{i_{meas}}$  is the measured values,  $x_{i_{est}}$  is the estimated values, and N is the total number of observations. The input and output data were normalized as follows:

$$\text{Normalized value} = \frac{\text{Original Value} - \text{Minimum Value}}{\text{Maximum Value} - \text{Minimum Value}} \tag{9}$$

The porosity estimation results illustrated that when both all and selected well logs were put into the network, 288, 62, and 62 data were assigned for training, validation, and testing, respectively. It was also observed that the estimation when all logs were inserted into model as the input, has high MSE and also low R-values (Table 5). Considering all other tested topologies, 6-7-1 was the chosen architecture, having the highest R-value and the lowest MSE (Table 7). The MSE and R values were the best model at epoch 103 (Fig. 6), when the optimal MLP model was gained for estimation of the porosity in the

**Table 4** The MLP network with a fixed topology (Tansig) replied to the different transfer functions

Transfer Functions	Training Data Set		Validation Data Set		Testing Data Set	
	MSE	R-value	MSE	R-value	MSE	R-value
Linear (purelin)	$7.65620 \times 10^{-6}$	0.9711	$6.98140 \times 10^{-6}$	0.9215	$6.14881 \times 10^{-5}$	0.8536
Saturating Linear (satlin)	$5.28117 \times 10^{-6}$	0.9756	$7.10514 \times 10^{-5}$	0.9278	$6.84528 \times 10^{-6}$	0.8533
Log-Sigmoid (logsig)	$7.15336 \times 10^{-5}$	0.9654	$4.95746 \times 10^{-4}$	0.9284	$4.53892 \times 10^{-5}$	0.7819
Triangular Basis (tribas)	$6.33542 \times 10^{-5}$	0.9561	$6.85332 \times 10^{-5}$	0.9347	$6.91255 \times 10^{-5}$	0.4584
Tan-Sigmoid (tansig)	$3.00125 \times 10^{-4}$	0.9990	$3.72011 \times 10^{-3}$	0.9984	$3.74141 \times 10^{-3}$	0.9995
Hard-Limit (hardlim)	$4.01228 \times 10^{-4}$	0.9437	$4.98411 \times 10^{-3}$	0.9265	$5.00352 \times 10^{-3}$	0.7765
Symmetric Hard-Limit (hardlims)	$7.43754 \times 10^{-4}$	0.9812	$6.12504 \times 10^{-3}$	0.9445	$6.84312 \times 10^{-3}$	0.8798

**Table 5** Mean squared error values and correlation coefficient for all logs for porosity estimation

Topology	Training Data Set		Validation Data Set		Testing Data Set	
	MSE	R-value	MSE	R-value	MSE	R-value
9-1-1	$7.91002 e^{-8}$	0.8369	$7.56740 e^{-8}$	0.8351	$5.62996 e^{-8}$	0.8378
9-1-1-1	$5.57123 e^{-8}$	0.8859	$6.01566 e^{-8}$	0.9009	$5.96234 e^{-9}$	0.8596
9-2-1	$5.37781 e^{-8}$	0.8448	$8.03052 e^{-8}$	0.8448	$6.96929 e^{-8}$	0.9019
9-1-2-1	$6.83208 e^{-8}$	0.9092	$5.60017 e^{-8}$	0.9246	$7.89825 e^{-9}$	0.8408
9-3-1	$8.70518 e^{-8}$	0.8135	$9.29469 e^{-8}$	0.8246	$1.23906 e^{-8}$	0.8228
9-1-3-1	$6.45184 e^{-8}$	0.9136	$7.23164 e^{-8}$	0.9382	$8.19125 e^{-8}$	0.8848
9-4-1	$6.82980 e^{-9}$	0.8471	$6.99655 e^{-8}$	0.8561	$5.33850 e^{-9}$	0.8471
9-1-4-1	$5.32485 e^{-8}$	0.9035	$8.11126 e^{-8}$	0.9193	$6.88546 e^{-8}$	0.8961
9-5-1	$8.01085 e^{-8}$	0.8467	$7.68625 e^{-8}$	0.8886	$7.36588 e^{-9}$	0.8365
9-1-5-1	$7.11672 e^{-8}$	0.8864	$8.00114 e^{-8}$	0.9136	$6.28320 e^{-8}$	0.8781
9-6-1	$6.84132 e^{-8}$	0.8232	$5.16285 e^{-8}$	0.9059	$6.75098 e^{-9}$	0.8471
9-7-1	$7.14580 e^{-8}$	0.8164	$5.04534 e^{-8}$	0.8374	$8.67665 e^{-8}$	0.8225
9-8-1	$7.27324 e^{-8}$	0.8258	$6.74763 e^{-8}$	0.8117	$8.00868 e^{-8}$	0.8511
9-9-1	$5.38482 e^{-8}$	0.8364	$5.32811 e^{-8}$	0.8337	$5.69783 e^{-8}$	0.9054
9-10-1	$6.93440 e^{-8}$	0.8337	$8.93726 e^{-8}$	0.8325	$5.59001 e^{-8}$	0.8166

**Table 6** The Pearson correlation matrix to determine the selected well logs in the ANN to estimate porosity

	GR	DT	CALI	RHOB	NPHI	SGR	PEF	CGR	MSFL	LLD	LLS	SP	∅
GR	1.00												
DT	0.96	1.00											
CALI	-0.08	-0.04	1.00										
RHOB	-0.80	-0.81	-0.37	1.00									
NPHI	0.75	0.75	-0.51	-0.32	1.00								
SGR	0.39	0.50	-0.71	-0.02	0.80	1.00							
PEF	0.02	0.04	0.92	-0.50	-0.45	-0.74	1.00						
CGR	0.40	0.48	-0.76	0.04	0.81	0.94	-0.73	1.00					
MSFL	0.36	0.20	-0.48	-0.14	0.31	0.26	-0.48	0.15	1.00				
LLD	-0.96	-0.97	0.16	0.74	-0.79	-0.50	0.04	-0.55	-0.22	1.00			
LLS	-0.97	-0.98	0.15	0.75	-0.79	-0.51	0.04	-0.55	-0.23	1.00	1.00		
SP	-0.41	-0.46	-0.14	0.63	-0.10	0.00	-0.29	0.15	-0.40	0.37	0.38	1.00	
∅	-0.42	-0.25	-0.29	0.63	-0.26	0.31	-0.37	0.14	0.04	0.35	0.35	-0.12	1.00

Kashafrud Formation. Table 8 compares the results of networks' trainings with one and two hidden layers. MSE and R values of  $2.78731 E^{-4}$  (Fig. 6) and 0.9999 (Fig. 7), respectively, were found with an optimal MLP model for the porosity estimation in the Kashafrud Formation.

Finally, as a major step, to estimate the porosity in the whole interval of the Kashafrud Formation from the core data that were not used since this step (10 samples), the optimal MLP network was employed. The average output which indicates the estimated porosity was equal to 0.13%.

### The application of MLP network for the volume of shale estimation

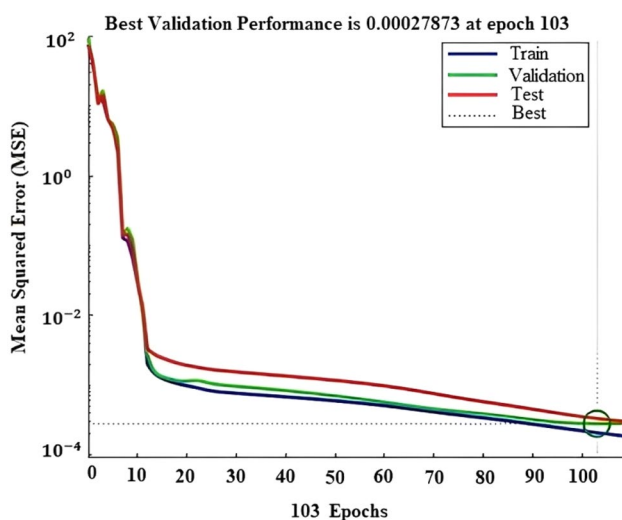
The core laboratory measured  $V_{sh}$  contains 10 data. Since the artificial intelligence-based methods needs a large number of data to well train the network, the number of data has been increased to 702 based on the geostatistical algorithms. It was observed that the  $V_{sh}$  estimation when all logs were inserted into model as the input has high MSE and also low R-values (Table 11). In the selection of proper inputs for the ANN model, the Pearson Correlation Coefficient (Table 8)

**Table 7** Mean squared error and correlation coefficient values for the selected well logs

Topology	Training Data Set MSE	Training Data Set R-value	Validation Data Set MSE	Validation Data Set R-value	Testing Data Set MSE	Testing Data Set R-value
6-1-1	$1/00686 e^{-4}$	0/9982	$9/32391 e^{-4}$	0/9871	$1/03811 e^{-5}$	0/9844
6-1-1-1	$5/63259 e^{-5}$	0/9979	$1/86091 e^{-4}$	0/9810	$1/80566 e^{-5}$	0/9942
6-2-1	$7/51577 e^{-4}$	0/9849	$9/75866 e^{-4}$	0/9937	$9/21556 e^{-4}$	0/9896
6-1-2-1	$1/56951 e^{-6}$	0/9912	$1/33013 e^{-7}$	0/9817	$5/21167 e^{-6}$	0/9946
6-3-1	$2/00632 e^{-5}$	0/9965	$3/03791 e^{-5}$	0/9929	$2/51106 e^{-5}$	0/9940
6-1-3-1	$4/08831 e^{-6}$	0/9875	$2/13467 e^{-5}$	0/9825	$4/48762 e^{-6}$	0/9954
6-4-1	$1/01417 e^{-7}$	0/9946	$2/50835 e^{-7}$	0/9934	$7/27836 e^{-7}$	0/9989
6-1-4-1	$1/99623 e^{-7}$	0/9881	$4/94222 e^{-4}$	0/9837	$2/51420 e^{-6}$	0/9761
6-5-1	$6/36613 e^{-7}$	0/9852	$2/14714 e^{-7}$	0/9955	$4/45812 e^{-7}$	0/9968
9-1-5-1	$2/76758 e^{-5}$	0/9949	$3/22903 e^{-5}$	0/9914	$8/34676 e^{-4}$	0/9832
6-6-1	$2/39039 e^{-5}$	0/9717	$4/94626 e^{-5}$	0/9875	$4/14706 e^{-5}$	0/9959
6-1-6-1	$4/99232 e^{-4}$	0/9965	$1/93628 e^{-6}$	0/9927	$2/17671 e^{-7}$	0/9957
6-7-1	$2/06925 e^{-4}$	0/9999	$2/78731 e^{-4}$	0/9999	$3/31220 e^{-4}$	0/9998
6-1-7-1	$2/12551 e^{-5}$	0/9873	$1/51897 e^{-6}$	0/9939	$2/66016 e^{-7}$	0/9960
6-8-1	$3/11573 e^{-8}$	0/9936	$3/92132 e^{-8}$	0/9842	$3/78817 e^{-6}$	0/9830
6-9-1	$6/24967 e^{-5}$	0/9971	$1/62983 e^{-4}$	0/9821	$5/25848 e^{-5}$	0/9936
6-10-1	$2/82603 e^{-7}$	0/9878	$3/00237 e^{-7}$	0/9978	$2/77961 e^{-7}$	0/9923

has been used extensively. An outline of the optimal MLP network model for estimating volume of shale is shown in Fig. 8. The input, output, parameters of the network, and their symbols are summarized in Table 9. Compared to the rest of the different transfer functions, the Tan-Sigmoid function showed a better performance (Table 10).

Similar to the previous estimation, before using the results of the Pearson correlation matrix, the values of R and MSE had significant errors (Table 11). However, after

**Fig. 6** Training, validation, and testing curves of the samples with the chosen topology 6–7-1

considering the sensitivity analysis, a significant improvement was seen in the trains (Table 12). The conventional methods for the estimation of  $V_{sh}$  are used which produce inconsistent results. Though, comparing all the tested networks, 6-8-1 had the least mean squared error and the highest correlation coefficient (Table 12). Through this, MSE and R-values of  $1.28701 E^{-9}$  (Fig. 9) and 0.9999 (Fig. 10), respectively, were found at epoch 1000 with an optimal MLP model for the  $V_{sh}$  estimation in the Kashafrud Formation.

Finally, as a major step, to estimate the  $V_{sh}$  in the whole interval of the Kashafrud Formation from the core data that were not used since this step (10 samples), the optimal MLP network was employed. The average output which indicates the estimated  $V_{sh}$  was equal to 8.34%.

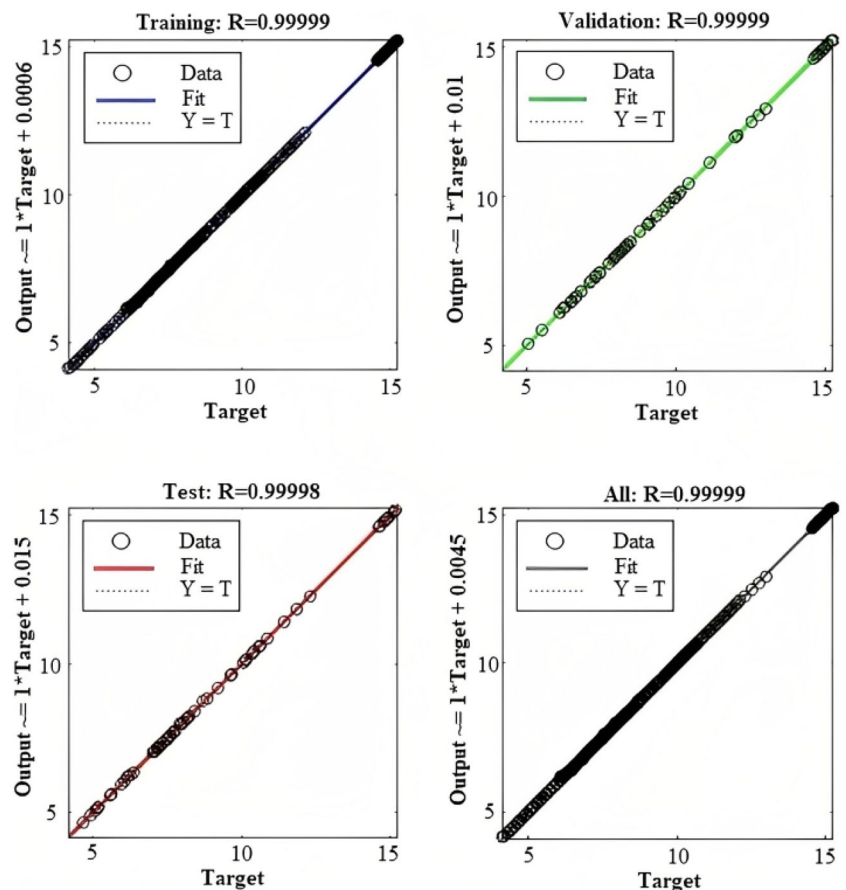
## Results and discussion

In this study, the laboratory analyses including the Powder X-ray diffraction (PXRD), and densitometry were performed. Based on the results of XRD analysis, the minerals in order of abundance are quartz, clay minerals, alkali feldspars, plagioclase, ankerite, and pyrite. Since the percentage of the clay minerals in the XRD test is weight percentage, this result should be compared with petrophysical data expressed in terms of volume percentage. To calculate the volume percent, a densitometry test was performed and its results showed that the highest and lowest amounts of clay minerals were 12.8 and 6.7 volume percent, respectively

**Table 8** The Pearson correlation matrix to determine the selected well logs in the ANN to estimate  $V_{sh}$

	GR	DT	CALI	RHOB	NPHI	SGR	PEF	CGR	MSFL	LLD	LLS	SP	$V_{sh}$
GR	1.00												
DT	0.96	1.00											
CALI	-0.08	-0.04	1.00										
RHOB	-0.80	-0.81	-0.37	1.00									
NPHI	0.75	0.75	-0.51	-0.32	1.00								
SGR	0.39	0.50	-0.71	-0.02	0.80	1.00							
PEF	0.02	0.04	0.92	-0.50	-0.45	-0.74	1.00						
CGR	0.40	0.48	-0.76	0.04	0.81	0.94	-0.73	1.00					
MSFL	0.36	0.20	-0.48	-0.14	0.31	0.26	-0.48	0.15	1.00				
LLD	-0.96	-0.97	0.16	0.74	-0.79	-0.50	0.04	-0.55	-0.22	1.00			
LLS	-0.97	-0.98	0.15	0.75	-0.79	-0.51	0.04	-0.55	-0.23	1.00	1.00		
SP	-0.41	-0.46	-0.14	0.63	-0.10	0.00	-0.29	0.15	-0.40	0.37	0.38	1.00	
$V_{sh}$	0.34	0.37	-0.36	-0.09	0.31	0.57	-0.56	0.49	0.44	-0.35	-0.35	0.04	1.00

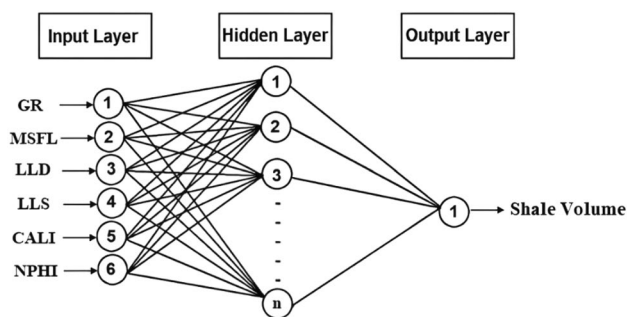
**Fig. 7** The estimation error curves for the samples with the chosen topology 6–7-1



(Table 1). Then, having the density, the laboratory  $V_{sh}$  of all samples was obtained. The results are presented in Table 13.

Figure 11 plots the data from the density log in terms of the natural gamma ray. The lowest amount of the natural GR of the studied well was chosen as the clean sand baseline, and the highest amount of the natural GR was chosen as the

clay baseline, based on the data of the natural GR of the studied field. The range of all samples lead to the clay baseline, and most core samples have a density between 2.68 and 2.72 g/cm<sup>3</sup> and subsequently, their radioactivity is relatively high. Thus, this indicates a high amount of the clay minerals, and this was confirmed in the laboratory studies.



**Fig. 8** The outline of the optimal MLP network model for estimating  $V_{sh}$  in the Kashafrud Formation

Furthermore, SEM studies were performed to determine the distribution pattern of the clay minerals. Based on these studies, the distribution pattern of clay minerals was mainly of pore filling type and in some limited cases, pore coating type was identified. The size of clay minerals varied from 1.4  $\mu\text{m}$  to 40  $\mu\text{m}$ . The performed EDAX analysis confirmed the minerals identified by XRD method. The most important achievements of this research include the following findings:

The results of volume of shale estimation showed that the natural gamma ray can be used as a criterion for further petrophysical studies and mathematical analytical methods (such as neural network) due to the lower error rate. Stieber calibration and combination of gamma-density logs also had the lowest mean error. The MLP neural network recorded an acceptable and appropriate performance for estimating the  $\varnothing$  and  $V_{sh}$  based on the selected logs. Though, when all the logs were imported as input, the error values prevented the appropriate topology from being selected as the best performance.

The challenge with the utilized method is the effort required to carefully select the appropriate training data, which is a common requirement for all models that use real well-logging data. However, the ANN helps to fuse different data and acquire complete and accurate information. Furthermore, this approach minimizes computing time, saving both time and money that would have been spent on core sampling without any prior knowledge of the matrix material or pore fluid.

The most important results of this study including following:

**Table 9** The input and output data used of the MLP network for estimating the volume of shale

Data Type	Parameter	Symbol	Unit	Range	Mean	Standard Deviation	
Input		DRHO	$\text{g/cm}^3$	-0.01319–0.013065	-0.00234	0.00677	
	Density	RHOB	$\text{g/cm}^3$	2.52–3.05	2.75	0.07	
	Neutron	NPHI	V/V	0.81–19.27	5.20	2.96	
	Micro Spherically	MSFL	Ohmm	0.20–161.58	26.53	27.84	
	Deep Resistance Radius	LLD	Ohmm	0.25–1949.91	184.96	211.22	
	Shallow Resistance Radius	LLS	Ohmm	0.38–1212.56	171.10	180.07	
	Spontaneous	SP	mV	-11.7904 to -28.0451	-32.59	3.61	
	Natural Gamma Ray	GR	API	28.92–140.35	84.12	32.01	
	SUM Gamma Ray	SCR	API	7.72–155.23	82.58	33.61	
	Standard Gamma Ray	CGR	API	-15.86 to 154.90	74.68	37.11	
	Caliper	CALI	IN	5.48–6.85	5.99	0.23	
	Photoelectric	PEF	B/E	-6.36–10.03	-2.45	3.70	
	Output	Volume of shale	$V_{sh}$	V/V	9.35–14.56	9.21	3.06

**Table 10** The MLP network with a fixed topology (Tansig) replied to the different transfer function

Transfer Functions	Training data set		Validation data set		Testing data set	
	MSE	R-value	MSE	R-value	MSE	R-value
Linear (purelin)	$7.65620 \times 10^{-5}$	0.9655	$6.98140 \times 10^{-4}$	0.9322	$6.14881 \times 10^{-4}$	0.8476
Saturating Linear (satlin)	$7.8546 \times 10^{-6}$	0.9416	$6.11458 \times 10^{-7}$	0.9347	$7.39827 \times 10^{-6}$	0.9112
Log-Sigmoid (logsig)	$7.35981 \times 10^{-5}$	0.9534	$6.31259 \times 10^{-6}$	0.9127	$6.13944 \times 10^{-6}$	0.8971
Triangular Basis (tribas)	$6.41226 \times 10^{-7}$	0.9614	$6.13147 \times 10^{-7}$	0.9172	$6.88342 \times 10^{-7}$	0.7694
Tan-Sigmoid (tansig)	$4.11676 \times 10^{-4}$	0.9982	$4.89541 \times 10^{-4}$	0.9988	$5.24151 \times 10^{-4}$	0.9993
Hard-Limit (hardlim)	$7.27814 \times 10^{-6}$	0.9619	$6.88545 \times 10^{-5}$	0.9584	$6.13836 \times 10^{-5}$	0.7925
Symmetric Hard-Limit (hardlims)	$5.98762 \times 10^{-6}$	0.9436	$5.17513 \times 10^{-7}$	0.9887	$6.44125 \times 10^{-6}$	0.8311

**Table 11** Mean squared error values and correlation coefficient for all well logs

Topology	Training Data Set		Validation Data Set		Testing Data Set	
	MSE	R-value	MSE	R-value	MSE	R-value
9-1-1	$8.56741 e^{-13}$	0.8891	$8.62475 e^{-14}$	0.8099	$8.29856 e^{-13}$	0.8540
9-1-1-1	$8.22139 e^{-13}$	0.8321	$7.95785 e^{-13}$	0.8221	$8.51259 e^{-14}$	0.8301
9-2-1	$9/13052 e^{-13}$	0.8679	$8.96939 e^{-13}$	0.8156	$7.07274 e^{-13}$	0.8431
9-1-2-1	$7/10773 e^{-13}$	0.8327	$7.71861 e^{-13}$	0.8235	$7.31949 e^{-14}$	0.8311
9-3-1	$7/29479 e^{-13}$	0.8657	$7.23907 e^{-13}$	0.8239	$7.51679 e^{-13}$	0.8533
9-1-3-1	$7/18049 e^{-13}$	0.8454	$8.57123 e^{-13}$	0.8446	$8.84628 e^{-13}$	0.8325
9-4-1	$9/99655 e^{-14}$	0.8574	$7.33954 e^{-13}$	0.8143	$7.93335 e^{-13}$	0.8336
9-1-4-1	$7/37131 e^{-14}$	0.8458	$7.22186 e^{-13}$	0.8112	$8.32872 e^{-13}$	0.8447
9-5-1	$7/68625 e^{-14}$	0.8570	$7.36588 e^{-13}$	0.8236	$7.89259 e^{-14}$	0.7990
9-1-5-1	$8/62776 e^{-13}$	0.8363	$8.84827 e^{-13}$	0.8224	$8.21433 e^{-13}$	0.8560
9-6-1	$7/16285 e^{-14}$	0.8566	$8.75098 e^{-13}$	0.8154	$7.17637 e^{-13}$	0.8527
9-7-1	$8/04534 e^{-14}$	0.8561	$7.67665 e^{-13}$	0.8152	$8.24428 e^{-14}$	0.8524
9-8-1	$7/74763 e^{-13}$	0.8742	$8.15868 e^{-15}$	0.8227	$8.82621 e^{-14}$	0.8673
9-9-1	$7/32811 e^{-14}$	0.8723	$8.69783 e^{-13}$	0.8224	$8.78067 e^{-14}$	0.8501
9-10-1	$8/93726 e^{-13}$	0.8711	$8.59327 e^{-13}$	0.8147	$7.18787 e^{-14}$	0.8491

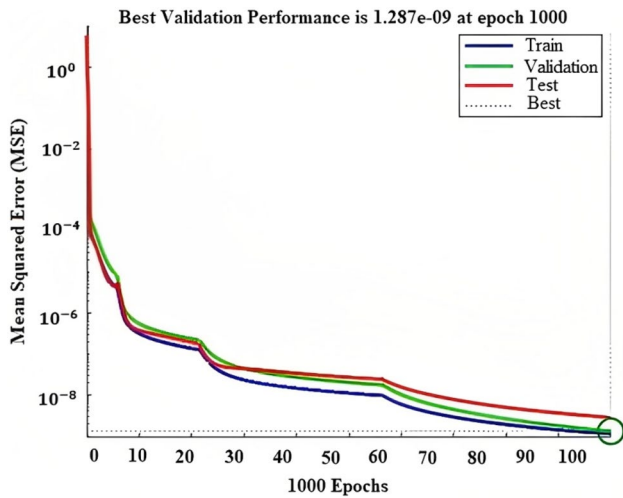
**Table 12** Mean squared error and correlation coefficient values for the selected well logs

Topology	Training Data Set		Validation Data Set		Testing Data Set	
	MSE	R-value	MSE	R-value	MSE	R-value
6-1-1	$1.22255 e^{-9}$	0.9994	$1.64936 e^{-9}$	0.9997	$5.03397 e^{-10}$	0.9877
6-1-1-1	$1.41151 e^{-9}$	0.9905	$3.00731 e^{-9}$	0.9715	$1.79318 e^{-9}$	0.9923
6-2-1	$2.11646 e^{-11}$	0.9989	$2.71556 e^{-11}$	0.9816	$3.56141 e^{-11}$	0.9984
6-1-2-1	$1.41151 e^{-9}$	0.9911	$3.00731 e^{-9}$	0.9964	$7.79318 e^{-7}$	0.9934
6-3-1	$5.44506 e^{-9}$	0.9821	$5.31570 e^{-9}$	0.9984	$4.22742 e^{-9}$	0.9979
6-1-3-1	$5.98975 e^{-10}$	0.9914	$5.95757 e^{-10}$	0.9967	$2.37883 e^{-7}$	0.9891
6-4-1	$3.96992 e^{-10}$	0.9984	$9.44087 e^{-10}$	0.9980	$9.04711 e^{-10}$	0.9971
6-1-4-1	$3.12938 e^{-10}$	0.9924	$5.80521 e^{-6}$	0.9970	$5.24420 e^{-6}$	0.9942
6-5-1	$8.67375 e^{-11}$	0.9980	$2.36584 e^{-10}$	0.9978	$3.90108 e^{-11}$	0.9875
6-1-5-1	$6.53578 e^{-10}$	0.9877	$4.53723 e^{-8}$	0.9814	$5.53134 e^{-8}$	0.9946
6-6-1	$2.00402 e^{-11}$	0.9828	$2.80022 e^{-11}$	0.9857	$4.19939 e^{-11}$	0.9836
6-7-1	$5.95261 e^{-10}$	0.9824	$1.43539 e^{-9}$	0.9979	$1.75258 e^{-9}$	0.9824
6-8-1	$1.10214 e^{-9}$	0.9999	$1.28701 e^{-9}$	0.9999	$2.78935 e^{-9}$	0.9999
6-9-1	$4.30431 e^{-10}$	0.9822	$8.59774 e^{-10}$	0.9974	$1.02560 e^{-9}$	0.9821
6-10-1	$3.10115 e^{-9}$	0.9973	$5.49104 e^{-9}$	0.9846	$1.48319 e^{-8}$	0.9964

1. The average laboratory value of  $V_{sh}$  for all the core samples was %8.88.
2. The average value of  $V_{sh}$  based on the petrophysical relationships in the entire under-studied well was %0.88.
3. The average value of  $V_{sh}$  based on the conducted MLP-ANN in the entire under-studied well was %8.34.
4. The average value of  $\emptyset$  based on the conducted MLP-ANN in the entire under-studied well was %0.13.

The aim of this study was to estimate the porosity and volume of shale in the Kashfrud gas reservoir in the

Khangiran field. The validation conducted in the research was highly valuable as it compared the results obtained from two distinct approaches, enabling authors to accurately and reliably calculate the percentages of improvement in estimating the two parameters. The superiority of the present investigation compared to other similar published studies in the field of oil exploration is the integration of conventional petrophysical methods (calculation of parameters methods using traditional relationships), and intelligent ML methods (MLP—ANN) to improve the accuracy of reservoir parameters estimation.



**Fig. 9** Training, validation, and testing curves of the samples with the chosen topology 6-8-1

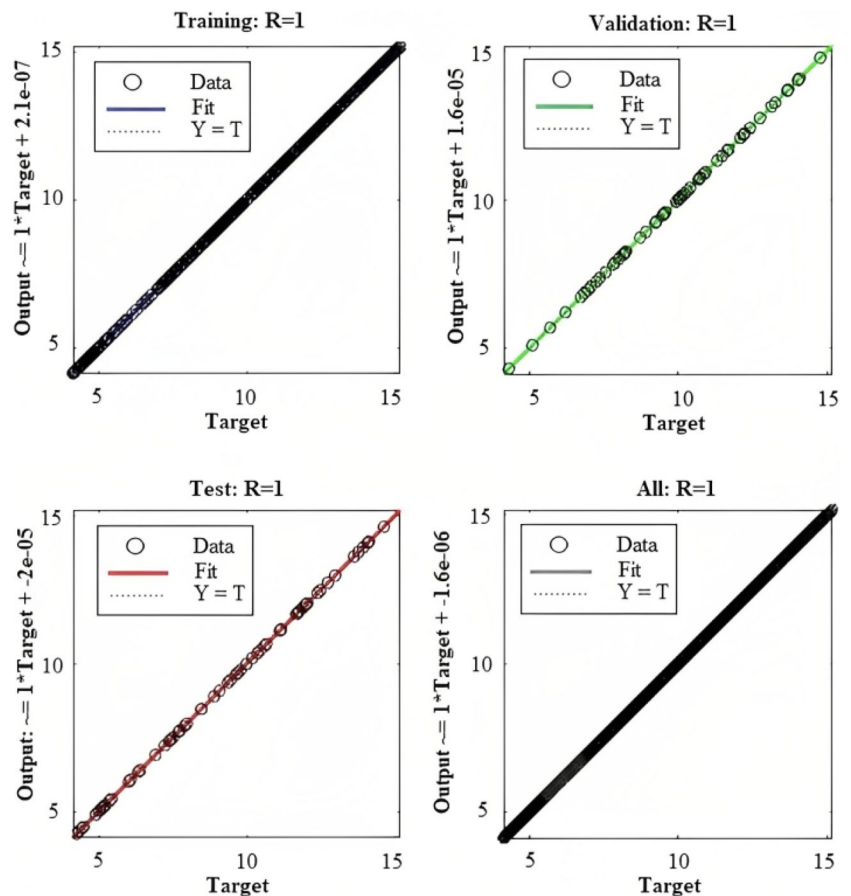
The ANN developed in the study can be employed and examined to estimate porosity and volume of shale values in other wells within the gas field with reliable accuracy, when real well log or core samples data are not available. Given

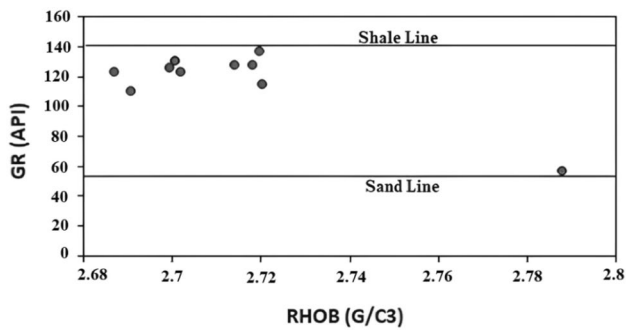
**Table 13** The values of density and laboratory volume of shale for each sample

The number of samples	The density of samples (g/cm <sup>3</sup> )	The laboratory volume of shale (%)
1	2.712	7.85
2	2.896	6.66
3	2.624	7.85
4	2.666	7.73
5	2.628	12.33
6	2.726	7.08
7	2.623	11.74
8	2.574	7.81
9	2.737	6.98
10	2.620	12.79

the high expenses associated with exploration operations in the oil industry, the method in this article presents a golden opportunity to save time and money by intelligent and modern techniques like ANNs to quickly and accurately estimate reservoir parameters using initial data from the target field. This approach can greatly assist petroleum engineers in

**Fig. 10** The estimation error curves for the samples with the chosen topology 6-8-1





**Fig. 11** Density vs. natural gamma-ray

tackling time-consuming and challenging problems related to the petroleum engineering field.

## Conclusions

Traditionally, the porosity and volume of shale are estimated with a very high error rate, which was the main reason that the multilayer perceptron (MLP) artificial neural network (ANN) was conducted to reduce the error. The application of MLP resulted in the significant error percentage decrease in the estimation of two parameters including  $\emptyset$  and  $V_{sh}$  from 58.3% and 89.46% in the traditional petrophysical method to  $2.78731 \text{ E}^{-4}$  and  $1.28701 \text{ E}^{-9}$ , respectively. According to the validation of obtained results from the application of the MLP method with the core analysis data, the porosity and volume of shale in the understudy field has been assessed to be highly accurate.

The correlation coefficient (R-value) and mean squared error (MSE) in the estimation process were improved considerably in comparison with conventional methods. Furthermore, the correlation coefficient for the both estimations were 0.9999, using the MLP method. Besides, the ability of MLP-ANN made great percentage of improvements, which were 99.95% for  $\emptyset$ , and 99.99% for  $V_{sh}$ . The obtained results of this investigation using MLP-ANN made great percentage of improvements (99.95% for  $\emptyset$ , and 99.99% for  $V_{sh}$ ), which has greatly impacted the estimation of in place hydrocarbon.

**Acknowledgements** The authors would like to acknowledge the X-ray laboratory, School of Mining Engineering, College of Engineering, University of Tehran for kind participation and collaborations made throughout the present study.

**Funding** The authors received no financial support for the research, authorship, and/or publication of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abdelghany WK, Hamed MS, Radwan AE (2023) Implications of machine learning on geomechanical characterization and sand management: a case study from Hilal field, Gulf of Suez, Egypt. *J Pet Explor Prod Technol* 13(1):297–312. <https://doi.org/10.1007/s13202-022-01551-9>
- Adegbite JO, Belhaj H, Bera A (2021) Investigations on the relationship among the porosity, permeability and pore throat size of transition zone samples in carbonate reservoirs using multiple regression analysis, artificial neural network and adaptive neuro-fuzzy interface system. *Pet Res* 6(4):321–332. <https://doi.org/10.1016/j.ptlrs.2021.05.005>
- Al Al-Azazi NA, Albaroot M (2022) Effect evaluation of shale types on hydrocarbon potential using well logs and cross plot approach, Halewah oilfield, Sab'atayn Basin, Yemen. *Energy Geosci* 3(2):202–210. <https://doi.org/10.55699/ijogr.2023.0301.1037>
- Alessa S, Sakhaee-Pour A, Sadooni FN, Al-Kuwari HA (2021) Comprehensive pore size characterization of Midra shale. *J Petrol Sci Eng* 203:108576. <https://doi.org/10.1016/j.petrol.2021.108576>
- Alessa S, Sakhaee-Pour A, Sadooni FN, Al-Kuwari HA (2022) Capillary pressure correction of cuttings. *J Petrol Sci Eng* 217:110908. <https://doi.org/10.1016/j.petrol.2022.110908>
- Ali M (2021) Machine learning based shale volume prediction from the Norwegian North Sea (Master's thesis, uis)
- Alipour KM, Kasha A, Sakhaee-Pour A, Sadooni FN, Al-Kuwari HAS (2022) Empirical relation for capillary pressure in shale. *Petrophysics* 63(05):591–603. <https://doi.org/10.30632/PJV63N5-2022a2>
- Balaky SM, Al-Dabagh MM, Asaad IS, Tamar-Agha M, Ali MS, Radwan AE (2023) Sedimentological and petrophysical heterogeneities controls on reservoir characterization of the Upper Triassic shallow marine carbonate Kurra Chine Formation, Northern Iraq: Integration of outcrop and subsurface data. *Mar Pet Geol* 149:106085. <https://doi.org/10.1016/j.marpetgeo.2022.106085>
- Bhuyan K, Passey QR (1994) Clay estimation from GR and neutron-density porosity logs. In: SPWLA 35th Annual Logging Symposium. Tulsa, Oklahoma, OnePetro. SPWLA-1994-DDD
- Cheng Y, Pan Z (2020) Reservoir properties of Chinese tectonic coal: a review. *Fuel* 260:116350. <https://doi.org/10.1016/j.fuel.2019.116350>
- Clavier C, Hoyle W, Meunier D (1971) Quantitative interpretation of thermal neutron decay time logs: part I. Fundamentals and techniques. *J Pet Technol* 23(06):743–755. <https://doi.org/10.2118/2658-A-PA>
- Doveton JH (1986). Log analysis of subsurface geology: concepts and computer methods
- Ehsan M, Gu H, Ahmad Z, Akhtar MM, Abbasi SS (2019) A modified approach for volumetric evaluation of shaly sand formations from conventional well logs: A case study from the talhar shale, Pakistan. *Arab J Sci Eng* 44(1):417–428. <https://doi.org/10.1007/s13369-018-3476-8>



- El-Gendy NH, Radwan AE, Waziry MA, Dodd TJ, Barakat MK (2022) An integrated sedimentological, rock typing, image logs, and artificial neural networks analysis for reservoir quality assessment of the heterogeneous fluvial-deltaic Messinian Abu Madi reservoirs, Salma field, onshore East Nile Delta, Egypt *Mar Pet Geol* 145:105910. <https://doi.org/10.1016/j.marpetgeo.2022.105910>
- Ershadinia M, Ghaemi F, Homam SM (2023) Permian to recent tectonic evolution of the Palaeotethys suture zone in NE Iran. *J Asian Earth Sci*. <https://doi.org/10.1016/j.jseae.2023.105658>
- Gamal H, Elkatatny S (2021) Prediction model based on an artificial neural network for rock porosity. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-021-05912-0>
- Ghorbanpour Yami H, Naqibi A, Alaviyan SM, Bahari A (2023) Different qualities of cement banding in geological formations of Khangiran Gas Field NE, Iran. *J Pet Res*. <https://doi.org/10.22078/PR.2022.4901.3191>
- Gong B, Keele D, Toumelin E, Clinch S (2019) Estimating net sand from borehole images in laminated deepwater reservoirs with a neural network. *Petrophys SPWLA J Form Eval Reserv Descr* 60(05):596–604. <https://doi.org/10.30632/PJV60N5-2019a4>
- Hong DV, Tien HN (2022) Using artificial neural network to predict volume of shale from well logging data. *Молодые-Наукам о Земле*. [https://doi.org/10.46326/JMES.2021.62\(3\).06](https://doi.org/10.46326/JMES.2021.62(3).06)
- Hussain W, Ehsan M, Pan L, Wang X, Ali M, Din SU, Liang L (2023) Prospect evaluation of the cretaceous yageliemu clastic reservoir based on geophysical log data: A case study from the Yakela Gas Condensate Field, Tarim Basin, China. *Energies* 16(6):2721. <https://doi.org/10.3390/en16062721>
- Iltaf KH, Butt SEH (2023) Energy geoscience. *Energy* 4:100143. <https://doi.org/10.1016/j.engeos.2022.100143>
- Iqbal MA, Rezaee R (2020) Porosity and water saturation estimation for shale reservoirs: an example from Goldwyer formation Shale, Canning Basin. *Western Austr Energ* 13(23):6294. <https://doi.org/10.3390/en13236294>
- Ismail A, Zeinel-Din MY, Radwan AE, Gabr M (2023) Rock typing of the Miocene Hammam Faraun alluvial fan delta sandstone reservoir using well logs, nuclear magnetic resonance, artificial neural networks, and core analysis, Gulf of Suez, Egypt. *Geol J*. <https://doi.org/10.1002/gj.4747>
- Jozanikohan G, Abarghoeei MN (2022) The Fourier transform infrared spectroscopy (FTIR) analysis for the clay mineralogy studies in a clastic reservoir. *J Pet Explor Prod Technol*. <https://doi.org/10.1007/s13202-021-01449-y>
- Khamees LA, Alrazzaq AAAA, Humadi JI (2022) Different methods for determination of shale volume for Yamama formation in an oil field in southern Iraq. *Mater Today Proc*. <https://doi.org/10.1016/j.matpr.2022.01.455>
- Larionov VV (1969) Borehole radiometry. *Nedra, Moscow* 127:813
- Mashayekhi Z, Kadkhodaei A, Solgi A, Baba Zadeh A, Aleali M (2022) Facies analysis, diagenesis processes and sedimentary environment of Shurijeh Formation in Khangiran gas field. *J Pet Res* 32:36–59. <https://doi.org/10.22078/PR.2022.4650.3090>
- Miri M, Bagheri R, Akhlaghi MR, Sotohan F (2018) Geochemical evolution of saline formation water of the Mozduran gas reservoir. *J Stratigr Sedimentol Res* 34(4):39–56. <https://doi.org/10.22108/jssr.2019.114104.1074>
- Mohammadinia F, Ranjbar A, Kafi M, Shams M, Haghight F, Maleki M (2023) Shale volume estimation using ANN, SVR, and RF algorithms compared with conventional methods. *J Afr Earth Sci*. <https://doi.org/10.1016/j.jafrearsci.2023.104991>
- Mohammed AKA (2020) A review: controls on sandstone permeability during burial and its measurements comparison—example, Permian Rotliegend Sandstone. *Model Earth Syst Environ* 6(2):591–603. <https://doi.org/10.1007/s40808-019-00704-w>
- Mohavvel S, Jozanikohan G (2022) The application of neural network method in petrophysical evaluation of asmari formation in a producing well in southwest of Iran. *J Min Eng* 17(54):1–13. <https://doi.org/10.22034/ijme.2021.113997.1751>
- Okon AN, Adewole SE, Uguma EM (2021) Artificial neural network model for reservoir petrophysical properties: porosity, permeability and water saturation prediction. *Model Earth Syst Environ* 7(4):2373–2390. <https://doi.org/10.1007/s40808-020-01012-4>
- Poursoltani MR, Gibling MR (2011) Composition, porosity, and reservoir potential of the Middle Jurassic Kashafud Formation, northeast Iran. *Mar Pet Geol* 28(5):1094–1110. <https://doi.org/10.1016/j.marpetgeo.2010.11.004>
- Radwan AE, Abudeif AM, Attia MM (2020) Investigative petrophysical fingerprint technique using conventional and synthetic logs in siliciclastic reservoirs: a case study, Gulf of Suez basin. *Egypt J Afr Earth Sci* 167:103868. <https://doi.org/10.1016/j.jafrearsci.2020.103868>
- Radwan AE, Wood DA, Radwan AA (2022) Machine learning and data-driven prediction of pore pressure from geophysical logs: a case study for the Mangahewa gas field, New Zealand. *J Rock Mech Geotech Eng* 14(6):1799–1809. <https://doi.org/10.1016/j.jrmge.2022.01.012>
- Rajabi M, Beheshtian S, Davoodi S, Ghorbani H, Mohamadian N, Radwan AE, Alvar MA (2021) Novel hybrid machine learning optimizer algorithms to prediction of fracture density by petrophysical data. *J Pet Explor Prod Technol* 11:4375–4397. <https://doi.org/10.1007/s13202-021-01321-z>
- Rajabi M, Hazbeh O, Davoodi S, Wood DA, Tehrani PS, Ghorbani H, Radwan AE (2023) Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. *J Pet Explor Prod Technol* 13(1):19–42. <https://doi.org/10.1007/s13202-022-01531-z>
- Saikia P, Baruah RD, Singh SK, Chaudhuri PK (2020) Artificial Neural Networks in the domain of reservoir characterization: a review from shallow to deep models. *Comput Geosci* 135:104357. <https://doi.org/10.1016/j.cageo.2019.104357>
- Sakhaee-Pour A, Bryant SL (2012) Gas permeability of shale. *SPE Reserv Eval Eng* 15(04):401–409. <https://doi.org/10.2118/146944-PA>
- Sakhaee-Pour A, Bryant SL (2015) Pore structure of shale. *Fuel* 143:467–475. <https://doi.org/10.1016/j.fuel.2014.11.053>
- Sakhaee-Pour A, Li W (2016) Fractal dimensions of shale. *J Nat Gas Sci Eng* 30:578–582. <https://doi.org/10.1016/j.jngse.2016.02.044>
- Saleh AH, Hemimey WAE, Leila M (2023) Integrated geological and petrophysical approaches for characterizing the pre-neomanian nubian sandstone reservoirs in ramadan oil field, Central Gulf of Suez, Egypt. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-023-07743-7>
- Schlumberger C, Thommes M (2021) Characterization of hierarchically ordered porous materials by physisorption and mercury porosimetry—A tutorial review. *Adv Mater Interfaces* 8(4):2002181. <https://doi.org/10.1002/admi.202002181>
- Shah MS, Khan MHR, Rahman A, Islam MR, Ahmed SI, Molla MI, Butt S (2021) Petrophysical evaluation of well log data for reservoir characterization in Titas gas field, Bangladesh: a case study. *J Nat Gas Sci Eng* 95:104129. <https://doi.org/10.1016/j.jngse.2021.104129>
- Solanki P, Baldaniya D, Jogani D, Chaudhary B, Shah M, Kshirsagar A (2021) Artificial intelligence: new age of transformation in petroleum upstream. *Pet Res*. <https://doi.org/10.1016/j.ptlrs.2021.07.002>
- Sondergeld, C. H., Newsham, K. E., Comisky, J. T., Rice, M. C., & Rai, C. S. (2010, February). Petrophysical considerations in evaluating and producing shale gas resources. In SPE unconventional gas conference. OnePetro. <https://doi.org/10.2118/131768-MS>

- Steiber RG (1973) Optimization of shale volumes in open hole logs. *J Petrol Technol* 31(1973):147–162
- Syed FI, AlShamsi A, Dahaghi AK, Neghabhan S (2022) Application of ML & AI to model petrophysical and geomechanical properties of shale reservoirs—a systematic literature review. *Petroleum* 8(2):158–166. <https://doi.org/10.1016/j.petlm.2020.12.001>
- Szabó NP (2011) Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophys* 59(5):935–953. <https://doi.org/10.2478/s11600-011-0034-0>
- Taheri M, Ciabeghods AA, Nikrouz R, Kadkhodaie A (2021) Modeling of the shale volume in the hendijan oil field using seismic attributes and artificial neural networks. *Acta Geol Sin English Edition* 95(4):1322–1331. <https://doi.org/10.1111/1755-6724.14739>
- Tali AH, Farman GM (2021) Use conventional and statistical methods for porosity estimating in carbonate reservoir in southern Iraq, Case study. *Iraqi Geol J*. <https://doi.org/10.46717/igj.54.2D.3Ms-2021-10-22>
- Tran H, Sakhaee-Pour A (2018a) Slippage in shale based on acyclic pore model. *Int J Heat Mass Transf* 126:761–772. <https://doi.org/10.1016/j.ijheatmasstransfer.2018.05.138>
- Tran H, Sakhaee-Pour A (2018b) Critical properties ( $T_c$ ,  $P_c$ ) of shale gas at the core scale. *Int J Heat Mass Transf* 127:579–588. <https://doi.org/10.1016/j.ijheatmasstransfer.2018.08.054>
- Tran H, Sakhaee-Pour A, Bryant SL (2018) A simple relation for estimating shale permeability. *Transp Porous Media* 124:883–901. <https://doi.org/10.1007/s11242-018-1102-6>
- Yu C, Tran H, Sakhaee-Pour A (2018) Pore size of shale based on acyclic pore model. *Transp Porous Media* 124:345–368. <https://doi.org/10.1007/s11242-018-1068-4>
- Zhou Y, Yang W, Yin D (2022) Experimental investigation on reservoir damage caused by clay minerals after water injection in low permeability sandstone reservoirs. *J Pet Explor Prod Technol* 12(4):915–924. <https://doi.org/10.1007/s13202-021-01356-2>