RESEARCH

Improved porosity estimation in complex carbonate reservoirs using hybrid CRNN deep learning model

Amirreza Mehrabi[1] · Majid Bagheri[2] · Majid Nabi Bidhendi[2] · Ebrahim Biniaz Delijani[1] · Mohammad Behnoud[1]

Abstract

This paper aims to improve porosity estimation in complex carbonate reservoirs by proposing a hybrid CRNN deep learning model. The objectives include addressing the challenges associated with porosity estimation in heterogeneous carbonate reservoirs and evaluating the performance of the CRNN model in accurately predicting porosity based on well-log data. The overall approach involves integrating CNN and RNN architectures within the CRNN model to effectively extract and combine relevant information from well logs. The model is trained using a dataset consisting of well-log and core analysis data from an Iranian carbonate oil field. Well-log data is used as the input including GR, DT, RHOB, LLD, and NPHI for model training, while core data is utilized for model validation. The model's performance is compared with the traditional MLP model in terms of accuracy and generalization. The proposed hybrid CRNN model demonstrates superior performance in predicting porosity values at new locations where only well-log data are available. It outperforms conventional neural network models, as evidenced by the significant improvement in the correlation coefficient between the model predictions and core data (from 0.67 for the MLP model to 0.98 for the CRNN model). The CRNN model's ability to capture complex spatial dependencies within heterogeneous carbonate reservoirs leads to more accurate porosity estimations and valuable insights into reservoir characterization. This paper presents novel and additive information to the existing body of literature in the petroleum industry. The hybrid CRNN model, combining CNN and RNN architectures, offers a unique approach to porosity estimation in complex carbonate reservoirs. By effectively integrating spatial and temporal patterns from well-log data, the model demonstrates higher accuracy rates and improved generalization capabilities. The findings contribute to the state of knowledge by providing a robust and efficient tool for accurate porosity prediction, which can assist in reservoir characterization and enhance decision-making in the petroleum industry.

Keywords Porosity estimation · CNN · RNN · Hybrid Model · Well log · Core data

Introduction

The petrophysical properties of the reservoir are vital for reservoir characterization and decision-making in reservoir engineering. Porosity estimation is a crucial aspect of reservoir characterization, providing information about reservoir rocks' storage capacity and flow properties. In carbonate reservoirs, porosity estimation can be challenging due to their complex depositional history and diagenetic processes that affect pore structure and connectivity (Bust et al. 2011, Kharraa et al. 2013, Bagrintseva 2015). Therefore, accurate porosity estimation in carbonate reservoirs requires a comprehensive understanding of the rock types, depositional environment, and diagenetic history. Several methods are available for porosity estimation in carbonate reservoirs, including conventional core analysis, well logs, and seismic data (Kennedy 2015; Zare et al. 2020; Lin et al. 2021; Bagheri and Riahi 2015). Core analysis is considered the most reliable method for determining porosity because it directly measures pore space using laboratory techniques such as h mercury porosimeter, helium pycnometer, and image analysis (McPhee et al. 2015), (Tiab and Donaldson 2016) However, core analysis is expensive and time-consuming. Well logs indirectly measure porosity by

✉ Majid Bagheri
  majidbagheri@ut.ac.ir

1 Department of Petroleum Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

2 Institute of Geophysics, University of Tehran, Tehran, Iran

analyzing rocks' electrical, acoustic, or nuclear properties (Kharraa et al. 2013, Kennedy 2015, Bagheri et al. 2019). The most common well logs for porosity estimation are neutron logs, density logs, and sonic logs (Darling 2005; Kennedy 2015; Ghosh 2022).

Machine learning (ML) algorithms have become increasingly popular in reservoir characterization due to their ability to process and analyze large amounts of data quickly and efficiently (Okon and Anyadiegwu 2021). They have a wide range of applications in reservoir characterization, including lithology identification, petrophysical parameter estimation, and seismic interpretation (Bagheri 2015, Tavakolizadeh 2022, Ahmadi 2019).

Recent advancements in machine learning have made it possible to accurately predict porosity using logging data (Okon and Anyadiegwu 2021, Li et al. 2022). Machine learning algorithms can analyze large amounts of data from various sources, such as well logs and seismic surveys, to identify patterns and make predictions (Iturrarán-Viveros and Parra 2014; Elkatatny et al. 2018). By training these algorithms on historical data sets, they can learn to recognize the relationships between different variables and accurately predict porosity in new reservoirs (Moosavi et al. 2022a) This approach has been successfully applied in several carbonate reservoirs around the world, resulting in more accurate predictions of porosity and improved hydrocarbon recovery rates. As machine learning technology continues to advance, this method will likely become even more effective at predicting porosity for complex carbonate reservoirs (Edet Ita Okon 2021, Lin et al. 2021). Artificial neural networks (ANNs) have become the most widely used method in reservoir property estimation, particularly porosity and permeability estimation, due to their ability to handle complex and non-linear relationships between input and output variables (Singh et al. 2016).

Although machine learning has been used in reservoir characterization, the complexity of the data in carbonate reservoir requires more advanced techniques to extract meaningful insights (Ahmadi and Chen 2019). As such, recent studies have focused on hybrid machine learning and deep learning to improve the prediction accuracy of machine learning (Chen et al. 2020, Matinkia et al. 2022). Deep learning, a subset of machine learning, has emerged as a powerful tool to tackle this challenge (Alzubaidi et al. 2021). Deep learning with a genetic algorithm is used to provide a reliable and effective model for reservoir porosity estimation in heterogeneous reservoirs (Chen et al. 2021; Wang and Cao 2022). Moreover, data gathering in heterogenous carbonate rocks is a challenging task due to data quality issues and noises in measurements. Several data noise reductions have been suggested to reduce the complexity of data (Cao et al. 2015). For example, the combination of support vector regression (SVR) and fuzzy Logic algorithms shows good performance in estimating the porosity of the reservoir. They employ fuzzy techniques to reduce the noise in datasets (Moosavi, Bagheri et al. 2022a, b). Numerous studies indicate combining clustering and prediction algorithms can lead to more accurate predictions and better insights into complex datasets (Ahmadi and Chen 2019). Clustering can help identify patterns and relationships within a dataset that may not be immediately apparent. By grouping similar data points, clustering algorithms can reveal underlying structures and trends that can be used to make predictions more accurately (Sun et al. 2024; Kaydani et al. 2012).

Some authors have also suggested hybrid deep-learning model techniques to improve the accuracy of regression models. This approach typically involves combining neural network approaches, for example, hybrid architecture involving Convolutional Neural Networks (CNN) combined with Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM), to create more accurate and robust models (Cui and Fearn 2018; Khan et al. 2020). As such, this paper addresses the application of hybrid deep learning architecture for regression tasks in reservoir property estimation.

Researchers in 2024, introduced a Convolutional Neural Network (CNN) and Transformer model to improve the accuracy and generalization ability of logging porosity prediction. The model was trained on a well log dataset. The CNN-transformer model showed good superiority in the task of logging porosity prediction with R-squared of 0.95 (Sun et al. 2024). Researchers in 2021 proposed a machine learning-based workflow to convert seismic data to porosity models. They designed a ResUNet + + based workflow to take three seismic data in different frequencies (i.e., decomposed seismic data) and estimate their corresponding porosity model. The workflow was successfully demonstrated in the 3D channelized reservoir to estimate the porosity model with more than 0.9 in R2 score for training and validating data (Jo et al. 2021). Researchers proposed an end-to-end convolutional neural network (CNN) regression model that automatically predicts continuous porosity at a millimeter scale resolution using two-dimensional whole core CT scan images. A CNN regression model was trained to learn from routine core analysis (RCA) porosity measurements. The linear models were outperformed by the CNN, indicating the capability of the CNN model in extracting textures that are important for porosity estimations (Chawshin et al. 2022). Tran Nguyen Thien Tam & Dinh Hoang Truong Thanh in 2023 used traditional methods in petroleum engineering and popular machine learning methods to estimate porosity and permeability via petrophysical data. Research data collected from the Volve field in Norway includes well logging and core logging data. They presented the prediction of porosity and permeability using an Artificial neural network (ANN) model, as compared to Least-squares support-vector machines (LSSVM) model and empirical model. The results

show that the ANN model could predict porosity and permeability with the highest R2 (coefficient of determination) of 0.9997 and lowest MSE (mean squared error) of 6.7769 (Tam & Thanh 2023). Researchers in 2023 carried out a comparative and statistical analysis of core-calibrated porosity with log-derived porosity for reservoir parameters estimation of the Zamzama GAS Field, Southern Indus Basin, Pakistan. They predicted porosity logs, filtered to different resolutions, using conventional and deep machine learning algorithms. Methods used include support vector regression (SVR), random forest (RF), and the multilayer perceptron (MLP) which all had accuracies above 0.96 (Munir et al. 2023). Sfidari et al. (2014) applied a machine learning approach to predict porosity, permeability and water saturation using an optimized nearest-neighbor, machine-learning and data-mining network of well-log data. The study was conducted on data from the South Pars Gas field. Jian Sun et al. in 2021, used machine learning approaches for the prediction of reservoir porosity and permeability while drilling. Their approach requires not only a high prediction accuracy but also short model processing and calculation times as new logging data are incorporated while drilling. In this paper, four machine learning algorithms were evaluated: The one-versus-rest support vector machine (OVR SVM), one-versus-one support vector machine (OVO SVM), random forest (RF) and gradient boosting decision tree (GBDT) algorithms and they showed accuracies ranging between 0.88 to 0.92. Delavar and Ramezanzadeh (2023) conducted a study titled "Pore Pressure Prediction by Empirical and Machine Learning Methods Using Conventional and Drilling Logs in Carbonate Rocks". The input models were conventional logs and drilling logs from four drilled wells in the carbonate reservoir of Asmari in the Middle East. The output model was the pore pressure of the Asmari reservoir. They used hybrid machine learning approaches including least-square support vector machine (LSSVM), artificial neural networks (ANN), and random forest (RF) approaches. The particle swarm optimization (PSO) and Bayesian method were applied to increase the accuracy of the ML procedures. The LSSVM-PSO and RF-Bayesian approaches showed the highest coefficient of determination, on average, 0.97 and 0.96, respectively, as well as the least average absolute percentage error (AAPE).

This work is conducted in response to the need for an accurate and reliable computational approach for reservoir property estimation. We use Convolutional-Recurrent Neural Networks (CRNN) hybrid deep learning model on porosity estimation in heterogeneous carbonate reservoirs. Traditional machine learning techniques rely on hand-engineered features to identify patterns in data, whereas deep learning models can automatically learn these features from raw data. The proposed method lies in its ability to learn a representation of the complex data pattern and structure that is specific to the heterogeneous carbonate reservoirs. This is achieved by capturing the spatial dependencies between features, which traditional methods cannot do with the same level of accuracy. By combining the strengths of CNNs and RNNs, CRNNs can extract more meaningful information from the input data, leading to more accurate porosity estimations. The CRNN architecture combines the power of CNNs and RNNs. CNNs are particularly good at capturing spatial features in data, while RNNs are effective at capturing temporal dependencies. The combination of these two networks allows the CRNN model to learn both spatial and temporal features simultaneously. The porosity estimation problem can be formulated as a regression problem, where the input is a well log, and the output is a core analysis porosity. In this case, the input well log can be preprocessed using CNN layers to extract relevant features, followed by RNN layers to model the spatial dependencies within the dataset. The final output can then be generated using a fully connected layer. This study was motivated by the idea that the CRNN model can be used to learn the spatial features of the log data and the temporal relationships between the log measurements. The input to the CRNN model is a sequence of log measurements, and the output is the predicted porosity value. The input data includes Gamma ray (GR), travel-time (DT), NPHI, ROHB, LLB logs, and core analysis data from a carbonate reservoir. The CRNN model was implemented using the TensorFlow platform in Python language programming.

Geology background

The raw input data set is from a carbonate reservoir in southern Iran (Fig. 1). This field was discovered in 1971 and contains two main carbonate oil formations. The reservoir formation of this field is very complex and rock compositions are dolomite, limestone, shale, anhydrite, salt, and sand (Ghazban 2007). Six wells are drilled in this field, and all the wells have been evaluated, but 2 wells (with available well-log and core analysis data) have been used due to a lack of evaluation results or weakness of interpretation.

The core data are shown in Fig. 2 while the well-log data include Gamma-ray (GR), compressional sonic travel-time (DT), bulk density (RHOB), neutron porosity (NPHI), and Deep Resistivity (LLD) are shown in Fig. 3.

For log analysis, well-logs were normalized (corrected for environmental effects) and calibrated with core data. The quality of porosity values for other wells was not satisfied and was not used. Because the results were poor, calibration was filed. Even for those two well with relatively good data, the match between core porosity and log porosity was poor due to strong heterogeneity, as shown in Fig. 4.

The porosity and grain density of the samples were determined by Ultra Porosimeter using Helium injection. This apparatus uses Boyle`s law to determine pore or grain
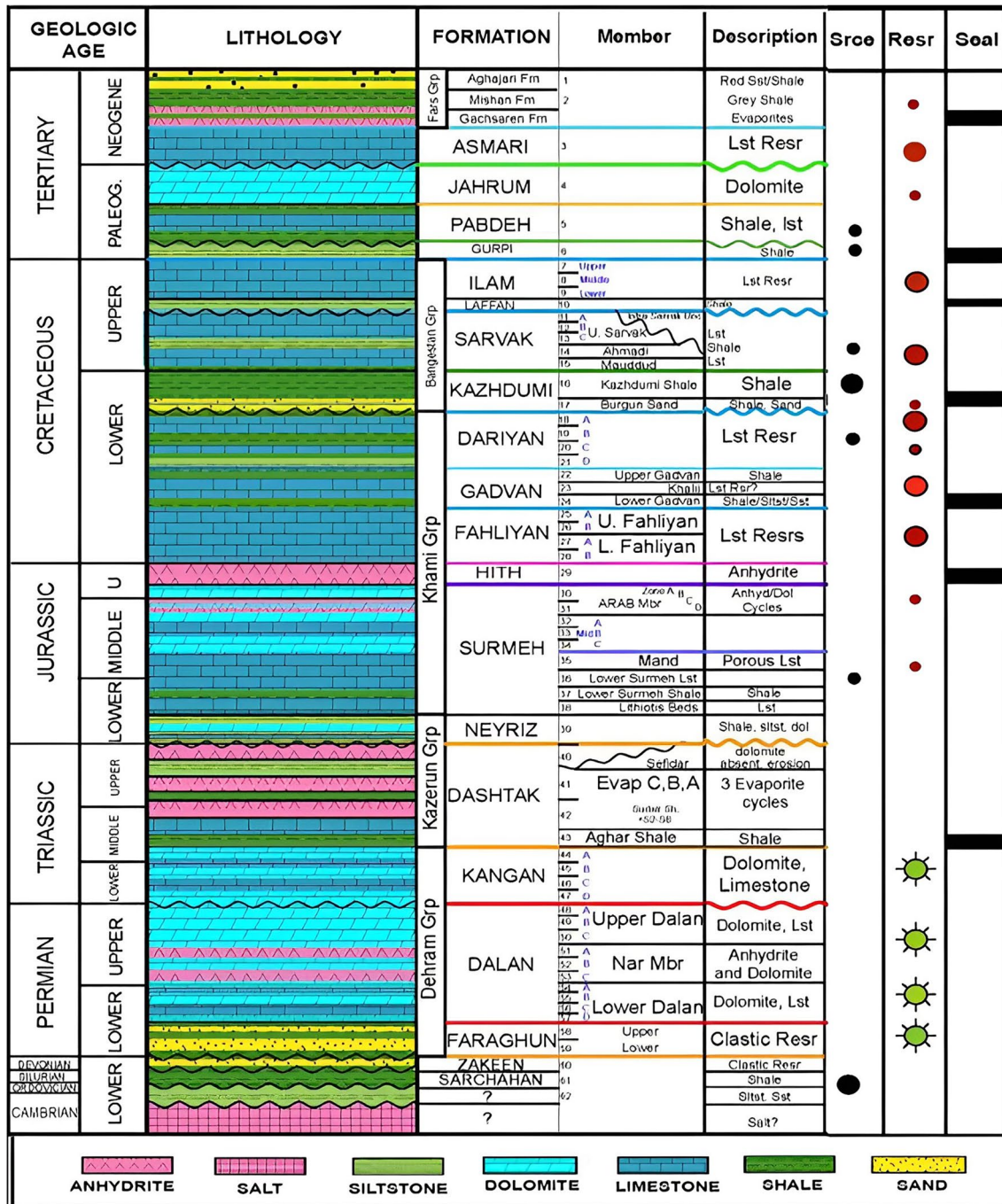
Fig. 1 General stratigraphic column in the Persian Gulf

volume from the expansion of a known mass of Helium into a calibrated sample holder.

Methodology

Hybrid deep neural network models combine the strengths of multiple types of neural networks to improve accuracy in predicting porosity from well-logs. The novel aspect of this work is a combination of CNN and RNN in which CNN is used to extract local features of the rock properties in well-logs and the RNN adaptively captures various dependencies of sequence features for regression modeling. Also, we use a Multi-Layer Perceptron (MLP) neural network as a conventional machine learning model to compare the results of the proposed hybrid deep learning model. MPL is a type of artificial neural network that consists of multiple layers of interconnected nodes, where each node represents

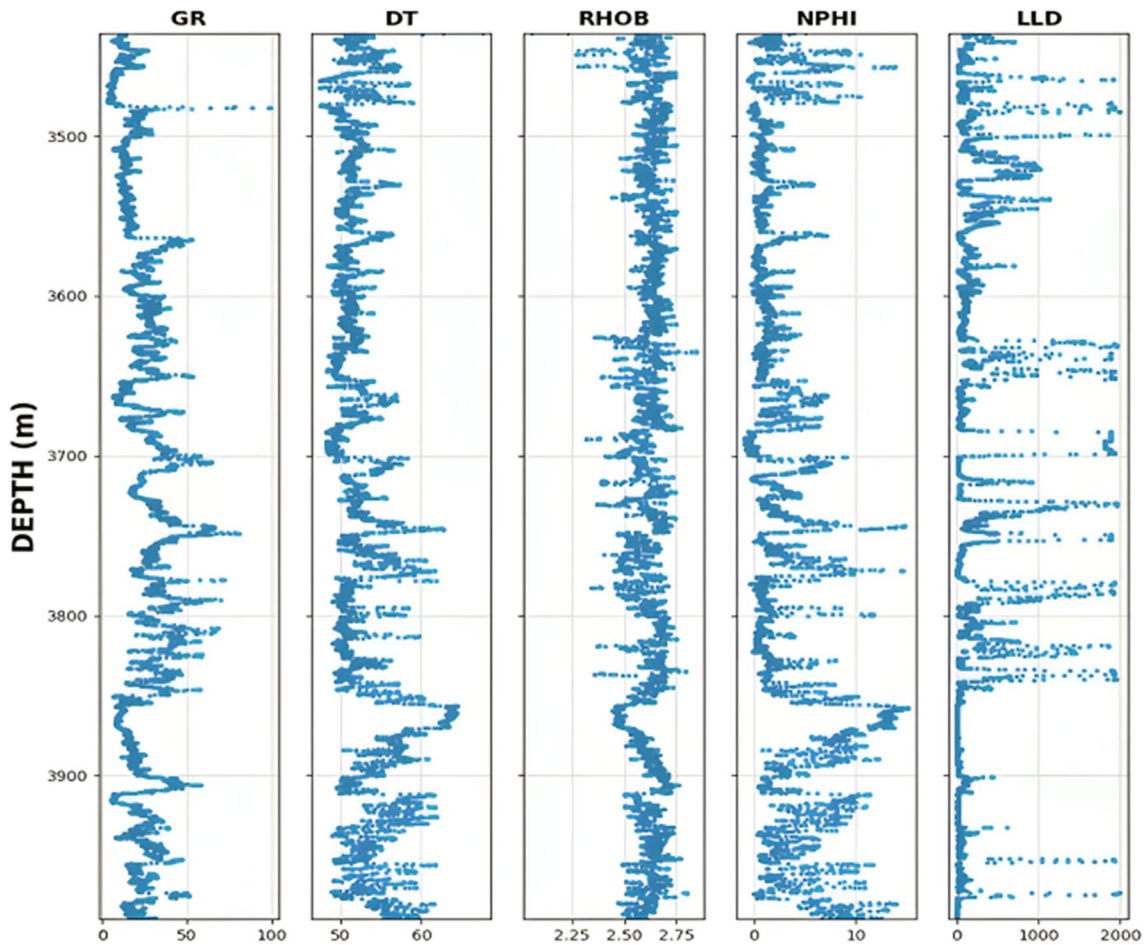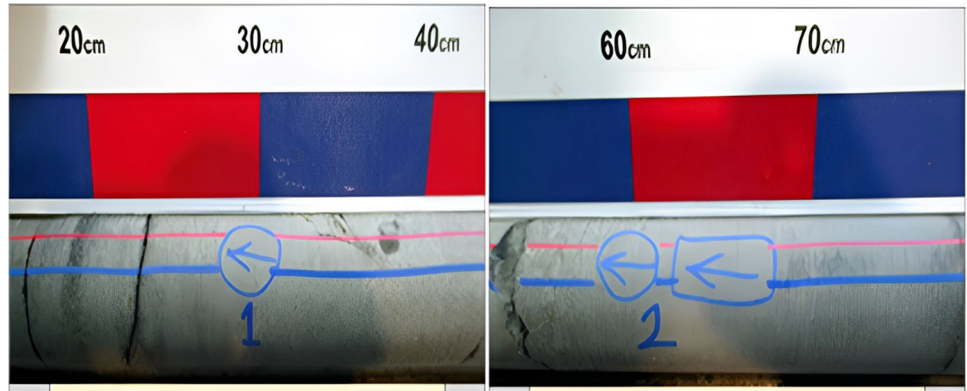**Fig. 2** Core sample data of well used in this study

Fig. 3 Well log tracks used in this study

a mathematical function. The primary objective of this research is to improve the accuracy of porosity estimation in complex carbonate reservoirs. The study specifically addresses the challenges associated with estimating porosity in heterogeneous carbonate formations. To achieve this, we evaluate the performance of a proposed Convolutional Recurrent Neural Network (CRNN) model in predicting porosity using well-log data.

The data utilized in this study comprises a combination of well-log and core analysis data collected from an Iranian carbonate oil field. The well-log data includes essential measurements such as Gamma Ray (GR), Sonic Transit
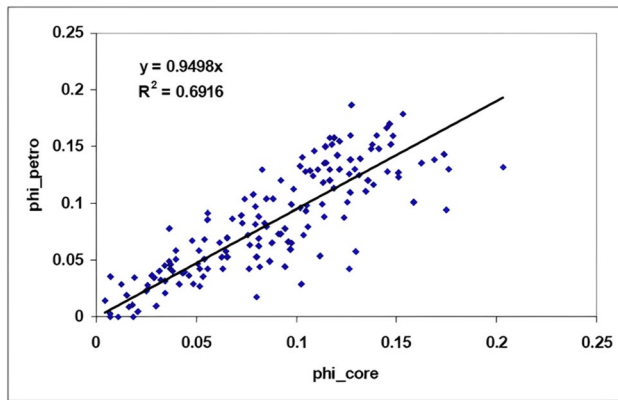
Fig. 4 Core porosity versus well-log porosity

Time (DT), Bulk Density (RHOB), Deep Resistivity (LLD), and Neutron Porosity (NPHI). Prior to modeling, the data undergoes preprocessing to handle missing values, outliers, and ensure overall consistency. The workflow of the research has been explained in the coming paragraph.

A hybrid CRNN model is developed by integrating Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architectures. The CNN component is designed to extract spatial features from the well-log data, capturing patterns across different depths, while the RNN component captures temporal dependencies by considering the sequential nature of well logs. For model training and validation, the dataset is partitioned into training and validation sets. The CRNN model is trained using the well-log data as input features, and core data is employed for model validation to assess performance. The CRNN model's accuracy and generalization capabilities are then compared to those of a traditional Multi-Layer Perceptron (MLP) model. The performance evaluation involves calculating the correlation coefficient between the model predictions and core data. The CRNN model demonstrates a significant improvement with a correlation coefficient of 0.98, compared to the 0.67 achieved by the MLP model. This substantial improvement illustrates the superior performance of the CRNN model. The CRNN model effectively combines spatial and temporal patterns from the well-log data, capturing the complex dependencies present within heterogeneous carbonate reservoirs. This integration leads to more accurate porosity estimations, highlighting the model's capability to address the complexities inherent in such geological formations. This research contributes novel insights to the existing petroleum literature by presenting a unique approach to porosity estimation. The hybrid CRNN model offers higher accuracy rates and improved generalization capabilities by integrating CNN and RNN methodologies. This approach not only enhances reservoir characterization but also improves

decision-making processes in the petroleum industry by providing more precise and reliable porosity predictions.

Convolutional neural networks (CNNs)

CNNs are a subset of deep learning models that have revolutionized computer vision and image recognition tasks. CNNs are inspired by the structure and function of the visual cortex in biological organisms, specifically the way individual neurons process small, overlapping regions of visual stimuli. The basic building block of a CNN is the convolutional layer (Khan et al. 2020). A convolutional layer applies a set of filters to the input data, resulting in a set of feature maps that highlight different aspects of the input. Each filter is a small matrix of weights that is convolved with the input data, producing an output value for each position in the resulting feature map. The weights in these filters are learned during training by optimizing a loss function that measures the discrepancy between the output of the model and the desired output. After one or more convolutional layers, the output is typically fed through a pooling layer, which reduces the dimensionality of the output while preserving important features. Pooling can be done using various methods, such as max pooling, average pooling, or L2-norm pooling. The reduced output from this stage is then passed through one or more fully connected layers, where the final classification decision is made (Ye and Wang 2023).

One of the most significant advantages of CNNs is their ability to automatically learn features from raw input data. In traditional machine learning models, engineers must manually select and extract relevant features from the data before feeding it to the model. With CNNs, however, the convolutional layers automatically learn salient features from the input, allowing the model to adapt to new types of data without requiring extensive pre-processing (Alzubaidi et al. 2021, Khan et al. 2020).

1D-CNN models are developed to analyze sequences of numerical data, in which the 1D-CNN model applies a series of convolutional filters to the input data, to extract key features and patterns from the sequence (Alzubaidi et al. 2021, Kaydani et al 2012). A typical 1D-CNN architecture for 1-dimensional data comprises 1-dimensional convolution layers, pooling layers, dropout layers, and activation functions.

The sequence of 1D convolutions (weighted sums of two 1D arrays) is a major operation in 1D-CNN. 1D propagation can be described by:

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} Conv1D(w_{ik}^{l-1}, s_i^{l-1})$$

where:

x_k^l input

x_k^l bias of the kth neuron at layer l

s_i^{l-1} output of the ith neuron at layer l-1

w_{ik}^{l-1} kernel from the ith neuron at layer l-1 to the kth neuron at layer l

Conv1D(.,.) is 1D convolution.

The forward propagation process involves passing input data through a series of convolutional layers, pooling layers, and activation functions. These layers form a hierarchy of features that are learned by the network. The output of the final convolutional layer is typically flattened into a vector and passed through one or more fully connected layers, which performs a non-linear transformation on the features learned from the convolutional layers. The final output of the network is usually a probability distribution over the class labels of the input data. During back-propagation, the error between the predicted output and the true label is propagated backward through the network, allowing the weights and biases of the network to be updated to minimize the error. This process involves computing the gradient of the loss function for each weight in the network, using the chain rule to propagate the error back through each layer (Alzubaidi et al. 2021, Ye and Wang 2023).

Figure 5 illustrates how Forward and Back-propagation works in a CNN model to adjust weights and bias.

These filters slide over the input data in a one-dimensional manner, capturing local interactions between adjacent data points and using them to build up higher-level representations of the overall pattern. The resulting feature maps are then processed further through additional layers of convolution, pooling, and other types of nonlinear transformations, before being fed into a final output layer for regression.

Recurrent neural network (RNN)

RNNs are a class of deep learning models that have shown great success in tasks such as prediction and regression. RNNs are designed to handle sequential data by maintaining a memory of past inputs. At their core, RNNs are neural networks that have loops in them, allowing information to persist over time. This loop enables the network to maintain an internal state or memory that can capture long-term dependencies in the input sequence (Gharehbaghi 2023).

The basic building block of an RNN is a recurrent neuron, which takes as input the current input and the output of the previous neuron in the sequence. This allows the network to maintain information about past inputs and use it to predict future inputs. Training an RNN involves optimizing the network weights to minimize the error between the predicted output and the actual output. This is done using backpropagation through time, a variant of the standard backpropagation algorithm that considers the temporal structure of the data. The critical advantage of RNNs over other types of feedforward neural networks is their ability to handle varying-length sequences, making them highly adaptable to a wide range of applications (Gharehbaghi 2023, Ye and Wang 2023).

Fig. 5 Functioning of Forward and Back-propagation within a CNN model
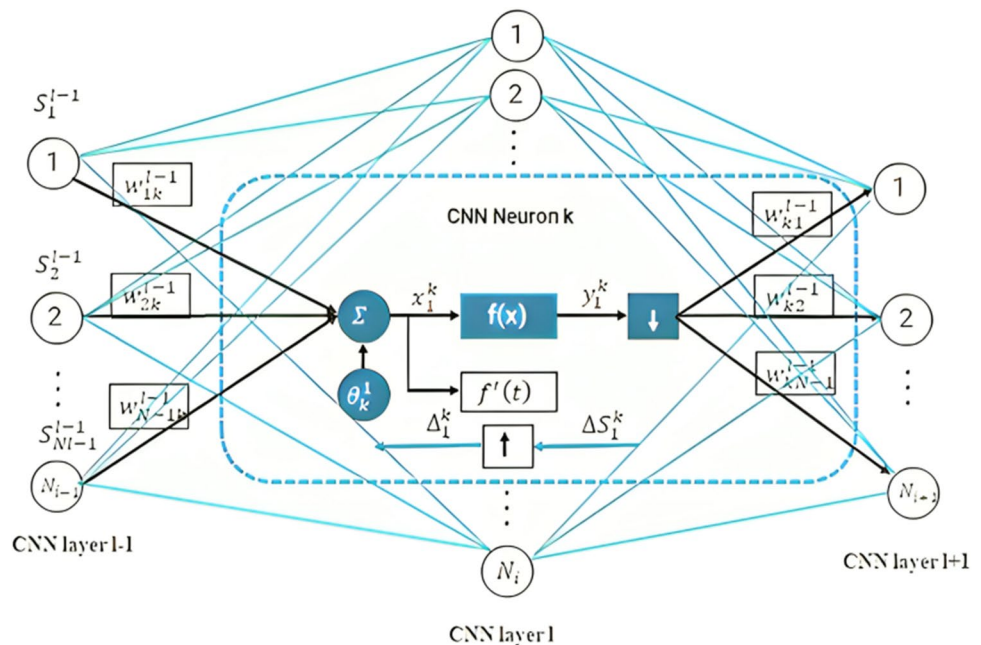
LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture that is commonly used in deep learning for sequence modeling and prediction tasks. Unlike traditional RNNs, LSTM networks have an internal memory state that can selectively retain or discard information over time, allowing them to process longer sequences more effectively (Gharehbaghi 2023).

LSTM can be mathematically expressed as follow (Iosifidis and Tefas 2022):

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

$$c_t\prime = tanh(W_{hc}h_{t-1} + W_{xc}x_t + b_c)$$

$$c_t = f_t c_{t-1} + i_t c_t\prime$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_0)$$

$$h_t = o_t tanh(c_f)$$

where f_t and i_t and o_t are activations at time step t for input, forget, and output gates. c_t is the protected cell activation at time t and h_t is activation to the next layer. W is the weight matrix and b is the bias vector (Iosifidis and Tefas 2022). Figure 6 depicts the Block diagram of the LSTM recurrent neural network cell unit.

In an LSTM model, the input sequence is fed into a chain of memory cells, each of which contains three gates: an input gate, an output gate, and a forget gate. The input gate controls how much of the input should be added to the memory cell, while the forget gate determines what information should be discarded from the cell's current state. Finally, the output gate controls how much of the cell's information should be outputted to the next layer of the network (Ye and Wang 2023, Gharehbaghi 2023).

Convolutional-recurrent neural networks (CRNNs) model

CRNNs are a type of neural network that combines the strengths of convolutional neural networks and recurrent neural networks. The idea behind this approach is to use CNNs for feature extraction from raw well-log data, followed by RNNs for capturing temporal dependencies in the data. This allows the model to learn complex patterns in the data and make accurate predictions (Gharehbaghi 2023).

To use a CRNN model for porosity prediction from well logs, we follow these general steps:

1. Data preprocessing: Preprocess the well-log data by cleaning, scaling, and normalizing it. Then split the data into training, validation, and testing sets.
2. Feature extraction: Extract features from the well-log data using convolutional layers. The output of the convolutional layers will be a set of feature maps that capture different aspects of the input data.
3. Sequence modeling: Feed the feature maps into a recurrent neural network to capture the sequential patterns in the data. The output of the recurrent layers will be a set of hidden states that represent the learned representations of the input sequence.
4. Prediction: Use a fully connected layer to predict the porosity value based on the final hidden state of the recurrent layers.
5. Model training and evaluation: Train the CRNN model on the training set and evaluate its performance on the validation set. Iterate on the model architecture and hyperparameters until satisfactory performance is achieved. Finally, test the model on the testing set to obtain a final measure of its performance.
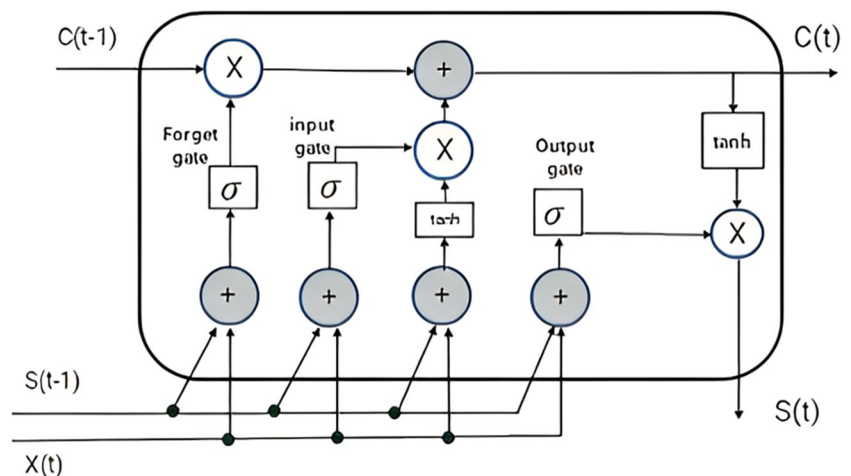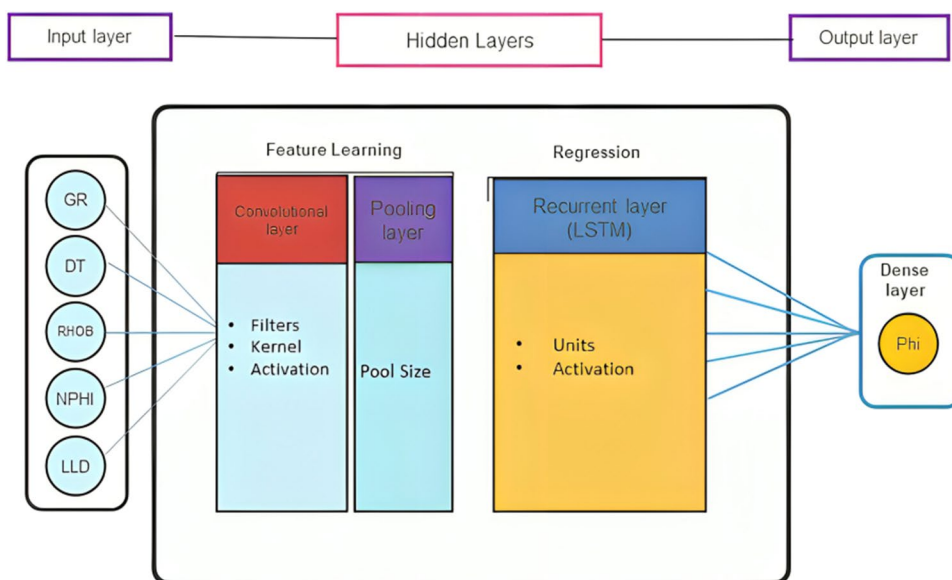
Fig. 6 Block diagram of the LSTM cell unit

Fig. 7 Architecture of Convolutional-Recurrent Neural Networks (CRNNs) model

The additional considerations specific to the dataset and problem, such as choosing appropriate loss functions, optimizing model hyperparameters, and dealing with missing or noisy data. The schematic of the proposed CRNN model is shown in Fig. 7.

The performance of the model is evaluated to measure how well the model can predict continuous numerical values based on input data. Here we use metrics such as mean squared error (MSE) and coefficient of determination (R-squared) to measure how well the model can predict the output values (Yousefmarzi et al. 2024). MSE was used to evaluate the performance of the models. It measures the average squared difference between the predicted and actual values. MSE defined as{Ye and Wang 2023#14}:

$$MSE = \frac{1}{N} \sum_{j=1}^{N} \left| y_j - \widetilde{y}_j \right|$$

The R^2 Coefficient of Determination, also known as the R-squared value, is a statistical measure used to evaluate the goodness of fit of a regression model. It indicates how well the regression line fits the data points in a scatter plot (Gharehbaghi 2023).

$$R^2 = 1 - \frac{sum\ squared\ regression(SSR)}{total\ sum\ of\ squares(SST)}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$\hat{y}_1, \hat{y}_2, \dots \hat{y}_n$ *are predicted values*
$y_1, y_2, \dots y_n$ *are observed values*
n *is the number of observations*

Multilayer perceptron (MLP) model

The Multilayer Perceptron (MLP) is a sophisticated type of artificial neural network (ANN) designed to model complex patterns in data. Characterized as a feedforward neural network, the MLP facilitates unidirectional flow of information—from input to output—without recursive loops or feedback mechanisms. The architecture of an MLP includes an input layer, one or more hidden layers, and an output layer. Each layer comprises multiple neurons, also referred to as nodes or units. Neurons in one layer are fully connected to every neuron in the subsequent layer, forming a dense network structure. The neurons within these layers utilize activation functions to introduce non-linearity, enhancing the model's capacity to capture complex relationships. Common activation functions employed are sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU). MLPs are trained using supervised learning techniques. The training process involves iterative adjustment of weights and biases connecting the neurons to minimize the error between predicted and actual values on a labeled dataset. Backpropagation, a widely used optimization algorithm, facilitates these weight updates during training.

In the context of reservoir porosity prediction, MLPs exhibit strong predictive capabilities. By learning from various input data such as well logs (resistivity, acoustic imaging), fracture parameters (porosity, density, length, width),

Table 1 Statistical ranges and parameters of Log data related to inputs/outputs employed for developing models

NO	Parameter (Units)	Count	Mean	Std	Min	25%	50%	75%	Max
1	GR (API units)	10872.0	23.396	12.381	2.2812	13.821	22.673	31.056	101.96
2	NPHI (%)	10872.0	0.1247	0.1391	-0.057	-0.031	0.1466	0.2580	0.4490
3	RHOB (g/cm3)	10872.0	2.5305	0.2859	1.6030	2.3043	2.4938	2.8084	3.1580
4	DT (μs/ft)	10872.0	75.761	17.188	39.717	63.950	72.855	86.595	145.45
5	LLD (Ω·m)	10872.0	92.030	496.72	0.000	1.0160	2.2450	33.798	2000
6	PHIE (%)	10872.0	0.0869	0.0806	0.000	0.000	0.0697	0.1522	0.3713

Table 2 Statistical ranges and parameters of Core data related to inputs/outputs employed for developing models

NO	Parameter	Count	Mean	Std	Min	25%	50%	75%	Max
1	Horizontal Permeability (mD)	236.00	54.818	131.19	0.0100	0.6000	6.200	39.300	1335.0
2	Coreing Depth (ft)	236.00	7504.6	1833.1	3933.0	6236.2	6814.0	9446.7	9516.5
3	Porosity %	236.00	15.862	7.8324	1.2900	10.225	16.500	21.685	44.690

Table 3 Performance comparison of MLP and CRNN model

Calculated parameter	MSE	R-Square
MLP model	0.00039	0.67
CRNN model	0.01342	0.98

and lithological information, MLPs can effectively capture the intricate relationships between these features and the target porosity values. The process begins with data preparation, constructing a dataset that includes fracture-related parameters and logging responses. During model training, the MLP learns from this dataset, understanding the complex relationships between fracture parameters and logging responses. Once trained, the MLP can predict porosity in reservoirs based on new input data. The model's accuracy is subsequently evaluated using regression metrics such as the correlation coefficient (R^2). Application of the MLP model aids in predicting key exploration horizons before drilling, thus supporting reservoir characterization.

Data set and input parameters

Data used for model developments and the Porosity range in both log and core are summarized in Table 1 and Table 2. All input parameters and other parameters of the dataset as shown in Table 2, which include GR, NPHI, RHOB, DRHO, DT and LLD along with PHIE(Porosity) as output, are used in the development of the models for prediction of porosity. On the other side, all input parameters as shown in Table 3 which are Horizontal Permeability, Coring Depth and Porosity, are used in development of models in order to predict porosity from core data. The depth range differences between Table 2 and Fig. 8 are due to the aggregation of well log data over different intervals. Table 2 shows specific core sample depths, while Fig. 8 represents a broader range from the entire well

log dataset. All models finally performed alongside each other as a part of the hybrid model which is going to be discussed further. The training data is loaded in the model and displayed logs in the plot as shown in Fig. 8. This figure illustrates the well log data post-preprocessing. The resulting data demonstrates improved consistency and quality, making it suitable for accurate model input. Figure 8 demonstrates the well-log data after these preprocessing steps, highlighting the enhanced clarity and uniformity achieved, which is crucial for the accurate prediction of porosity in carbonate reservoirs.

Data preprocessing

Before training the neural network model, the well-log data must be preprocessed to remove noise and data gaps and to normalize the data. Outlier detection and removal is one of the important steps in this process, especially when working with well log data where noise and bias are present. Noise in well log data can be caused by various factors such as measurement errors, environmental interference, or equipment failure. Knowledge of the domain-specific properties of well log data was used to identify outliers. This approach involves understanding the physical characteristics of the well and using this information to identify data points that do not conform to the expected behavior. Then, the identified outliers were removed from the dataset.

We used the principal components analysis (PCA) technique in multivariate analysis. PCA was employed to reduce the dimensionality of the input feature set, which helps in minimizing redundancy and capturing the most significant variance in the data. This preprocessing step is crucial for improving the efficiency of the CRNN model by focusing on the most informative features and reducing overfitting risks.

To explore the relationship between multiple independent variables and multiple dependent variables, multivariate analysis was performed, as shown in Fig. 9.

Fig. 8 Well-log track used in this study
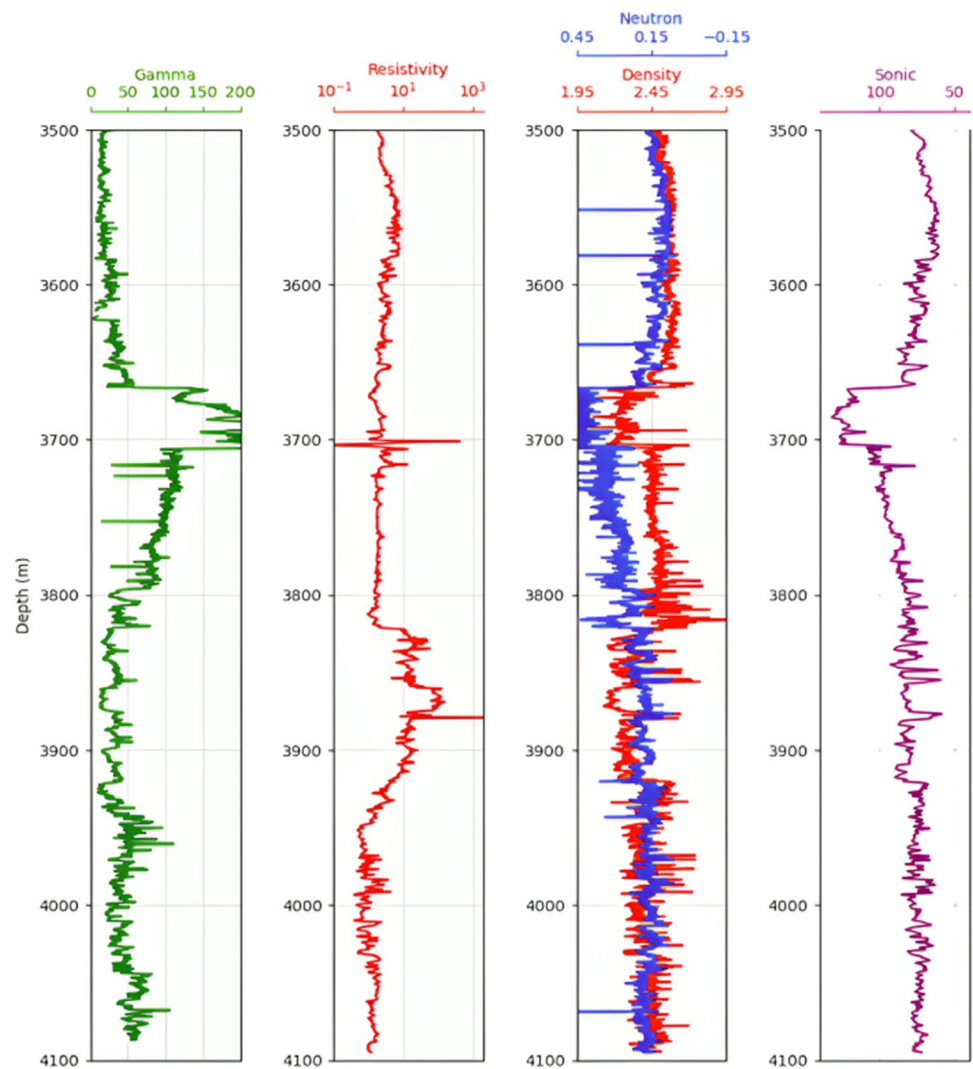
In this analysis, two types of relationships including positive correlation and negative correlation can be identified.

The frequency distributions of data are graphically represented by a histogram plot in Fig. 10.

After deleting the missing value in the data set, the training data distribution was checked as shown in Fig. 11.

Normalization is necessary because the range of values for each log varies widely, and normalization ensures that each variable contributes equally to the model. The dataset is randomly divided into training, test and validation sets.

Training the neural network model

The neural network model's training process utilizes backpropagation—a dynamic algorithm that adjusts the weights between neurons to minimize the error between predicted and actual porosity values. Optimal model performance necessitates careful tuning of hyperparameters, which critically influence the learning process. For the Convolutional Neural Network (CNN) component, several hyperparameters are meticulously fine-tuned. This includes varying the number of CNN layers, typically between 2 to 4, to assess their impact on feature extraction. The number of neurons per layer, ranging from 32 to 128, is adjusted to balance model complexity and efficiency. Filter sizes (e.g., 3×3, 5×5) are evaluated to capture relevant spatial patterns, and subsampling factors (stride for down sampling, e.g., 2×2) are determined to prevent overfitting.

Similarly, for the Long Short-Term Memory (LSTM) component, hyperparameters are carefully configured. The number of hidden layers is explored, with options between single or double layers, to effectively capture temporal dependencies. The number of LSTM units, typically 64 or 128, is varied to balance expressiveness and computational cost. Other critical parameters include the learning rate (e.g., 0.001, 0.01) for efficient convergence, batch sizes (e.g., 16, 32) for training stability, dropout rates (e.g., 0.2, 0.5) for regularization, and the number of epochs to ensure robust
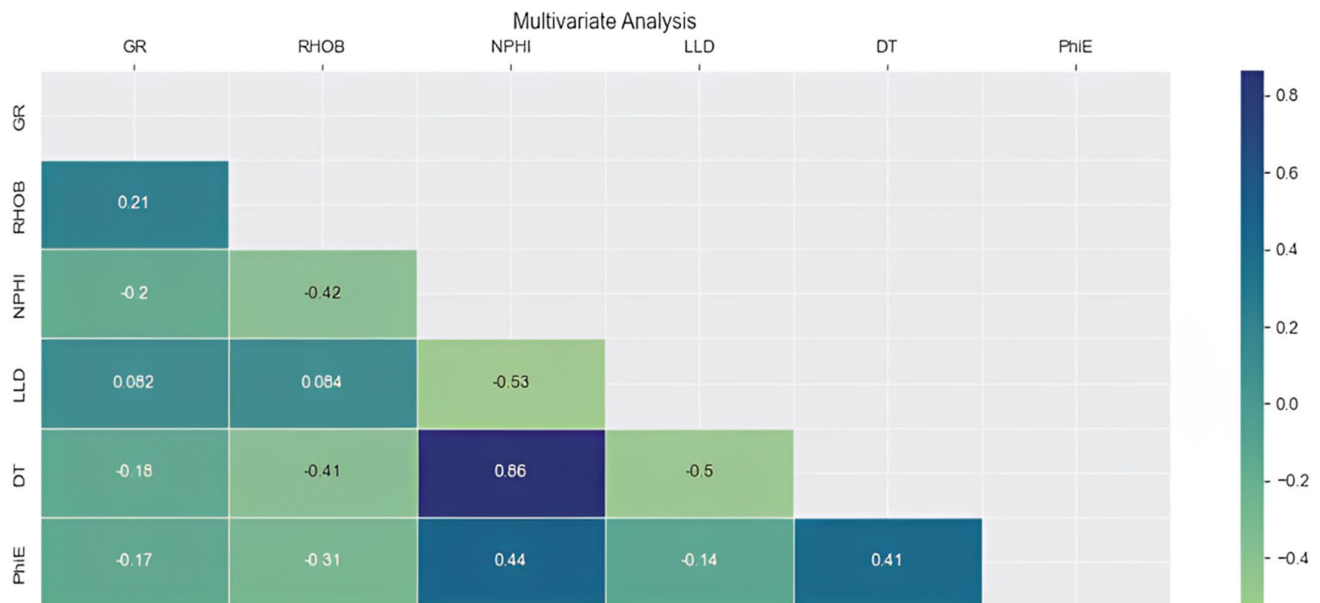
Fig. 9 Multivariate analysis plot

Experiment results

The data-driven model proposed in this work has rock porosity as the output, whereas the input combines the well-log data. A CRNNs regression model that combines the strengths of CNNs and RNNs was used for the sequence data analysis task to estimate the porosity of carbonate rock from well-log and core data.

Once the dataset was prepared, the model was built with the TensorFlow platform and trained on the dataset by Adam Optimizer. Then, a validation set was used to evaluate the model's performance. Moreover, the hyperparameters were adjusted using a Keras Tuner.

We first used the MPL neural network to train a dataset of well-log data and corresponding porosity values to learn the relationship between these variables. The input layer of the network receives the well log data as inputs, which include measurements of gamma-ray, sonic, neutron, resistivity, and density logs data. The output layer of the network produces a predicted porosity value. During the training of this model, the weights of the connections between nodes are adjusted to minimize the difference between the predicted porosity values and the actual porosity values in the training dataset. This process is performed using backpropagation, a method for iteratively adjusting the weights based on the error between predicted and actual outputs. Once trained, the MPL neural network was used to predict the porosity of new well-log data that it has not seen before. The performance of the MLP model is shown in Fig. 12.

The model was trained on the training dataset, and it was evaluated based on the validation data set. The results are poor as can be seen in Fig. 13 for the correlation coefficient between the model and core data.

Although the model learned the training data too well, it cannot generalize to new data. This revealed that the MLP model is too simple to capture the underlying structures in the data of this heterogenous carbonate rock. The performance evaluation of this model shows MSE is 0.0039 and the correction coefficient is 0.67.

To implement the CRNN model for porosity prediction from well logs, we first preprocess the well log data by normalizing it and dividing it into sequences of equal length. We then split the data into training, testing and validation sets. Next, we define the CRNN architecture using TensorFlow and Keras. The architecture consists of a 1D convolutional layer followed by a max pooling layer, an LSTM layer, a dropout layer to prevent overfitting, a time-distributed dense layer with a sigmoid activation function, and a flatten layer followed by a dense layer with a linear activation function. We compile the model using the Adam optimizer and mean squared error as the loss function, and train it on the training set for a specified number of epochs.
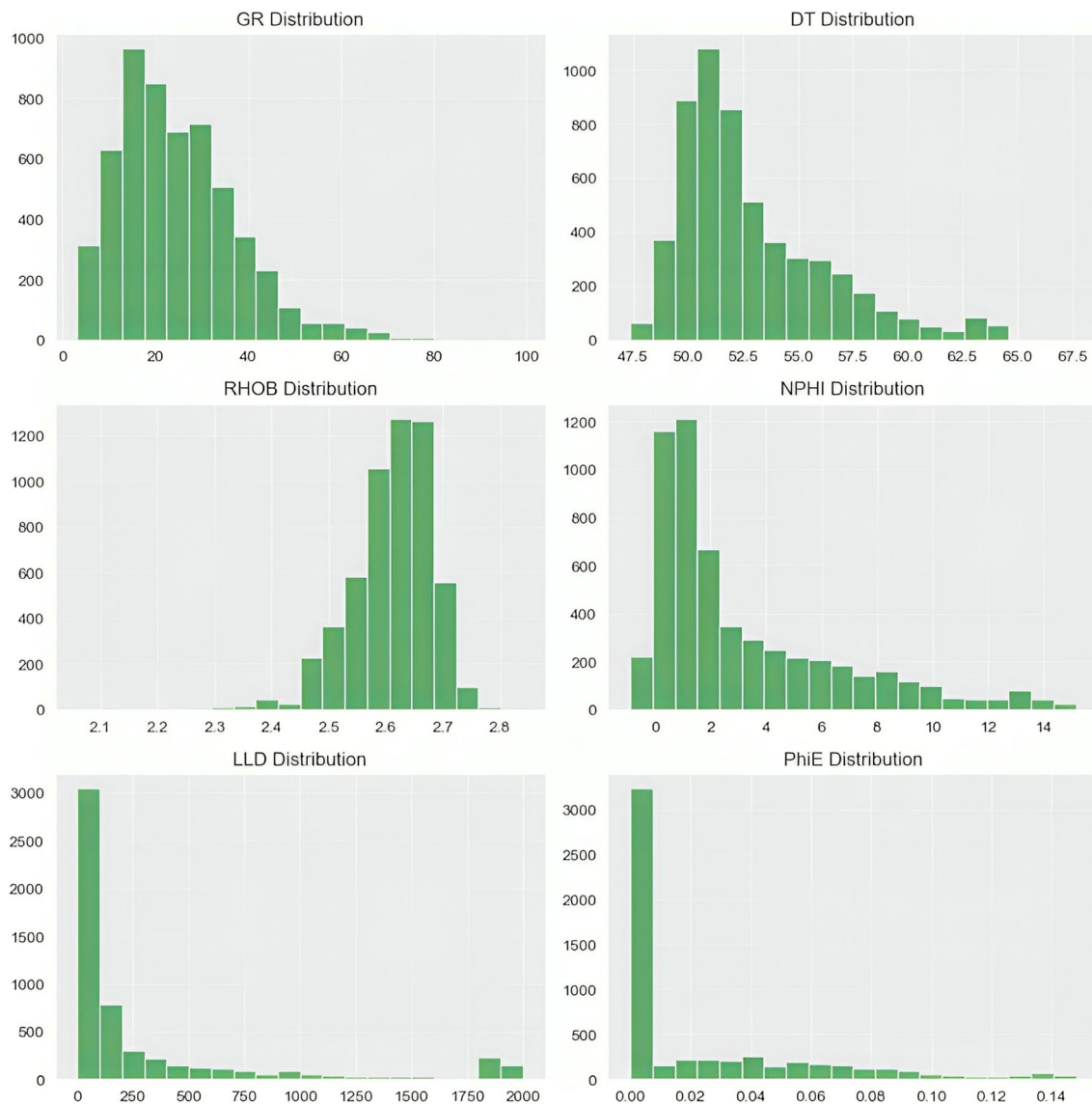
learning while balancing training time and convergence. To optimize these hyperparameters, the Keras Tuner library is employed, which provides various search algorithms, including Random Search and Bayesian Optimization. The primary goal of this process is to maximize model accuracy while avoiding overfitting, resulting in a robust hybrid CRNN model for accurate porosity prediction in complex carbonate reservoirs.

Fig. 10 Frequency distributions of data

We evaluate the performance of the model on the testing set using mean squared error as the evaluation metric. Figure 14 shows the performance of the CRNN model for the training and validation dataset.

As can be seen, A good fit is identified by a training and validation loss that decreases to a point of stability with a minimal gap between the two final loss values.

The coefficient of determination (R-squared) to measure how well the model can predict the output values for CRNN is shown in Fig. 15.

The correlation coefficient between the CRNN model and the core data obtained was 0.98, which shows excellent performance.

The models were tuned by hyperparameters configuration using Keras Tuner. The optimal set of hyperparameters was used to train the final model on the entire dataset. By using the parameter tuning method with Keras Tuner, the performance of models was improved and better accuracy on task was achieved. Table 3 presents the comparison of the performance of both tuned models.

The results showed that the CRNN model had a MAS of 0.01342, while the MLP model had an average MSE of 0.00039. This indicates that the CRNN model was able to predict porosity with higher accuracy than the MLP model.
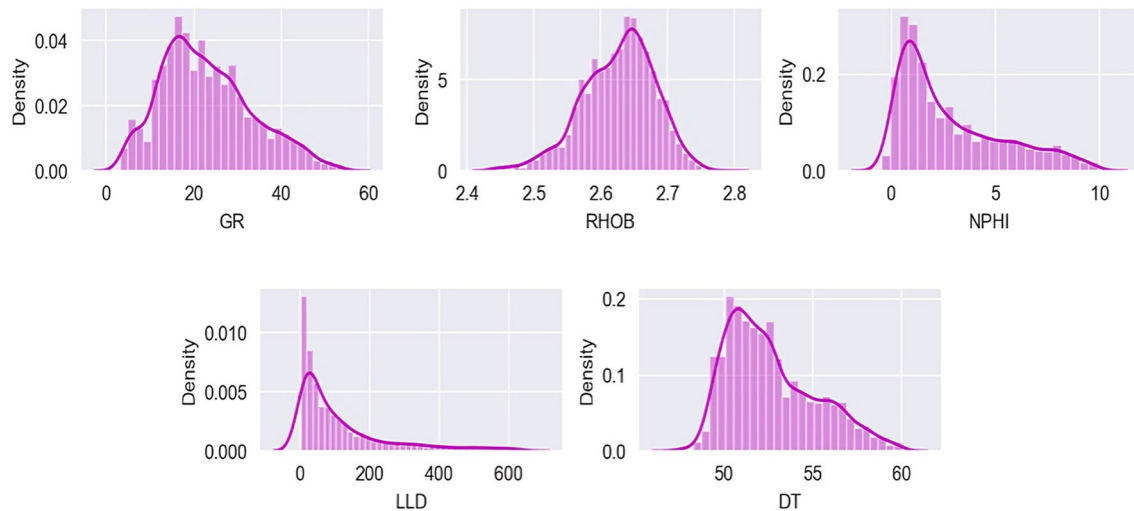
Fig. 11 Distribution of the training data

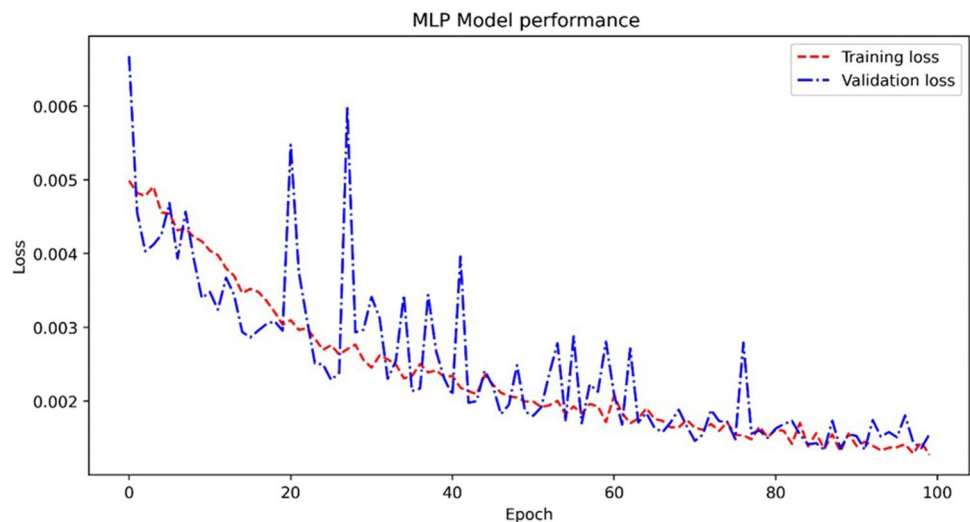**Fig. 12** Learning performance of the MLP model

Figure 16 depicts predicted porosity using a CRNN model correlated with core data (blue point).

It is observed that the CRNN model outperforms MLP algorithms for porosity prediction from well logs. The CRNN model can capture the complex spatial–temporal patterns in the log data and achieve better accuracy compared to other models.

It is clear that the CRNN model was efficiently able to capture the spatial dependencies between different logs using convolutional layers, and the temporal dependencies within each log using recurrent layers. This allowed the model to effectively learn the complex relationships between the input data and the target variable (porosity). Overall, CRNNs are a novel approach for porosity estimation in heterogeneous carbonate reservoirs. By combining the strengths of CNNs and RNNs, CRNNs can capture the complex spatial dependencies within the reservoir and learn a representation that is specific to the reservoir being analyzed. This approach has the potential to improve the accuracy of porosity estimation and provide valuable insights into reservoir characterization.

Conclusion

The study presents a comprehensive analysis of porosity estimation in complex carbonate reservoirs using a hybrid Convolutional Recurrent Neural Network (CRNN) deep learning model. Key findings highlight the superior performance of the CRNN model over traditional Multilayer Perceptron (MLP) models in predicting porosity from well logs. The CRNN model adeptly captures complex spatial
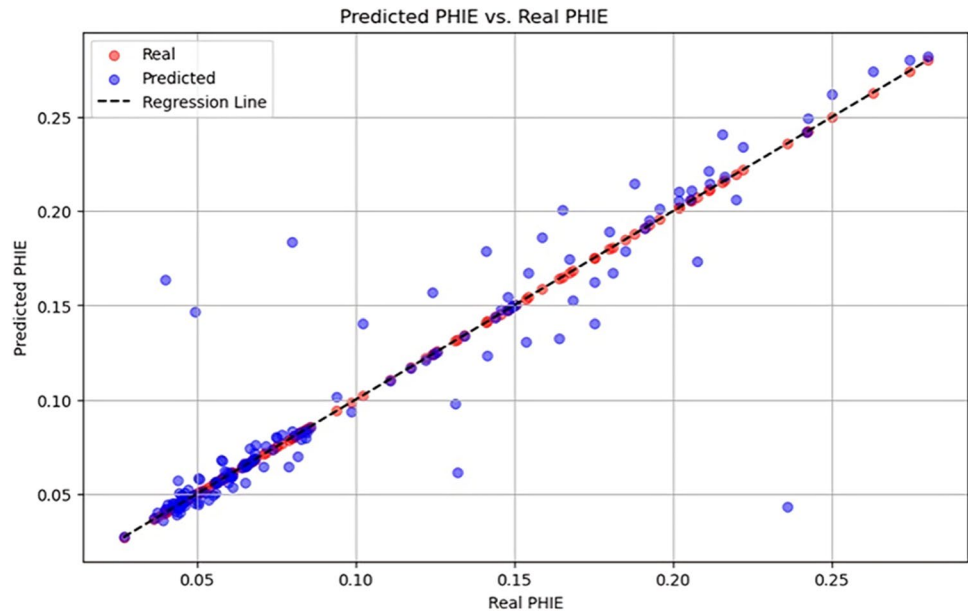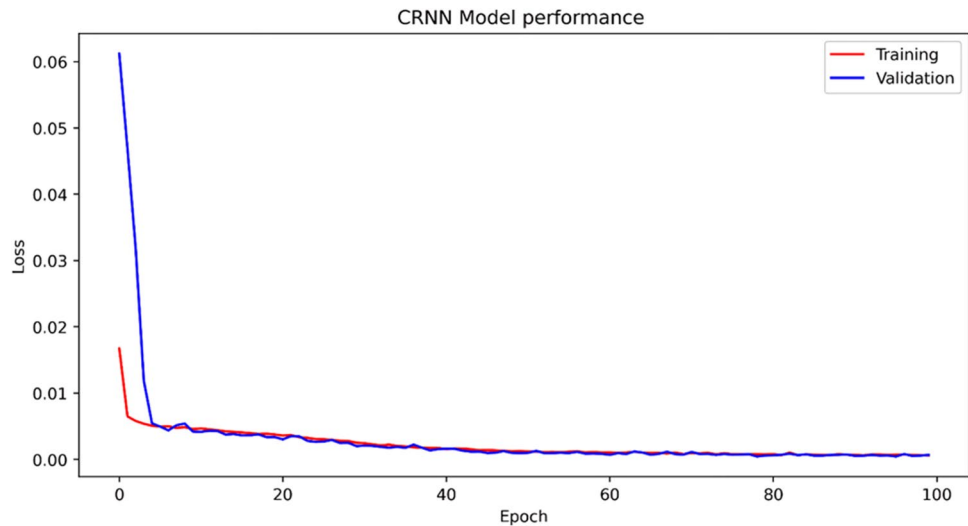
Fig. 13 R squared plot for MLP model

Fig. 14 Performance CRNN model

and temporal dependencies within the reservoirs, resulting in a significant improvement in correlation coefficients between model predictions and core data (0.98 for CRNN versus 0.67 for MLP). This ability to extract relevant features and model sequential data underscores the CRNN model's efficacy as an accurate tool for porosity estimation in heterogeneous carbonate reservoirs.

The research further illustrates that the CRNN model not only achieves higher accuracy rates but also operates with fewer model parameters compared to conventional models. This efficiency demonstrates the model's capacity to handle complex predictions of reservoir porosity, highlighting the importance of considering both spatial and temporal patterns in well-log data for accurate porosity prediction.

Despite the promising results, the research acknowledges certain limitations. The model's performance evaluation is based on data from an Iranian carbonate oil field, which may possess unique characteristics. Therefore, further validation using diverse datasets from various carbonate reservoirs worldwide is necessary to ensure the model's robustness and generalizability. Future research should also explore the integration of additional geological and petrophysical data sources to enhance the model's accuracy and broaden its applicability. Incorporating uncertainty quantification techniques and sensitivity analysis could provide valuable insights into the reliability of porosity predictions and identify critical input parameters.

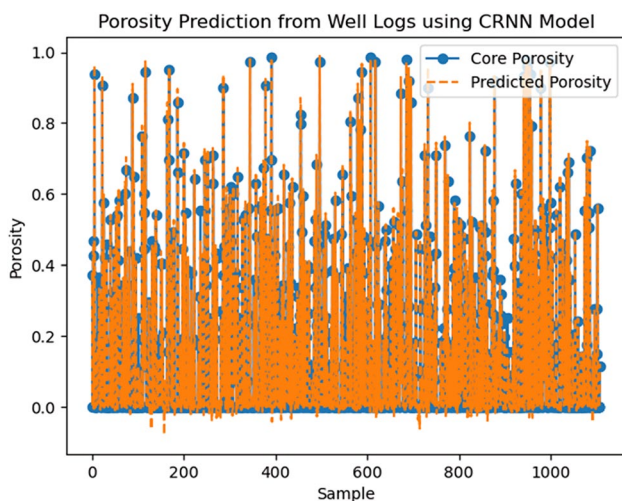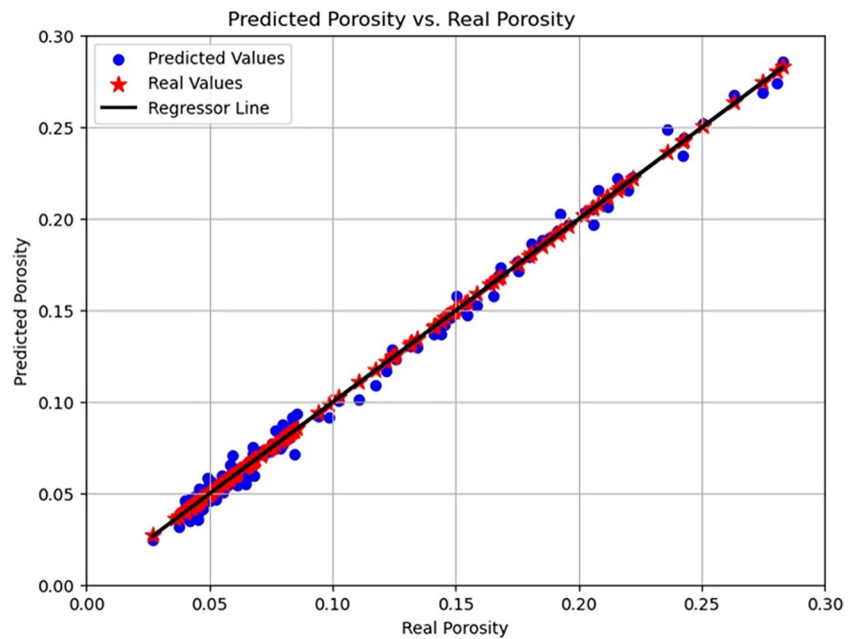**Fig. 15** R squared for CRNN
model

Fig. 16 Prediction porosity using CRNN model. This figure displays predicted versus actual porosity values. The total number of data points exceeds the core data count (236) because it includes interpolated well log data points, used to enhance the model's training dataset and improve prediction accuracy

In conclusion, this study introduces the hybrid CRNN deep learning model as a promising approach for accurately estimating porosity in complex carbonate reservoirs. By addressing the identified limitations and exploring suggested future research directions, the model's reliability can be further enhanced, its applicability expanded, and our understanding of porosity estimation in heterogeneous carbonate reservoirs advanced.

Declarations

Competing interests The authors declare no competing interests.

References

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

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L (2021) Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data 8(1):53

Bagheri M, Riahi MA (2015) Seismic facies analysis from well logs based on supervised classification scheme with different machine learning techniques. Arab J Geosci 8:7153–7161. https://doi.org/10.1007/s12517-014-1691-5

Bagheri M, Rezaei H (2019) Reservoir rock permeability prediction using SVR based on radial basis function kernel. Carbonates Evaporites 34:699–707. https://doi.org/10.1007/s13146-019-00493-4

Bagrintseva, K. I. (2015). Carbonate Reservoir Rocks. Scrivener Publishing LLC.

Bust VK, Oletu JU, Worthington PF (2011) The challenges for carbonate petrophysics in petroleum resource estimation. SPE Reservoir Eval Eng 14(01):25–34

Cao J, Yang J, Wang Y (2015) Extreme learning machine for reservoir parameter estimation in heterogeneous sandstone reservoir. Mathematical Problems in Engineering 2015:1–10. Hindawi

Chawshin K, Berg CF, Varagnolo D (2022) Automated porosity estimation using CT-scans of extracted core data. Comput Geosci 26:595–612. https://doi.org/10.1007/s10596-022-10143-9

Chen L, Lin W, Chen P, Jiang S, Liu L, Hu H (2021) Porosity prediction from well logs using back propagation neural network optimized by genetic algorithm in one heterogeneous oil reservoirs of Ordos Basin, China. J Earth Sci 32(4):828–838

Chen W, Yang L, Zha B, Zhang M, Chen Y (2020) Deep learning reservoir porosity prediction based on multilayer long short-term memory network. Geophysics 85(4):WA213–WA225

Cui C, Fearn T (2018) Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. Chemom Intell Lab Syst 182:9–20

Darling T (2005) Well logging and formation evaluation. Gulf Professional Publishing, Houston, Texas. https://doi.org/10.1016/B978-0-7506-7883-4.X5000-1

Delavar MR, Ramezanzadeh A (2023) Pore pressure prediction by empirical and machine learning methods using conventional and drilling logs in carbonate rocks. Rock Mech Rock Eng 56:535–564. https://doi.org/10.1007/s00603-022-03089-y

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

Gharehbaghi A (2023) Deep learning in time series analysis. CRC Press

Ghazban F (2007) Petroleum geology of the persian gulf, University of Tehran

Ghosh S (2022) A review of basic well log interpretation techniques in highly deviated wells. J Pet Explor Prod Technol 12(7):1889–1906

Iosifidis A, Tefas A (2022) Deep learning for robot perception and cognition. Iosifidis A, Tefas A (eds). Elsevier p 16

Iturrarán-Viveros U, Parra JO (2014) Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. J Appl Geophys 107:45–54

Jo H, Cho Y, Pyrcz M, Tang H, Fu P (2021) Machine learning-based porosity estimation from spectral decomposed seismic data. Preprint at https://arXiv.org/abs/2111.13581

Kaydani H, Mohebbi A, Baghaie A (2012) Neural fuzzy system development for the prediction of permeability from wireline data based on fuzzy clustering. Pet Sci Technol 30(19):2036–2045. https://doi.org/10.1080/10916466.2010.531345

Kennedy M (2015) Log analysis part I: Porosity, Editor(s): Martin Kennedy, developments in petroleum science. Elsevier 62:181–207. https://doi.org/10.1016/B978-0-444-63270-8.00007-4

Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53(8):5455–5516

Kharraa HS, Al-Amri MA, Mahmoud MA, Okasha TM (2013) Assessment of uncertainty in porosity measurements using NMR and conventional logging tools in carbonate reservoir, SPE, Saudi Arabia, Section Technical Symposium and Exhibition, pp 570–586. https://doi.org/10.2118/168110-ms

Li W, Fu L, AlTammar MJ (2022) Well log prediction using deep sequence learning, ARMA/DGS/SEG International Geomechanics Symposium, Abu Dhabi, UAE, November. https://doi.org/10.56952/IGS-2022-184

Lin T, Mezghani M, Xu C, Li W (2021) Machine learning for multiple petrophysical properties regression based on core images and well logs in a heterogeneous reservoir. SPE Annual Technical Conference and Exhibition, Dubai, UAE. https://doi.org/10.2118/206089-MS

Matinkia M, Hashami R, Mehrad M, Hajsaeedi MR, Velayati A (2022) Prediction of permeability from well logs using a new hybrid machine learning algorithm. Petroleum 9(1):108–123. https://doi.org/10.1016/j.petlm.2022.03.003

McPhee C, Reed J, Zubizarreta I (2015) Chapter 5 - routine core analysis. In: McPhee C, Reed J, Zubizarreta I (eds) Developments in Petroleum Science, vol 64. Elsevier, pp 181–268

Moosavi N, Bagheri M, Nabi-Bidhendi M, Heidari R (2022) Porosity prediction using fuzzy SVR and FCM SVR from well logs of an oil field in south of Iran. Acta Geophysica

Moosavi N, Bagheri M, Nabi-Bidhendi M et al (2022b) Fuzzy support vector regression for permeability estimation of petroleum reservoir using well logs. Acta Geophys 70:161–172. https://doi.org/10.1007/s11600-021-00700-8

Munir MN, Zafar M, Ehsan M (2023) Comparative and statistical analysis of core-calibrated porosity with log-derived porosity for reservoir parameters estimation of the Zamzama gas field, southern Indus Basin, Pakistan. Arab J Sci Eng 48:7867–7882. https://doi.org/10.1007/s13369-022-07523-9

Okon EI, Anyadiegwu DA (2021) Application of machine learning techniques in reservoir characterization. In Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers, Lagos, Nigeria, August. https://doi.org/10.2118/208248-MS

Sfidari E, Kadkhodaie-Ilkhchi A, Rahimpour-Bbonab H, Soltani B (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. J Pet Sci Eng 121:87–102. https://doi.org/10.1016/j.petrol.2014.06.013

Sun Y et al (2024) Porosity prediction through well logging data: A combined approach of convolutional neural network and transformer model (CNN-transformer). Phys Fluids, 36(2). https://doi.org/10.1063/5.0190078

Sun J, Zhang R, Chen M et al (2021) Identification of porosity and permeability while drilling based on machine learning. Arab J Sci Eng 46:7031–7045. https://doi.org/10.1007/s13369-021-05432-x

Singh S, Kanli AI, Sevgen S (2016) A general approach for porosity estimation using artificial neural network method: A case study from Kansas gas field. Stud Geophys Geod 60(1):130–140

Tam TNT, Thanh DHT (2023) Estimate petrophysical properties by using machine learning methods. In: PL Vo, DA Tran, TL Pham, H Le Thi Thu, NV Nguyen (eds) Advances in Research on Water Resources and Environmental Systems pp 511–529. Hanoi, Springer. https://doi.org/10.1007/978-3-031-17808-5_29

Tavakolizadeh N, Bagheri M (2022) Multi-attribute Selection for Salt Dome Detection Based on SVM and MLP Machine Learning Techniques. Nat Resour Res 31:353–370. https://doi.org/10.1007/s11053-021-09973-8

Tiab D, Donaldson EC (2016) Porosity and permeability. In: D Tiab, EC Donaldson (eds) Petrophysics (4th ed). Gulf Professional Publishing, Houston, pp 67–186

Wang J, Cao J (2022) Deep learning reservoir porosity prediction using integrated neural network. Arab J Sci Eng 47(9):11313–11327

Ye A, Wang Z (2023) Modern Deep Learning for Tabular Data: Novel Approaches to Common Modeling Problems. Apress, Seattle, WA, US

Yousefmarzi F, Haratian A, Mahdavi Kalatehno J et al (2024) Machine learning approaches for estimating interfacial tension between

oil/gas and oil/water systems: A performance analysis. Sci Rep 14:858. https://doi.org/10.1038/s41598-024-51597-4

Zare A, Bagheri M, Ebadi M (2020) Reservoir facies and porosity modeling using seismic data and well logs by geostatistical simulation in an oil field. Carbonates Evaporites 35:65. https://doi.org/10.1007/s13146-020-00605-5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.