چهل و دومین گردهمایی (همایش) ملی
علوم زمین
The 42nd National Geosciences Congress
www.geoconf.ir

Data-driven insights into gas reservoir depths determination: application of 1D-CNN algorithm in the Kangan and Upper-Dalan formations

Ali Gohari nezhad
M.SC. holder of Petroleum Production Engineering
University of Tehran, College of Engineering, Chemical Engineering faculty, Institute of Petroleum Engineering
Tehran, Iran
Ali.gohari.nezha@ut.ac.ir

Abstract— **Accurately determining the depths of gas reservoirs is a critical challenge, particularly in the Kangan and Upper Dalan formations of the South Pars gas field. Conventional methods, exemplified by Archie's equation, face limitations in such tight carbonate reservoirs, prompting the exploration of advanced techniques like NMR logging. However, the high costs and time-consuming nature of NMR logging necessitate alternative approaches. In this study, a solution grounded in data-driven insights by leveraging a 1D-CNN (One-Dimensional Convolutional Neural Network) algorithm has been proposed. This deep learning approach aims to provide precise depth determination while overcoming the challenges posed by traditional methods. The study methodology involves the individual implementation of the 1D-CNN algorithm and its integration into a comprehensive model for enhanced accuracy. By applying this algorithm, we intend to predict gas effective porosity profile based on well logs to determine productive zones and intervals in the Kangan and Upper Dalan formations. The dataset includes information from 5 wells, incorporating both training and testing wells, with an emphasis on validation through a blind well to ensure robustness. Unlike standard procedures, we go beyond mere prediction by comparing the algorithmic results with actual depths in geographically blind well. The study emphasizes the algorithm's industrial implementation capability by showcasing its effectiveness in predicting reservoir depths. Preliminary results indicate promising accuracy and stability, paving the way for a more intelligent model with practical applications in the delineation of production intervals. In conclusion, our research presents a data-driven approach to gas reservoir depth determination, specifically in the Kangan and Upper Dalan formations, utilizing the 1D-CNN algorithm. This study not only highlights the potential of this algorithm in overcoming traditional limitations but also underscores its practicality and cost-effectiveness as a valuable alternative to conventional methods and expensive logging techniques in complex reservoirs.**

Keywords: petrophysics, 1D-CNN, effective gas porosity, Dalan formation, Kangan formation

1. Introduction

porosity, a measure of empty space within a volume, is crucial in reservoir characterization studies. Effective porosity, the ratio of interconnected void space to the total medium volume, is more significant for Petro physicists. Core analysis measures porosity in rock samples but has limitations like limited sample representation, potential alteration, and time and cost issues for large-scale studies. Effective gas porosity is crucial in determining in-situ gas volume in depth intervals, identifying productive layers, and optimizing perforation intervals. It provides valuable information about a reservoir's ability to store and transmit gas, impacting production performance and recovery potential [1]. By having the hydrocarbon effective porosity profile across a well, operators can enhance well productivity by targeting these intervals for perforation.

The Archie's law [2] equation is the most common method for predicting hydrocarbon saturation and effective porosity profiles within a reservoir, but it encounters errors in tight carbonate reservoirs. NMR logs have various applications in petrophysics, including accurate estimation of reservoir water saturation, hydrocarbon saturation, volume, permeable porosity, and production layers. However, despite their high accuracy, methods based on NMR logs have exorbitant costs and often consume considerable time.

Although in recent years studies have been made to determine NMR log outputs using data driven models from conventional well logs [3], but the literature lacks a model capable of predicting effective gas porosity from NMR logs using machine learning techniques. By an overview into the facing challeng, it is evident that the South Pars gas field faces a significant challenge in defining production intervals. This study aims to address this by utilizing NMR logs, despite the high costs. The conventional method involves analyzing various well logs, but misinterpretation can lead to errors. This study introduces a 1D-CNN model to determine the NMR gas effective porosity across the well and focuses on model validation through the inclusion of a blind well in the field, providing insights into its measurement accuracy and contributing to the resolution of industrial challenges related to identifying productive depth intervals. An inteligent model has been developed to predict effective gas porosity from NMR log data, significantly reducing costs and time requirements compared to NMR logs. Figure 1 shows the study pipeline and the step by step procedure.

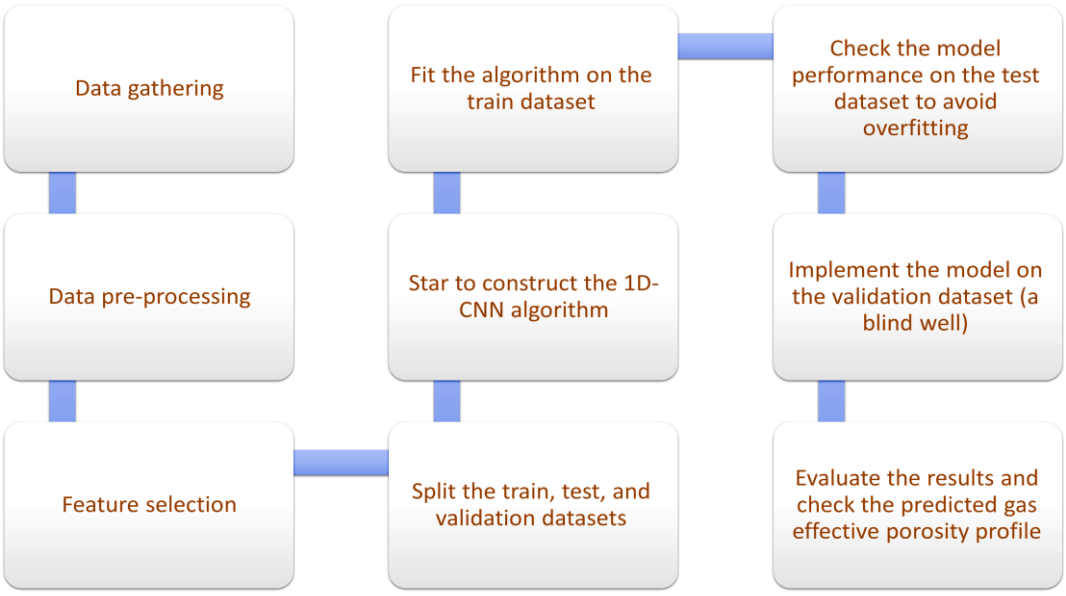


Figure. 1. A view of the study flowchart to construct the 1D-CNN intelligent model.

1.1 Case study geology overview

Tight carbonate gas reservoirs, like the South Pars gas field, are complex hydrocarbon reservoirs with low porosity and permeability, influenced by variations in grain size, mineralogy, and diagenetic alterations. These reservoirs present unique challenges and opportunities for exploration and production due to their intricate pore structure and heterogeneity of carbonate rocks. The South Pars gas field, located in Persian Gulf waters, is one of the largest gas fields in the world with a proven reserve of 441.5 tcf of gas in place [4]. The main reservoirs are the Dalan and Kangan formations, primarily composed of lime, anhydrite, and dolomite (Figure 2). Accurately predicting effective gas porosity is crucial for estimating hydrocarbon volumes and identifying suitable production layers for perforation. Conventional models often yield unrealistically low effective gas porosity values in these low-porosity intervals.

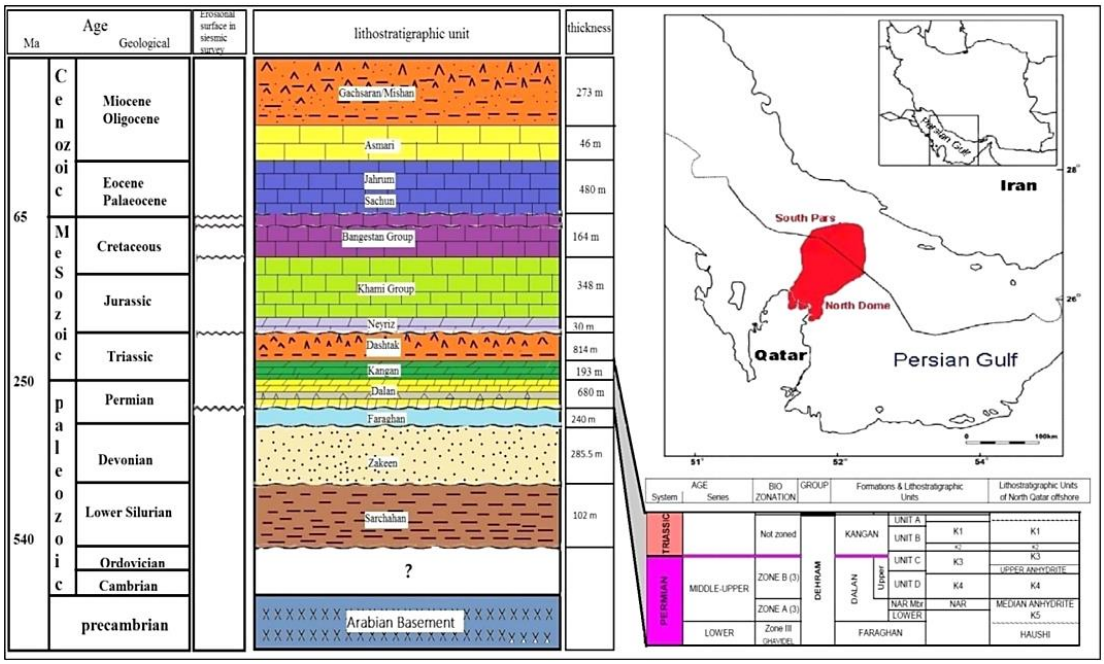


Figure. 2. Location map of South Pars gas field in the Persian Gulf (right) and stratigraphic position of the studied formations (left) [5].

The Dalan Formation, first discovered in the Dalan No. I well in the Dalan anticline southwest of Shiraz, is a significant geological unit that overlies the Furghun Formation. The lower section of the formation is marked by dark, massive fossiliferous limestones and dolomites, with an oolitic texture. This lower section thickens from the Zagros thrust zone to the southeast end of the Zagros geosyncline near Bander Abbas, indicating a Permian sea transgression. The lower part initially displays favorable secondary porosity, transitioning into predominantly anhydritic layers with notable anhydrite beds in the middle section. In the central part of Fars province, the Nar Member of the Dalan Formation undergoes specific lithological changes, transitioning from anhydrite to dolomite. Beyond Bander Abbas, the evaporite beds pinch out in the Furghun Formation. As we ascend, the Dalan Formation reverts to carbonate-rich, uniformly thick-bedded rock, primarily composed of micritic limestone and dolomite. However, an unspecified section is eroded by a post-Permian unconformity. In the upper part of the Permian section, there is a facies change to terrigenous rocks in the interior of Fars Province.

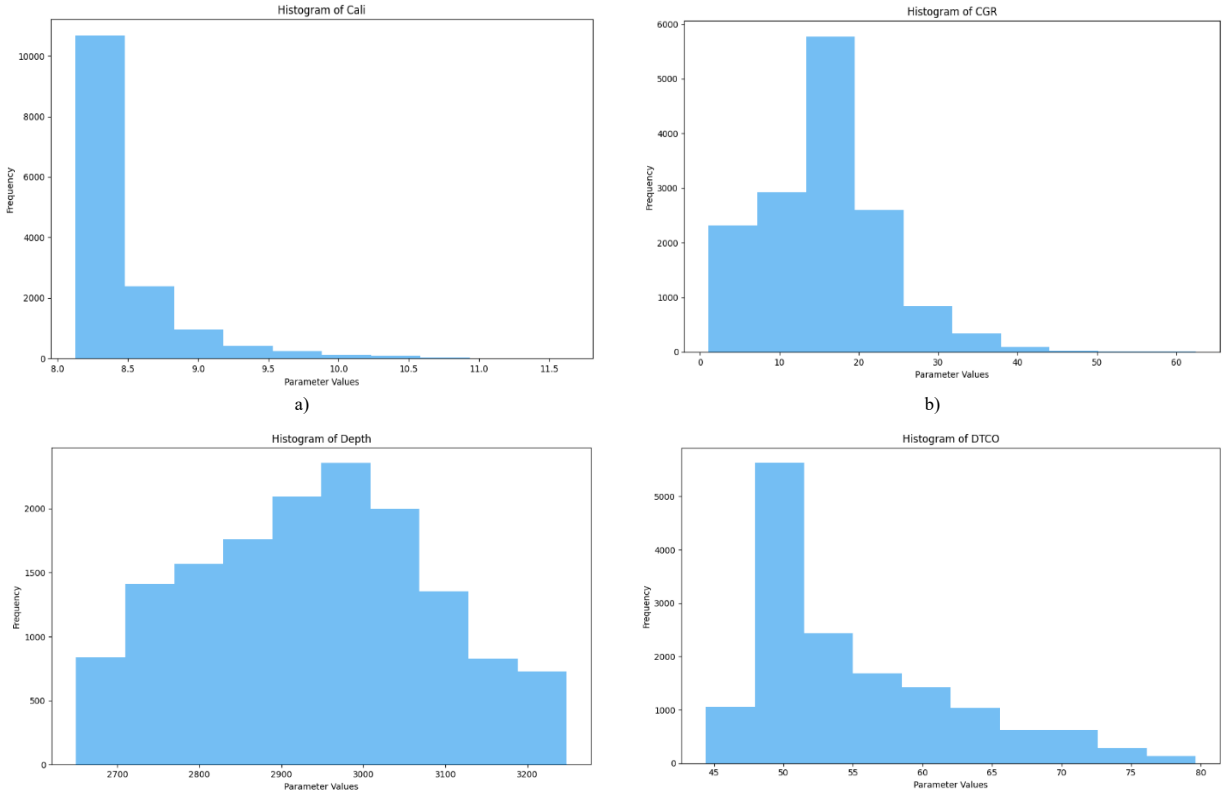
The Kangan Formation, located above the Dalan Formation, is characterized by oolitic carbonates with intermittent anhydrite beds. It is a significant structure in southern Iran, with a regional unconformity indicating the separation between the Permian and Triassic periods. The lower part of the Kangan Formation, featuring a

shale bed containing Claraia (pelecypod), suggests an Early Triassic age. The formation overlays Permian carbonates housing abundant fusulinids of Permian age. The Kangan anticline runs parallel to the main Zagros thrust, trending northwest in alignment with the Late Cretaceous-Paleocene compressional stress-field of the Zagros geosyncline. The deformation style of the Kangan structure is mainly concentric, utilizing Infra-Cambrian (Hormuz) and to a lesser extent Triassic evaporites as basal detachment surfaces. Isopach maps of the Permian section reveal the Permian Basin's elongate trough, trending NW-SE parallel to the Cretaceous-Tertiary Zagros geosyncline. Thickness decreases towards the SW in the Arabian Shelf area and to the NE towards the Iranian Plateau. The Permian depocenters were in the extreme SE portion of the trough, near the present Gulf of Oman, with more than 5,000 feet of Permian sediments deposited. The Kangan structure, located at the edge of the Permian trough, suggests a shallow-water interior shelf for the Permo-Triassic depositional environment, characterized by mainly intertidal and supratidal sedimentation [6].

3. **Methodology**

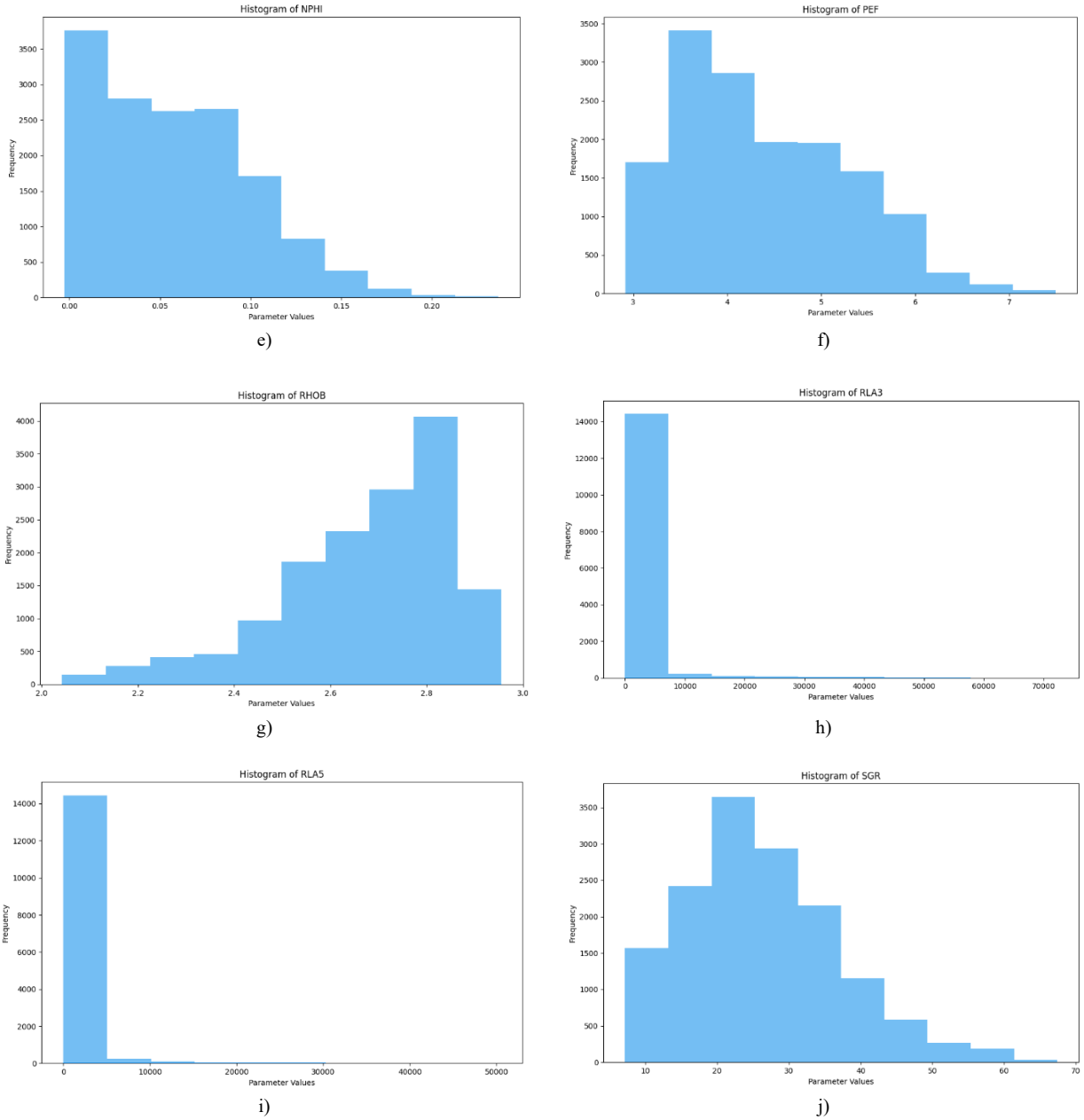
3.1 Data Gathering, pre-processing, and feature engineering

This study uses data from 5 gas well logging datasets in the South Pars field, including NMR log information, to enhance understanding of the reservoir's permeable porosity distribution and optimize gas production in challenging carbonate gas reservoirs. Data from 4 wells (A, B, C, D) were used for training and testing, while data from the wells F was used as the blind well for model validation. The primary input data is derived from common petrophysical logs, including DTCO, NPHI, RHOB, CAli, SGR, CGR, RLA3, and RLA5. The output data is the product of Phie (total effective porosity) and Sg_nmr (gas saturation estimated from nuclear magnetic resonance data). . Fig. 3 illustrates the cumulative histogram plot of all the available features.The cumulative frequencies or proportions of a dataset are shown in a cumulative histogram plot, derived from a histogram. It provides insights into data distribution, revealing the progression of frequencies and percentages. The plot can also determine percentiles and individual data points' relative positions within the dataset



a) b)

c) d)



e) f)

g) h)

i) j)

Figure. 3. Cumulative distribution histogram plot of a) Caliper. b) CGR. c) depth, d) DTCO, e) NPHI, f) PEF, g) Rhob, h) RLA3, i) RLA5, j) SGR.

Data preprocessing involves three main stages in this study: data cleaning, data denoising, and data normalization. Data cleaning removes meaningless readings from the log, while data denoising eliminates noise and outliers, facilitating a clearer understanding of the data. Data normalization scales data values uniformly using the Min-Max scaler, improving the efficiency of deep learning or machine learning algorithms

Feature engineering is a crucial aspect of machine learning and data analysis, transforming raw data into meaningful features that enhance model performance and provide valuable insights. Conventional methods

include univariate selection, Recursive Feature Elimination, Principal Component Analysis, feature importance methods, correlation analysis, L1 regularization, forward/backward selection, and feature scaling/normalization. In this study, petroleum engineering knowledge and heat map correlation were used to select input parameters, including Caliper (well diameter), DTCO (compressional wave transit time), NPHI (porosity derived from neutron log), PEF (photoelectric log reading), and RHOB (bulk density). These input features provide us the information for subsequent analysis and modeling, capturing important geological and petrophysical characteristics and contributing to a better understanding of subsurface conditions in the studied wells.

3.2 1D-CNN

Convolutional Neural Networks (CNNs) are a powerful and successful model in deep learning, particularly in image processing. Developed in the 1980s by Yan LeCun and Yann Bengio, CNNs serve as a robust tool for pattern recognition and feature extraction in images. They gained popularity and real development in subsequent decades, showcasing their success in image recognition competitions, especially in ImageNet competitions. Multilayer neural networks were among the earliest artificial neural network models in the history of deep learning, but they struggled to leverage spatial features and information from images. CNNs, introduced in the 1980s, were designed based on ideas independent of engineering and demands in image processing. They use convolutional layers to apply filters to images, extracting various spatial features, making them highly efficient and effective in pattern recognition and feature identification. 1D CNNs, unlike traditional 2D CNNs, focus on one-dimensional signals and are highly beneficial in tasks related to time-series data. They find widespread applications in signal processing, especially in extracting important features from audio signals, translating sound to text, or analyzing brain waves. Their ability to utilize spatial information in one-dimensional signals contributes to improving efficiency and accuracy in tasks related to time-series and sequential data [7]. The structure of a 1D CNN may include multiple convolutional layers, activation layers, and pooling layers, allowing the model to extract complex features from input data (Figure 4). This structure is suitable for various tasks in signal and time-series data processing, making it highly effective in pattern recognition, prediction, and data separation. The convolutional layer is responsible for feature extraction on the input vector, using a convolution kernel to execute local convolutions and define receptive fields. The size and type of the kernel significantly impact the features extracted. The training stage uses various kernels to extract multiple features, creating detailed input maps. The pooling layer is used to further extract features and reduce the dimensionality of the convolution result, similar to convolution. Common pooling functions include average and max pooling. The fully connected layer aligns with the traditional neural network structure and consists of multiple hidden layers. It combines the features extracted by the previous layers and integrates the information from the convolutional and pooling layers to produce the final output [8].

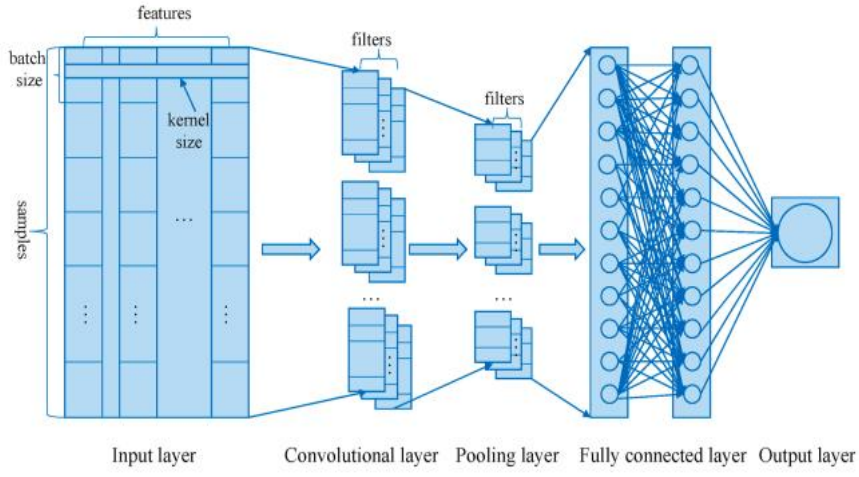


Figure. 4. A schematic to overview the structure of a 1D-CNN [9].

3.3 Algorithm impplementation and evaluation

The provided python code establishes a 1D Convolutional Neural Network (1D-CNN) using the Keras library with TensorFlow as the backend. The model architecture is configured sequentially, beginning with a 1D convolutional layer featuring 16 filters and a kernel size determined by the variable **kernelSize**. This layer, utilizing the Rectified Linear Unit (ReLU) activation function, processes input data with a shape of (nFeatures, 1). Subsequently, a dropout layer with a 20% dropout rate is incorporated, followed by the application of Leaky Rectified Linear Unit (Leaky ReLU) activation, introducing a small negative slope.

The model further incorporates three additional 1D convolutional layers with varying filter sizes (32, 64, 128) and kernel sizes (9, 9, 5). Another dropout layer, with a 20% dropout rate, is interposed, and Leaky ReLU is reapplied as the activation function. The fifth convolutional layer boasts 256 filters and a kernel size of 5, followed by Leaky ReLU activation. Subsequently, MaxPooling with a pool size of 2 is applied. The output from the convolutional layers undergoes flattening to facilitate input to the fully connected layers.

The succeeding layers consist of three fully connected layers with node sizes of 256, 128, and 64. Leaky ReLU is employed as an activation function for the first, and Batch Normalization is applied. The subsequent three fully connected layers exhibit decreasing node sizes (64, 32, 16). The final layer is a Dense layer with nodes equal to 1 and a sigmoid activation function (Figure 5). For compilation, the model utilizes the Adam optimizer with a learning rate of 0.001, employing Mean Squared Error (MSE) as the loss function and Mean Absolute Error (MAE) as a metric for assessment.

This 1D-CNN architecture is thoughtfully structured for regression tasks, incorporating strategic use of dropout layers and batch normalization to mitigate overfitting. Activation functions and optimizer choices align with established practices in neural network design, contributing to the network's stability and effectiveness during training.The provided code constructs a 1D-CNN tailored for regression tasks, incorporating convolutional and fully connected layers with activation functions, dropout, and normalization techniques to enhance its learning capabilities and generalization to new data. The structure of the model is organized to progressively extract features from the input data and make accurate predictions.

Models' performances were assessed using three mathematical expressions: R-squared (R^2), mean absolute error (*MAE*), and mean squared error Mean squared error (*MSE*). R^2 quantifies the proportion of dependent variable variation explained by independent variables, while MAE and MSE measure the average squared difference between anticipated and actual values. Both measures evaluate regression model performance, with lower values indicating greater performance (Table 1).

Table 1. The mathematical expressions of the performance metrics (loss functions).

Model evaluation parameters	Mathematical expression
R-squared (R^2)	$R^2 = 1 - \frac{\Sigma_i (y_i - \hat{y}_i)^2}{\Sigma_i (y_i - \bar{y}_i)^2}$
Mean squared error (*MSE*)	$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
Mean absolute error (*MAE*)	$MAE = \frac{1}{n}\sum_{j=1}^{n} \lvert y_i - \hat{y}_i \rvert$

Figure. 5. A schematic of the constructed 1D-CNN layers, depicting each layer's uinits number and activation function.

4. Results

After completing the model and algorithm construction, the next step involves implementation and examining the obtained results. To evaluate the performance, as mentioned earlier, three criteria have been utilized. The model has been implemented on three sets of data, and the results can be observed in Table 2. For a better understanding of the results, the predicted gas effective porosity values by the model are plotted against the NMR log-derived values (the target values) for each of the three data sets in Figure 6. The R-squared value obtained from each data set is also indicated. In Figure 7, a column chart for each of the three metrics in the study is drawn based on the obtained results from the model implementation. Finally, in Figure 8, the predicted gas effective porosity values are plotted alongside the values derived from the NMR log. Considering the trend of both plots, the high accuracy of the designed model in determining productive intervals (with suitable gas effective porosity) and non-productive intervals (with low gas effective porosity) can be observed.

Table 2. Statistical results of the implemented algorithm on the Train, Test, & Validation datasets.

	Train	Test	Validation
R-squared	0.989	0,981	0.905
MAE	0.237	0.395	0.934
MSE	0.487	1.02	2.04

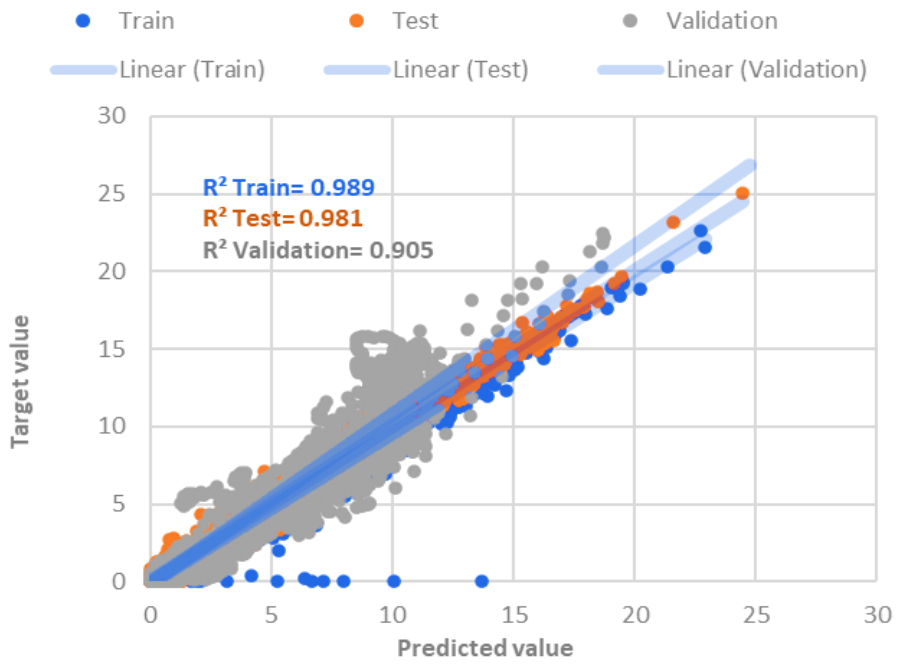


Figure 6. A cross plot of target value (NMR given) and algorithm predicted values of gas effective porosity for all 3 datasets..

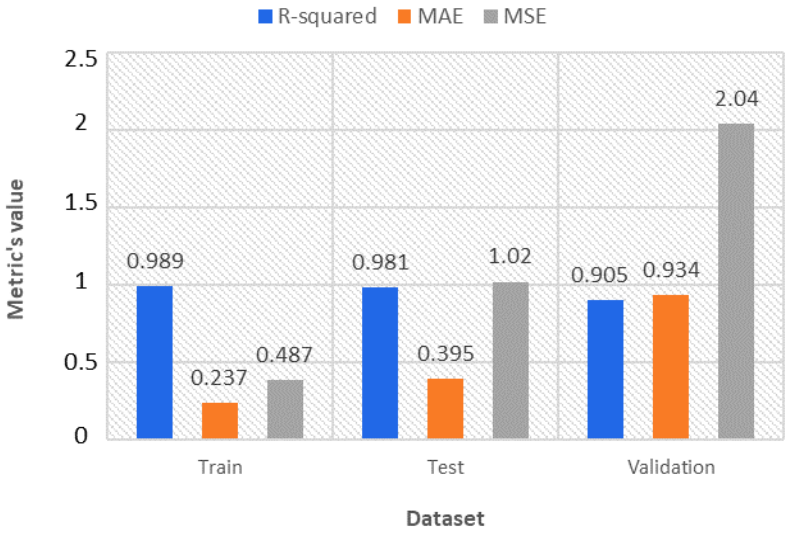


Figure 7. A column chart to compare the applied model accuracy on the train and test, and validation datasets in terms of R-squared (R^2), mean absolute error (MAE), mean squared error (MSE).

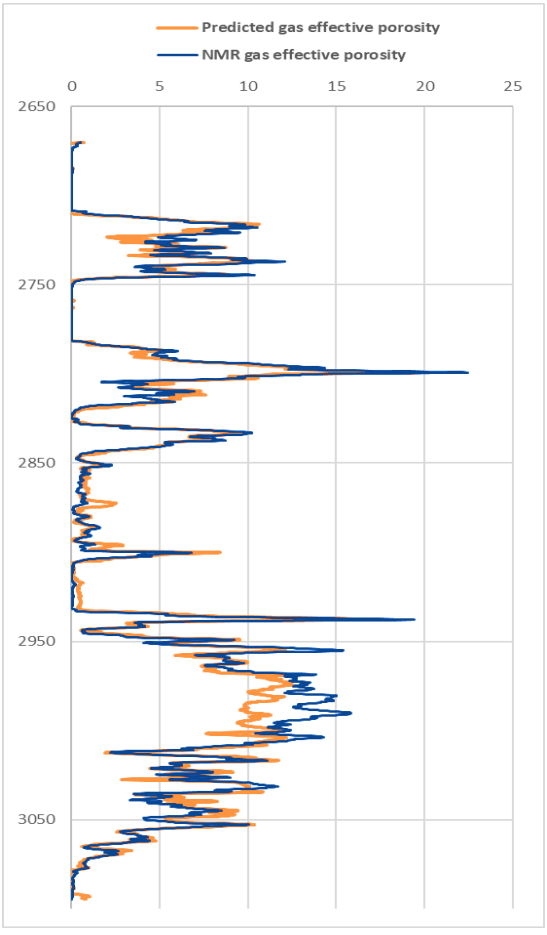


Figure 8. Comparison of NMR gas effective porosity profile and the algorithm predicted profile.

Productive intervals generally refer to depth intervals in which hydrocarbons first of all have a significant volume, so that production from it is economical, and then in these intervals, hydrocarbons must have the ability to move and transfer. It means that it is stored in holes that are connected to each other and have the ability to transfer hydrocarbons. In Figure No. 8, it can be seen that the model has been able to identify well the intervals that are prone to production, as well as the intervals that have a small volume of gas with the ability to transfer, and determine their boundaries as well, just like the NMR log.

5 Conclusion

In the South Pars field, the Kangan and Dalan Upper formations, characterized by a dolomite and calcite structure interwoven with anhydrite veins, exhibit a constrained porous environment, leading to frequently non-permeable porosities. The presence of these unique geological conditions introduces inaccuracies in conventional log readings, complicating the precise differentiation between productive and non-productive intervals. The NMR log stands out as the most accurate method for determining gas production depth intervals in these formations. However, its high costs and occasional lack of economic feasibility limit its widespread use. To address these challenges, this study employed a 1D-CNN algorithm, leveraging machine learning, to interpret common well logs continuously and predict gas effective porosity derived from NMR logs in the well. Consequently, the model predicts production and non-production layers. Data from five locations underwent meticulous pre-processing and feature engineering, entering the algorithm for analysis in three categories: training, testing, and validation. The R-squared results, 0.989, 0.981, and 0.905 for training, testing, and parameter measurement, respectively, affirm the commendable performance of the intelligent model in solving the presented problem. This study underscores the potential of artificial intelligence and data-driven algorithms in overcoming challenges in the fields of geology and petrophysics.

References

[1] Koponen A, Kataja M, Timonen J. Permeability and effective porosity of porous media. Phys. Rev. E. 1997 Sep;56(3):3319-3325.
[2] Archie GE. The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. Transactions of the AIME. 1942;146(01):54-62.
[3] Jamshidian, M., Hadian, M., Mansouri Zadeh, M., Kazempoor, Z., Bazargan, P., & Salehi, H. (2015). Prediction of free-flowing porosity and permeability based on conventional well logging data using artificial neural networks optimized by Imperialist competitive algorithm – A case study in the South Pars gas field. *Journal of Natural Gas Science and Engineering*, *24*, 89-98.
[4] Aali, J., Rahimpour-Bonab, H., & Kamali, M. R. (2006). Geochemistry and origin of the world's largest gas field from Persian Gulf, Iran. *Journal of Petroleum Science and Engineering*, *50*(3), 161-175.
[5] Ebrahim Sfidari, Ali Kadkhodaie-Ilkhchi, Hossain Rahimpour-Bbonab, & Behzad Soltani (2014). A hybrid approach for litho-facies characterization in the framework of sequence stratigraphy: A case study from the South Pars gas field, the Persian Gulf basin. *Journal of Petroleum Science and Engineering*, *121*, 87-102.
[6] Kashfi, S. M. (2008). Geology of the Permian "super-giant" gas reservoirs in the Greater Persian-Gulf Area. *Journal of Petroleum Geology*, *15*, 465-480.
[7] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing. 2021; 151:107398.
[8] Pires de Lima R, Duarte D, Nicholson C, Slatt R, Marfurt KJ. Petrographic microfacies classification with deep convolutional neural networks. Computers & Geosciences. 2020; 142:104481.
[9] Zhang Y, Zhang C, Ma Q, Zhang X, Zhou H. Automatic prediction of shear wave velocity using convolutional neural networks for different reservoirs in Ordos Basin. Journal of Petroleum Science and Engineering. 2022; 208:109252.