University of Tehran Press

Journal of the Earth and Space Physics

Homepage: https://jesphys.ut.ac.ir

Vol. 50, No. 4, Winter 2025, P. 91-113

(Research Article)

Enhancing Porosity Prediction Accuracy in Oil Reservoirs: Evaluating Hybrid Machine Learning Approaches Integrating Well Log and Core Data

Mehrabi, A. R.[1] | Bagheri, M.[2] ✉ | Nabi Bidhendi, M.[2] | Biniaz Delijani, E.[1] | Behnood, M.[1]

1. Department of Petroleum Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
2. Department of Earth Physics, Institute of Geophysics, University of Tehran, Tehran, Iran.

Corresponding Author E-mail: majidbagheri@ut.ac.ir

(Received: 8 June 2024, Revised: 19 Jan 2025, Accepted: 5 March 2025, Published online: 15 March 2025)

Abstract

Accurate prediction of porosity holds significant importance across various domains within the oil and gas sector, encompassing activities such as reservoir delineation, well design, and production enhancement. However, conventional methodologies often encounter difficulties in capturing the intricate relationships among diverse data streams and porosity metrics. This study introduces a novel hybrid model framework aimed at refining the precision and resilience of porosity forecasts by integrating multiple machine learning methodologies and exploiting complementary data modalities. This hybrid architecture enables flexible and intricate integration of diverse models and data sources, potentially leading to enhanced overall porosity prediction accuracy. Notably, the proposed model incorporates several innovative elements, including the amalgamation of ensemble techniques and deep learning models tailored for sequential data, as well as the utilization of complementary data sources, such as well log and core data, to facilitate automatic feature learning and representation, thereby bolstering robustness and generalization capabilities. Experimental outcomes underscore the hybrid model's potential to achieve notable prediction accuracies, with R-squared values surpassing 0.93 on log data and 0.88 on core data sets, outperforming individual models. The model also exhibits commendable robustness and training efficiency, leveraging advanced methodologies such as ensemble techniques. In conclusion, this study underscores the promise of hybrid machine learning models as dependable tools for porosity prediction from core data. The insights gleaned from this research hold the potential to advance the understanding and optimization of porosity forecasting, thereby facilitating the formulation of more efficient reservoir management strategies.

Keywords: Porosity, Log and core data, Hybrid model, Gradient Boosting Regression.

1. Introduction

Porosity is a fundamental characteristic of hydrocarbon reservoirs, exerting a significant impact on the volume and mobility of fluids within porous media. It represents the void spaces within rock formations that can contain fluids such as oil and gas. Elevated porosity is directly associated with a greater capacity for hydrocarbons in a reservoir, thereby profoundly affecting reservoir operations and management strategies (Al-Khafaji et al., 2024; Hussain et al., 2023; Nazari and Hajizadeh, 2023). In the domain of oil and gas exploration and production, reservoir porosity is a crucial parameter that dictates the fluid retention capacity of a reservoir. Accurate prediction of porosity is essential for successful exploration and production operations (Tiab and Donaldson, 2016). Traditionally, porosity determination relies on expensive methods such as core analysis and well testing, which are often challenging due to the absence of cores in many field wells. As a result, techniques that assess reservoir petrophysical properties, including porosity, using well logging charts are indispensable. Fortunately, well logs are readily available for most wells within a field (Ahmadi et al., 2014). In the oil and gas industry, porosity is more than just a technical detail; it is a strategic asset that guides various stages of exploration and production (Tariq et al., 2023).

The microscopic characteristics of pore structures within oil and gas reservoirs,

Cite this article: Mehrabi, A. R., Bagheri, M., Nabi Bidhendi, M., Biniaz Delijani, E., & Behnood, M. (2025). Enhancing Porosity Prediction Accuracy in Oil Reservoirs: Evaluating Hybrid Machine Learning Approaches Integrating Well Log and Core Data. *Journal of the Earth and Space Physics*, 50(4), 91-113. DOI: http//doi.org/10.22059/jesphys.2025.377351.1007611

E-mail: (1) amirreza.mehrabi@srbiau.ac.ir | biniaz@srbiau.ac.ir | m.behnood@gmail.com (2) mnbhendi@ut.ac.ir

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/jesphys.2025.377351.1007611

Print ISSN: 2538-371X
Online ISSN: 2538-3906

including pore shape, size distribution, and connectivity, intricately regulate fluid movement and hydrocarbon flow (Xu et al., 2022). Effective porosity, which represents the connected pore space that facilitates fluid flow and movement, is particularly significant. Innovative techniques have been developed to enhance permeability in low-permeability shale reservoirs by inducing controlled fractures, thereby increasing hydrocarbon production. Accurate porosity estimation is crucial for precisely assessing reservoir attributes within the oil and gas sectors (Gamal and Elkatatny, 2022).

Optimal porosity prediction is crucial for reducing reliance on logging and core tests, underscoring porosity's essential role in various applications within the oil and gas industry (Gamal and Elkatatny, 2022; Hussain et al., 2023). This highlights its paramount importance in reservoir assessment and management. Porosity measurement in the oil and gas sector involves diverse methodologies. The predominant approach relies on laboratory measurements of core samples acquired during drilling operations. This process, known as Routine Core Analysis (RCA), includes stages such as core cutting, handling, preservation, transport, sampling, and subsequent testing (Moosavi et al., 2023). Determining the bulk volume of a sample can be performed through physical measurement or displacement methods. Physical measurement is suitable for core samples with regular geometric shapes, whereas displacement methods involve immersing the core sample in mercury within a pycnometer or graduated cylinder. Besides these direct techniques, indirect methods based on void space properties are also utilized (Maniscalco et al., 2022; Li et al., 2021). These include assessing the electrical conductivity of an electrically conductive fluid within the void space of the sample or monitoring the absorption of radioactive particles by a fluid occupying the void space. Furthermore, advanced technologies such as nuclear magnetic resonance (NMR), electrical resistivity, and well logging are employed for porosity evaluation (Moosavi et al., 2023; Nasseri and Mohammadzadeh, 2017).

Well logging, also known as borehole logging, is a critical process in the oil and gas industry that involves recording detailed information about the geological formations penetrated by a borehole. This process entails lowering specialized instruments, known as logging tools, into the wellbore to capture subsurface data related to rock properties and fluid characteristics at various depths (Yu et al., 2008). The data obtained from well logging provides valuable insights into the types of rock present, their ability to retain oil and gas, and the reservoir's potential productivity. By carefully analyzing this data, geologists and engineers can make informed decisions regarding drilling operations, production strategies, and overall reservoir management. Well log analysis plays a crucial role in identifying optimal drilling locations, accurately estimating reserves, and optimizing production methods to maximize oil and gas extraction from underground reservoirs (Fu, 2003; Shiri et al., 2011).

Well logs offer a concise yet comprehensive graphical representation of formation parameters against depth. These plots enable interpreters to distinguish between different lithologies, identify porous and non-porous rocks, and swiftly recognize lucrative pay zones within subsurface formations. The skill of log interpretation depends on understanding the significance of each measurement (Hussain et al., 2023). As mentioned before, both log data and core data are crucial for determining porosity. Although conventional methods can be resource-intensive, adopting efficient methodologies is essential to minimize costs (Yousefmarzi et al., 2024). Looking ahead, emerging technologies and techniques hold significant promise for advancing porosity calculations. The integration of nanotechnology, advanced imaging and characterization techniques (Gupta et al., 2022), computational modeling, and machine learning can greatly enhance our understanding of interfacial behavior and improve the accuracy of porosity predictions (Esteghlal et al., 2023). Machine learning and neural networks have increasingly demonstrated their utility in the oil and gas industry, particularly in applications such as reservoir characterization (Mahzad and Riahi, 2024) and seismic image reconstruction (Mahzad and Bagheri, 2025). These approaches enable computers to learn from data without requiring explicit programming. They excel at identifying complex patterns and relationships within data and can make accurate predictions based on new information (Dargi et al., 2023; Zamani et al., 2023; Kirch et al., 2020). Exploring previous research on porosity estimation provides valuable insights into various methods, challenges, opportunities, and innovative approaches in this field.

Sun et al. (2024) developed a Convolutional Neural Network (CNN) and Transformer model to enhance the accuracy and generalization of logging porosity predictions. This innovative CNN-Transformer model, trained on a well log dataset, demonstrated significant superiority in logging porosity prediction tasks, achieving a notable R^2 value of 0.95 (Sun et al., 2024). Jo et al. (2021) proposed a machine learning-based workflow to transform seismic data into porosity models. Their approach involved a ResUNet++-based workflow capable of processing three seismic datasets at different frequencies (decomposed seismic data) to generate corresponding porosity models. The effectiveness of this workflow was validated in a 3D channelized reservoir, showing R^2 scores above 0.9 for both training and validation datasets (Jo et al., 2021).

In 2023, Tam and Thanh integrated traditional petroleum engineering methods with widely-used machine learning techniques to estimate porosity and permeability using petrophysical data. Their study, utilizing data from the Volve field in Norway and incorporating well logging and core logging data, focused on predicting porosity and permeability using an Artificial Neural Network (ANN) model, which was compared to the Least-Squares Support Vector Machines (LSSVM) model and an empirical model. The ANN model demonstrated exceptional performance, achieving the highest coefficient of determination (R^2) of 0.9997 and the lowest Mean Squared Error (MSE) of 6.7769 for porosity and permeability prediction (Tam and Thanh, 2023).

In another study, Munir et al. (2023) conducted a comparative and statistical analysis of core-calibrated porosity versus log-derived porosity for estimating reservoir parameters in the Zamzama GAS Field, Southern Indus Basin, Pakistan. They employed conventional and deep machine learning algorithms such as Support Vector Regression (SVR), Random Forest (RF), and Multilayer Perceptron (MLP), all of which showed accuracies above 0.96 (Munir et al., 2023).

A 2021 study by Jian Sun et al. utilized machine learning methodologies for real-time prediction of reservoir porosity and permeability during drilling operations. This study emphasized both high prediction accuracy and rapid model processing to accommodate the integration of new logging data while drilling. Four machine learning algorithms—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)—were evaluated, with accuracies ranging from 0.88 to 0.92 (Sun et al., 2021).

This paper presents a novel and comprehensive study on modeling porosity of four oil wells in the Foroozan oil field, using intelligent hybrid machine learning techniques. The results are compared with previous studies that have used different approaches. Table 1 summarizes the main features and findings of the reviewed studies. As shown in Table 1, the current study offers several advantages and innovations over the existing literature, which are discussed below. Unlike most previous studies, which have either neglected some key parameters or focused on specific cases while primarily using single-model approaches, the current study considers a broader range of input parameters affecting reservoir porosity.

It aims to capture the complex relationship between log data and core data and their integration to achieve highly accurate predictions, particularly from core data. This study employs Long Short-Term Memory (LSTM) networks and Gradient Boosting Regressor (GBR) separately for log data and integrates them into a hybrid model to predict porosity from both log and core data, ensuring more precise results in porosity prediction.

Table 1. Comparison of reviewed studies and their Modeling Results.

Authors (year)	Input models	Output models	Model types	Results	Accuracy
Youzhuang Sun, Shanchen Pang, Junhua Zhang, and Yongan Zhang (2024)	Well log dataset	Logging porosity prediction	Convolutional Neural Network (CNN) and Transformer model	The CNN-transformer model showed good superiority in the task of logging porosity prediction	All accuracies are above 0.96
Honggeun Jo, Yongchae Cho, Michael J. Pyrcz, Hewei Tang, and Pengcheng Fu (2021)	Three seismic data in different frequencies (i.e., decomposed seismic data)	Porosity	ResUNet++ based workflow	The workflow was successfully demonstrated in the 3D channelized reservoir to estimate the porosity model.	More than 0.9 in R2 score for training and validating data.
Tran Nguyen Thien Tam and Dinh Hoang Truong Thanh (2023)	well logging and core logging data	porosity and permeability	Artificial neural network (ANN) model, as compared to Least-squares support-vector machines (LSSVM) model and empirical model	Ability to handle both core dataset and log data set and impressive prediction of both Porosity and Permeability	Metrics R- Squared (coefficient of determination) of 0.9997 and lowest MSE (mean squared error) of 6.7769
Muhammad Nofal Munir, Mohammad Zafar and Muhsan Ehsan (2023)	well logging and core logging data	core-calibrated porosity and log-derived porosity	support vector regression (SVR), random forest (RF), and the multilayer perceptron (MLP)	Ability to handle both core dataset and log data set and impressive prediction of core-calibrated porosity and log-derived porosity	$R^2 = 0.96$
Jian Sun et.al (2021)	logging data	Reservoir porosity and permeability	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)	Highly accurate in predicting reservoir porosity and permeability	Accuracies ranging between 0.88 to 0.92

2. Methodology

2-1. Dataset and geological setting

The datasets utilized in this study were obtained from the Foroozan Oil Field, located in the Persian Gulf. Four oil wells were analyzed in this investigation: Well F-02, Well F-03, Well F-09, and Well F-14. The Foroozan Oil Field, also referred to as Forouzan, is situated in the Persian Gulf, approximately 100 kilometers southwest of Kharg Island, Iran. It spans the Iran-Saudi Arabia border, with the majority of its hydrocarbon reserves–accounting for over 80%–located within Saudi Arabian waters. The field was discovered in 1966 and is managed by the National Iranian Oil Company (NIOC) along with its subsidiary, the Iranian Offshore Oil Company (IOOC). Initially, the field was equipped with extensive infrastructure, including 66 wells, two production platforms, one processing unit, twelve wellhead platforms, three separators, one desalting tower, and two living quarter platforms. Over the years, the Foroozan Oil Field has undergone several phases of renovation and redevelopment. These efforts were undertaken to boost its crude oil production capacity and improve the output of gas and condensates, as documented by Hassanzadeh et al. (2019).

The Foroozan Oil Field is structurally divided into several areas, designated as F1, F2, F3, F5, and F8. These areas are separated by faults or synclines, creating distinct structural domains. Among them, the F1 and F5 areas exhibit limited production potential, while F2, F3, and F8 are the primary productive regions. Notably, the F2 and F3 areas share a border with the Marjan field, highlighting their significance in regional hydrocarbon production.

The main hydrocarbon-bearing reservoirs in the Foroozan field include the Burgan A and B members, the Lower Dariyan member, the Yamama Formation, the Manifa Formation, and the Arab Formation.

The location map of the studied wells in the Foroozan Oil Field is presented in Figure 1. This map illustrates the spatial distribution of the selected wells throughout the field. The wells were chosen strategically to reflect geological and petrophysical variations within the reservoir. Their scattered distribution ensures a comprehensive assessment of the changes in formation properties across different parts of the field.

Core sampling has been extensively conducted within the primary reservoirs to support routine core analysis. Specifically, the Burgan Member is well-represented, with core samples from eight wells across all productive areas, encompassing its full thickness. In contrast, the Lower Dariyan Member has limited core coverage, with samples obtained from the upper section in three wells (F-03-00, F-06-00, and F-14-00) within the F2 and F3 areas. Similarly, the Mauddud Member has minimal core data, with a small interval cored in two wells (F-02-00 and F-09-00), primarily within its limestone section. A limited number of samples from a sandy layer in the Mauddud Member, at 1900 mMD in Well F-09-00, provide additional insights.

Figures 2 to 5 display the well logs obtained from the selected wells (F-02, F-03, F-09, and F-14). Each figure illustrates the complete suite of logs for a single well, including Borehole Size (BS), Caliper (CALI), Gamma Ray (GR), Neutron Porosity (NPHI), Bulk Density (RHOB), Density Correction (DRHO), Slowness (DT), Resistivity (RT), and Effective Porosity (PHIE). These logs provide essential data for interpreting subsurface geology and reservoir characteristics, capturing key variations in lithology, fluid content, and reservoir quality. The well logs were selected to complement the spatial analysis shown in the location map, ensuring consistency in the evaluation of reservoir conditions across the field.

Overall, an analysis of the available core data reveals minimal variation in rock properties within the primary reservoir intervals, particularly in the Burgan and Yamama Members, where data coverage is both adequate and well-distributed. These formations, along with the Arab and Manifa Members, constitute the key hydrocarbon-bearing units within the field.

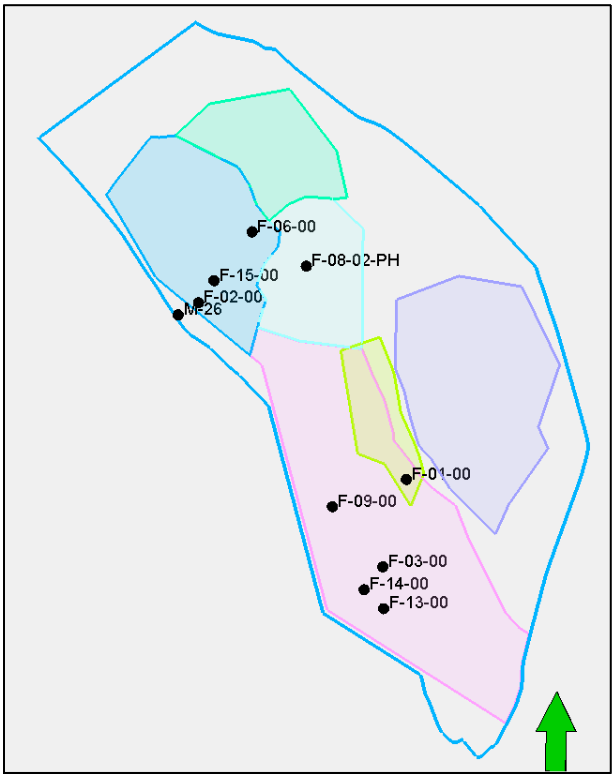


Figure 1. Location map of wells in the Foroozan Field. Source: CAPE Consultant Group, Special Core Analysis Report prepared for the Iranian Offshore Oil Company (IOOC), August 2021.

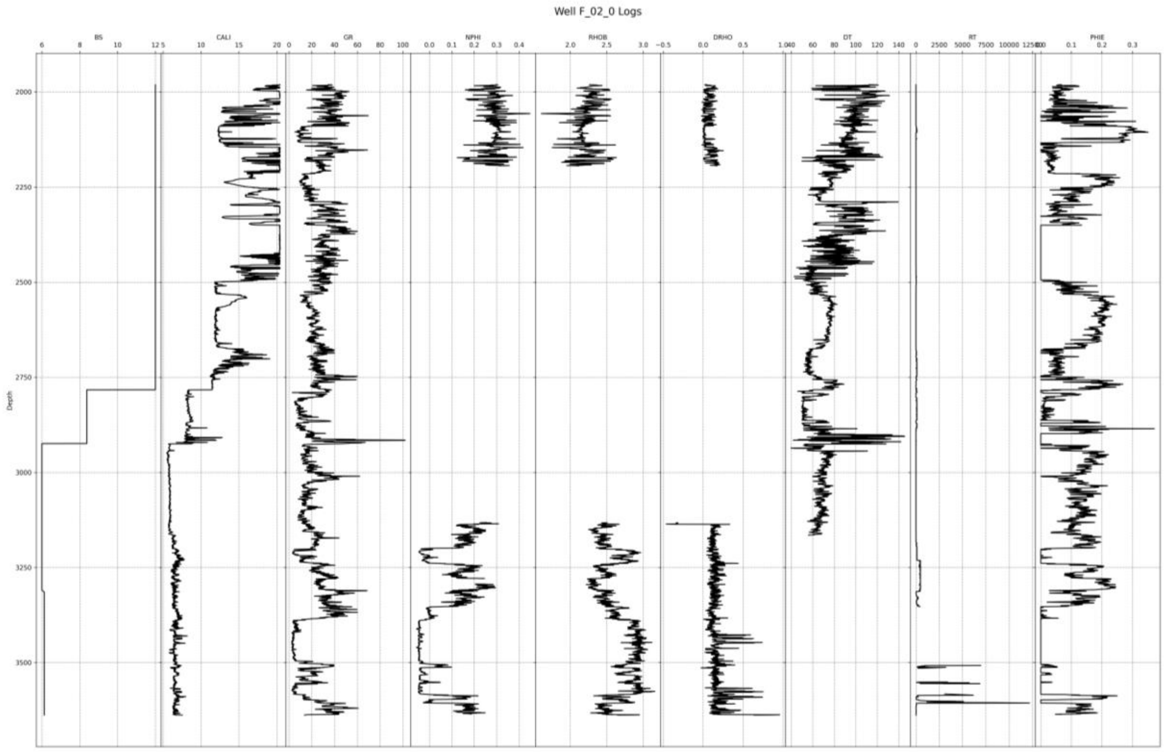


Figure 2. Graphical representation of well logs for Well F-02-0. The logs include Borehole Size (BS), Caliper (CALI), Gamma Ray (GR), Neutron Porosity (NPHI), Bulk Density (RHOB), Density Correction (DRHO), Slowness (DT), Resistivity (RT), and Effective Porosity (PHIE). These logs provide a detailed assessment of the subsurface geological and petrophysical properties.

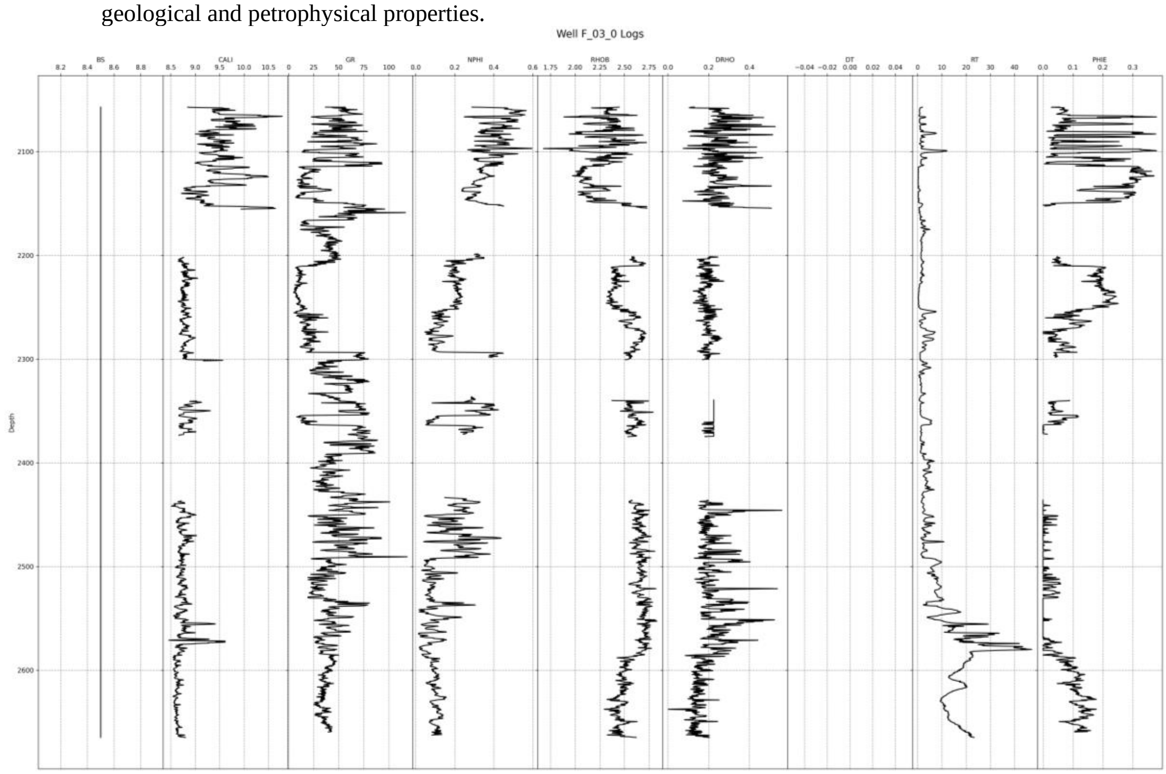


Figure 3. Well log data from Well F-03-0, showing Borehole Size (BS), Caliper (CALI), Gamma Ray (GR), Neutron Porosity (NPHI), Bulk Density (RHOB), Density Correction (DRHO), Slowness (DT), Resistivity (RT), and Effective Porosity (PHIE). The data offers critical insights into formation evaluation and reservoir characterization.

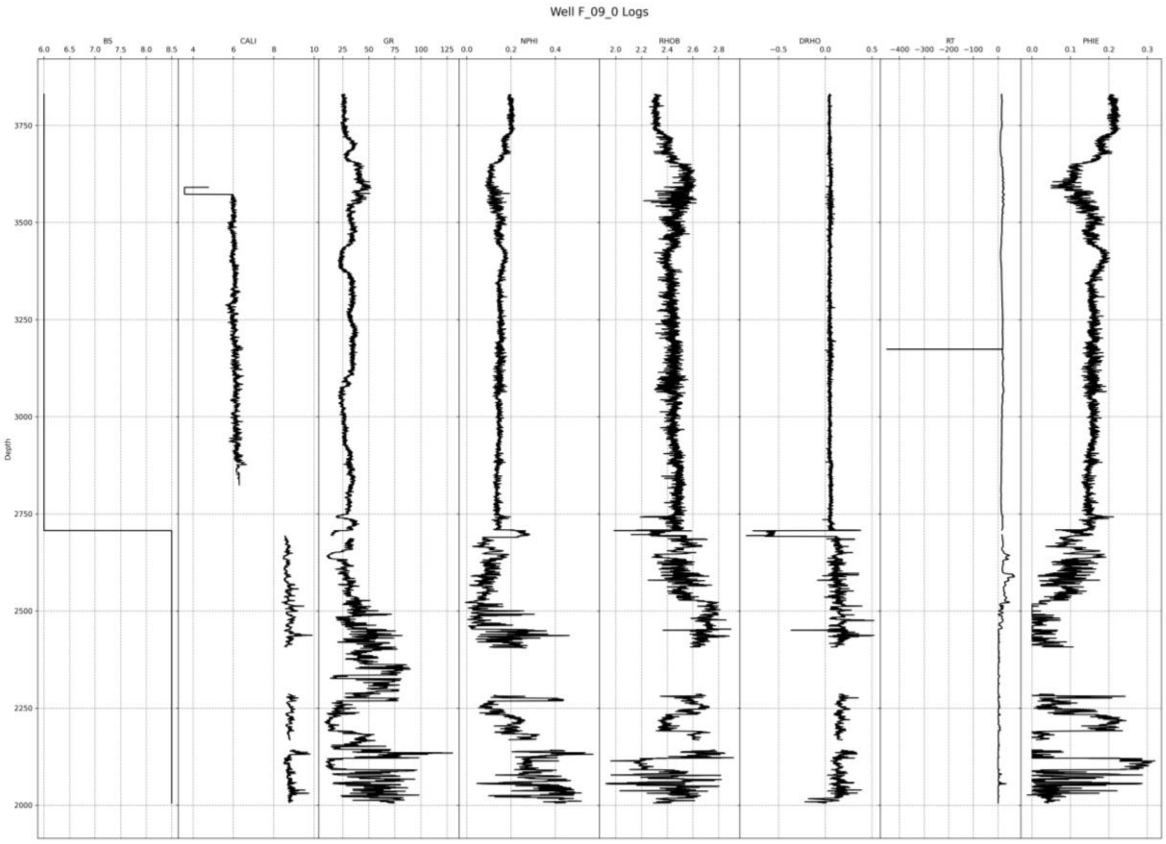


Figure 4. Well log suite for Well F-09-0, illustrating Borehole Size (BS), Caliper (CALI), Gamma Ray (GR), Neutron Porosity (NPHI), Bulk Density (RHOB), Density Correction (DRHO), Slowness (DT), Resistivity (RT), and Effective Porosity (PHIE). These measurements enable detailed analysis of lithology and fluid content.

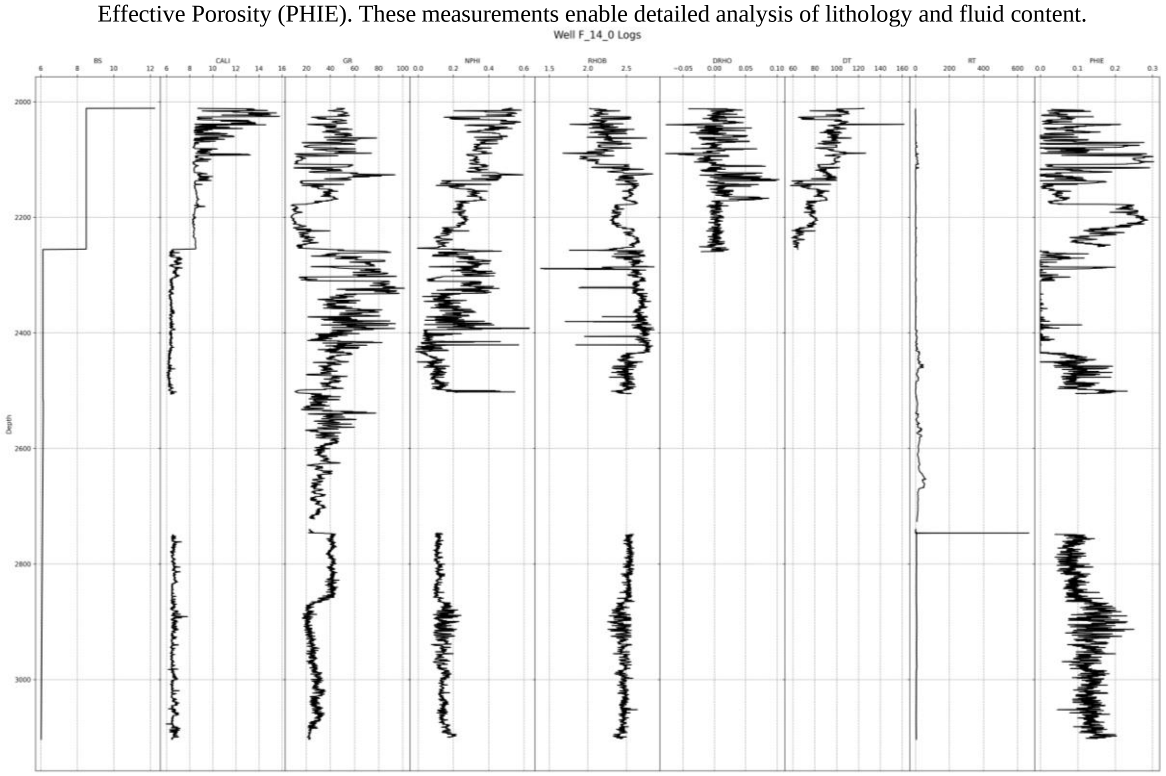


Figure 5. Visualization of well logs from Well F-14-0, depicting Borehole Size (BS), Caliper (CALI), Gamma Ray (GR), Neutron Porosity (NPHI), Bulk Density (RHOB), Density Correction (DRHO), Delta-T (Slowness) (DT), Resistivity (RT), and Effective Porosity (PHIE). The suite aids in interpreting subsurface geology and assessing reservoir quality.

2-2. Data preparation

Data preparation is a critical component of the machine learning workflow, as it directly impacts both data quality and model performance (Carey et al., 2015; Jo, 2019). In this study, the data preparation steps include handling missing values, imputing incomplete records, addressing outliers, and normalizing the dataset to ensure uniform feature scaling. These processes are crucial for creating a robust and reliable dataset for machine learning models (Talebkeikhah et al., 2021; Al Shalabi and Shaaban, 2006).

Missing values were managed through imputation techniques, ensuring that incomplete records did not compromise the dataset's integrity. Each feature was evaluated to identify missing values, and appropriate imputation strategies were applied to replace them. This step ensured that the dataset remained complete and consistent, minimizing the risk of bias during model training.

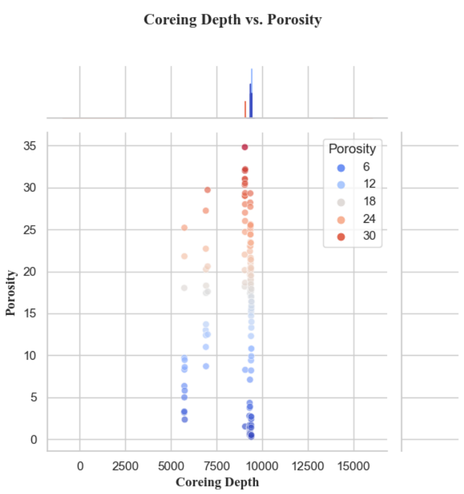
Outlier detection and handling were conducted using a rigorous and systematic approach. Statistical methods, including the Interquartile Range (IQR), were applied to identify potential outliers. For each quantitative feature, the IQR was calculated, and observations falling below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR were flagged as potential outliers. Once identified, these potential outliers underwent a thorough examination. Instances linked to data entry errors or clear discrepancies were carefully removed. For potential outliers representing genuine data variability, a detailed review ensured these critical data points were retained. Domain experts provided contextual assessments to avoid the inadvertent exclusion of significant observations. This careful balance between removing erroneous outliers and preserving valuable data enhanced the dataset's reliability and ensured it reflected the full range of variability.

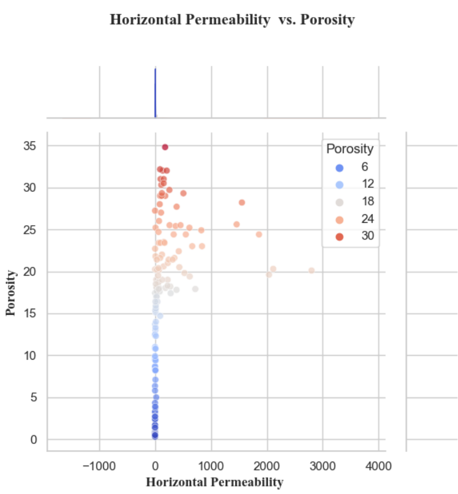
Normalization was applied to scale numerical features consistently across the dataset. This process involved transforming each feature to a scale ranging from 0 to 1. Mathematically, normalization was performed using the formula:

$$\text{Normalized } X = (X - X\_\text{minimum}) / (X\_\text{maximum} - X\_\text{minimum}) \tag{1}$$

Here, 'X_minimum' and 'X_maximum' represent the minimum and maximum values of the feature 'X', respectively. By standardizing feature magnitudes, normalization ensured fair comparisons and combinations of variables with different units or scales. This step not only facilitated the training process of machine learning models but also enhanced their predictive accuracy by reducing biases associated with inconsistent feature scaling (Pan et al., 2016; Jo, 2019).

Figures 6 and 7 present data joint plots to identify outliers within the dataset. For simplicity, the joint plots focus on data from well F-02. These visualizations provide a detailed perspective on the dataset's distribution, enabling a comprehensive assessment of potential anomalies and their impact on the analytical process. This systematic approach to data preparation ensures the creation of high-quality datasets that support the development of robust and effective machine learning models.





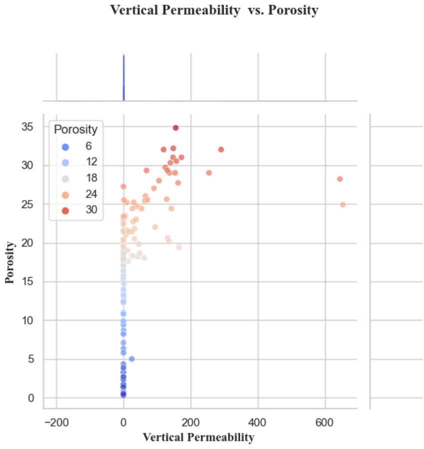
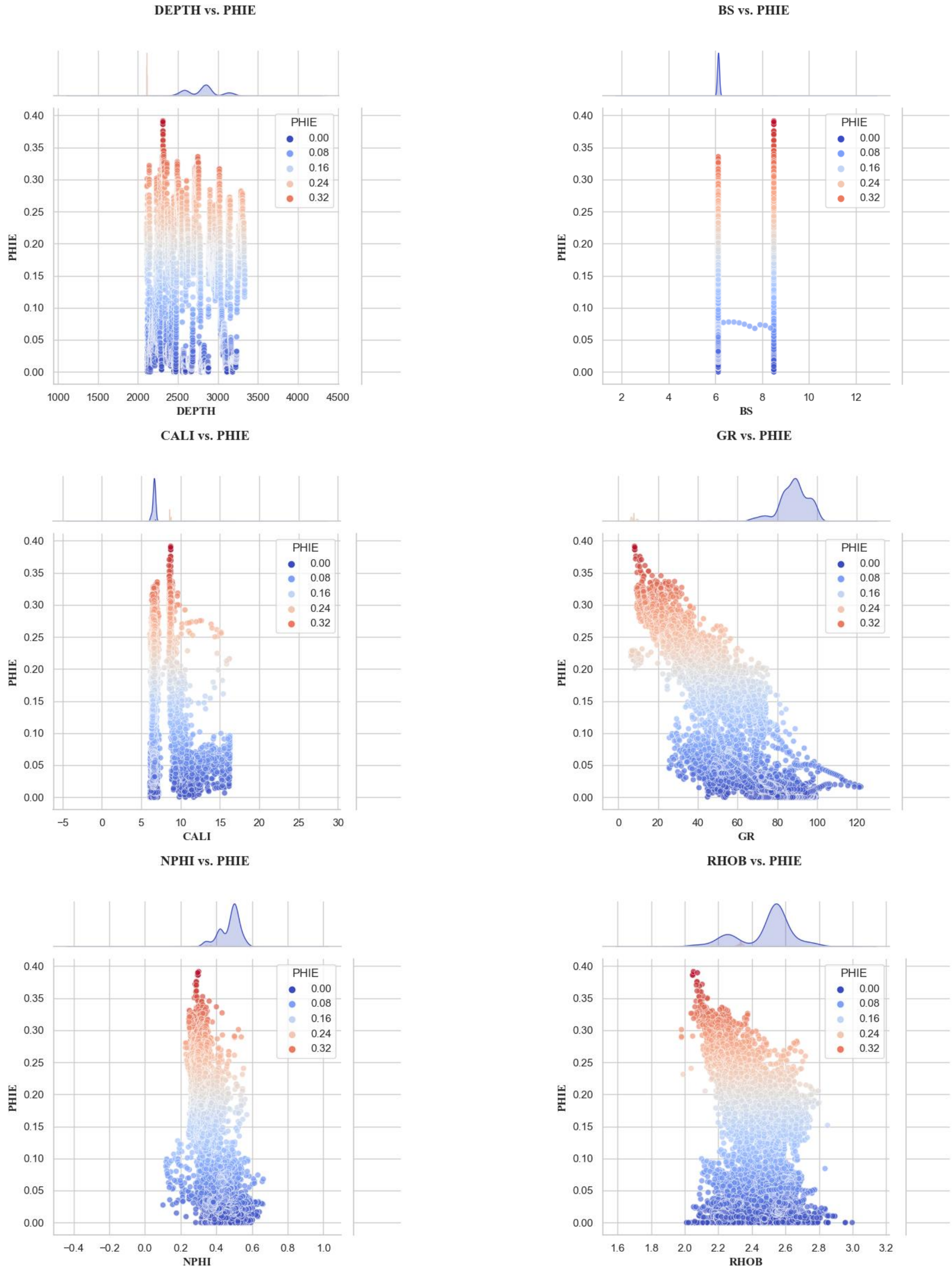


Figure 6. Joint plots for examination of different outliers in Core data- Well F-02.



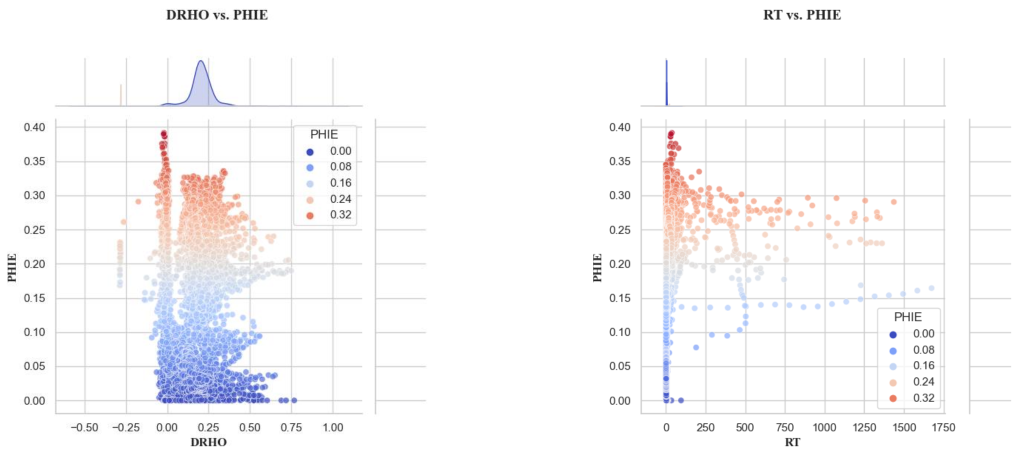


Figure 7. Joint plots for examination of different outliers in Log data- Well F-02 (see text for symbols).

2-3. Model Design and Training Process

In this research, advanced long short-term memory (LSTM) networks were utilized alongside computational techniques to develop models for porosity estimation using well log data. Additionally, a Gradient Boosting Regressor (GBR) model was developed to estimate porosity based on core data from four oil wells. Subsequently, a hybrid model was created by combining the well log-based LSTM model with the GBR model for core data, aiming to achieve optimal porosity forecasting accuracy by leveraging both log and core data sources. The study utilized a comprehensive dataset comprising well logs and corresponding core data from four oil wells, offering detailed reservoir property characterization.

All input parameters–DEPTH, BS, CALI, GR, NPHI, RHOB, DRHO, and RT–along with PHIE (porosity) as the output, were utilized in developing the well log-based LSTM model for porosity prediction. Conversely, the GBR and final hybrid models were developed using input parameters such as horizontal permeability, coring depth, vertical permeability, and porosity (as the output) to predict porosity from core data. Both models were ultimately integrated into a hybrid framework, which will be discussed in subsequent sections.

These parameters derived from both log and core data were selected based on their significant role in system behavior and their contribution to accurate predictions of production performance. The experimental dataset was randomly divided into two subgroups, with 80% allocated for training the models and the remaining 20% used to test the models' efficiency and reliability on blind cases. This data allocation method is widely recognized for producing reliable and desirable results.

Table 2 presents the hyperparameters as control parameters for each modeling technique used in this study, including long short-term memory (LSTM) networks and Gradient Boosting Regressor (GBR) models. Hyperparameters are critical settings that affect the behavior and performance of machine learning models. They are not derived from the data but chosen prior to training and can significantly impact the model's ability to learn and generalize. Proper hyperparameter tuning is essential for optimizing both model accuracy and efficiency. The hyperparameters for each well were tuned using a cross-validation and other techniques in LSTM, GBR, and hybrid models. The rationale behind the selected hyperparameters for each model is critical for their optimization and overall effectiveness. Each model's parameters were carefully adjusted to ensure a robust predictive capability.

For Well F-02, the architecture of the LSTM model was designed to accommodate the specific characteristics of the training data. The input shape of the LSTM layer was defined as (1, num_features), where num_features represents the total number of input features in the dataset. The LSTM layer consisted of multiple units, with the exact number determined through hyperparameter tuning. Following the LSTM layer, a dense layer with a single output unit was employed for regression tasks. A linear activation function was used by default, as no explicit activation function was specified. The model optimization was performed using the

RMSprop optimizer, while the Mean Squared Error (MSE) function was utilized to calculate the loss during training.

The hyperparameters for the LSTM model included several critical variables. The number of LSTM units was chosen from the set {32, 64, 96, 128}, while the learning rate for the RMSprop optimizer was selected from the values {0.01, 0.001, 0.0001}. A batch size of 32 was used for training, with the number of epochs set to 100 by default. To prevent overfitting, early stopping was applied with a patience of 10 epochs. The results of these hyperparameter configurations are summarized in Table 3, which also provides control parameters for the core data used in the final hybrid model.

The hybrid model for Well F-02 integrated the LSTM model and the Gradient Boosting Regressor (GBR) model to leverage both well log and core data. The LSTM component retained the same architecture, with an input shape of (1, num_features), multiple LSTM units, and a dense output layer. The RMSprop optimizer was employed, with the learning rate determined during hyperparameter tuning, and MSE function was applied as the loss function. For the GBR component, the model was initialized with hyperparameters such as the learning rate, number of estimators, and maximum depth. These parameters were optimized using the hyperopt library to minimize the MSE.

Table 2. Control parameters used for the development and application of soft computing techniques for Porosity estimation in long short-term memory networks (LSTM) and Gradient Boosting Regressor (GBR) models.

Model	Parameters	Well F02	Well F03	Well F09	Well F14
LSTM	Learning Rate	0.01	0.01	0.01	0.01
	Dense	Single dense layer with one output unit, which is used .for regression	Single dense layer with one output unit, which is used .for regression	Single dense layer with one output unit, which is used .for regression	Single dense layer with one output unit, which is used .for regression
	Epochs	100	100	100	100
	Batch size	32	32	32	64
	Optimizer	RMSprop	RMSprop	Rmsprop	Adam
	Loss Function	Mean squared error (MSE)	Mean squared error (MSE)	Mean squared error (MSE)	Mean squared error (MSE)
	Early Stopping	10	10	10	10
	Units	32	32	32	64
Gradient Boosting	Learning rate	0.1	0.1	0.05	0.2
	Max depth	6	6	6	6
	Max features	1	0.3	0.3	1
	Min samples leaf	3	5	3	3
	n estimators	500	600	600	600
	CV	5	5	5	5

Hyperparameter tuning for the LSTM and GBR models yielded specific configurations. For the LSTM model, the number of units was optimized to 32, and the best learning rate was determined to be 0.01 from the set {0.01, 0.001, 0.0001}. For the GBR model, the learning rate was optimized within the range [0.01, 0.5], the number of estimators was chosen from {50, 100, 150, 200}, and the maximum depth was selected from {3, 4, 5, 6}. These configurations ensured optimal performance for both components of the hybrid model.

The evaluation metrics for the hybrid model included R-squared, root mean squared error (RMSE), and mean absolute error (MAE), calculated separately for both well log and core data. For well log data, the predicted porosity values were compared with the actual measurements, while similar evaluations were performed for core data predictions. The results were visualized using scatter plots comparing actual versus predicted porosity values. Separate plots were generated for the training and test sets, with the training set visualized using triangle markers and the test set by diamond markers.

Table 3. Control parameters used for the development and application of soft computing techniques for Porosity estimation in final Hybrid model.

Model	Parameters	Well F02	Well F03	Well F09	Well F14
Gradient Boosting	Learning rate	0.05	0.1	0.01	0.05
	Max depth	4	4	10	4
	Max features	1	0.5	1	1
	Min samples leaf	9	3	5	3
	n estimators	200	100	500	100
	CV	5	5	5	5
Final Hybrid Model	Estimator	Meta learner regressor	Meta learner regressor	Meta learner regressor	Meta learner regressor
	Activation	Relu	Relu	Relu	Relu
	Optimizer	Adam	Adam	Rmsprop	Adam
	dense__0__units	64	128	64	32
	dense__1__units	32	32	64	32
	dense__2__kernel _regularizer	None	None	(0.01) L1	(0.01) L1
	Epochs	50	50	100	100
	Batch size	32	64	64	64
	Weights	0.6	0.6	0.4	0.6
	Ensemble Model	Voting Regressor	Voting Regressor	Voting Regressor	Voting Regressor

In this research, the well log data inherently exhibits a sequential structure, with measurements recorded at various depths along the wellbore. To effectively capture depth-wise dependencies and sequential patterns within this data, LSTM networks are employed (Wu et al., 2021). These networks enable the modeling of relationships between measurements at different depths. Gradient boosting, on the other hand, is an ensemble learning technique that combines multiple weak learners, typically decision trees, to develop a robust predictive model. Unlike LSTM, which is commonly utilized for sequential data analysis, gradient boosting operates by iteratively fitting decision trees to the residuals of the preceding trees. This approach enables the modeling of relationships between measurements at varying depths without relying solely on the sequential nature of the data (Zou et al., 2021; Subasi et al., 2022).

The proposed hybrid model structure aims to capitalize on the advantages of both GBR and LSTM to enhance porosity prediction based on well log data. Predictions generated by the GBR and LSTM models are merged, which could be a neural network or another ensemble model. This model is trained to effectively merge the predictions from the GBR and LSTM models, utilizing the strengths of each component. It is capable of learning intricate non-linear relationships between the GBR and LSTM predictions, potentially resulting in enhanced overall accuracy of porosity prediction. Through the integration of GBR and LSTM models, this hybrid approach can capture both the non-linear associations among well log features and porosity (via GBR) and the sequential and spatial characteristics within the well log data (via LSTM).

To effectively leverage both well log data and core data, we propose a hybrid model that integrates core-based gradient boosting with a well log-based LSTM model to predict porosity from both data sources (Bittar et al., 2021; Alyaev and Elsheikh, 2022; Abbas et al., 2023; Hadavimoghaddam et al., 2021). The hybrid model consists of two main components. The first is the LSTM component, designed to processes sequential well log data and capture depth-wise dependencies. By learning from variations in well log features across different depths, the LSTM model effectively models the sequential relationships inherent in well log data. The second component is the Gradient Boosting Regressor (GBR), which leverages ensemble techniques to model the complex, non-linear relationships present in both well log and core data. The GBR is particularly effective for capturing intricate dependencies between core features and porosity.

The final hybrid model integrates these two components, with the LSTM model trained on well log data, and the GBR model, trained specifically on core data. By merging the sequential modeling capabilities of the LSTM with the predictive power of gradient boosting, the hybrid model benefits from the complementary strengths of both techniques. This integration aims to improve overall porosity prediction accuracy by utilizing ensemble techniques, deep learning for sequential data, and the complementary information provided by well log and core data.

The selection of these models was a critical step in the research methodology, guided by several critical criteria. These criteria include established efficacy, ensuring that the chosen models have demonstrated proven performance in similar contexts; diversity in learning approaches, enabling the capture of varied data patterns; and optimization capabilities, which ensure adaptability to different datasets and hyperparameter configurations. Additionally, achieving a balance between interpretability and performance was a key consideration. Finally, the relevance of the selected models to the dataset was a key consideration, ensuring that the techniques used align with the characteristics of the well log and core data.

3. Results and discussion

This section presents a comparative analysis of the performance of each model separately and, at last, the final results of the hybrid model for all four wells in predicting porosity. These models were introduced in the previous section. First, we delve into the results related to models developed based on log data. Figure 8 illustrates the cross-plot results obtained from LSTM model, Figure 9 for GBR model and Figure 10 and Figure 11 show the results from the hybrid model for log and core data, respectively. Figure 12 and Figure 13 also demonstrate the R^2 values obtained from these three models. R^2, also known as the coefficient of determination, serves as a statistical metric reflecting the extent to which independent variables account for the variance in a regression model, thereby indicating the model's adequacy in fitting the data. Ranging from 0 to 1, with 0 denoting that the model fails to explain any variance around the mean of the response data and 1 implying a complete explanation of such variance, R^2 quantifies the model's explanatory capability. Typically, an R^2 value of 0.7 or higher is deemed satisfactory in practical applications, suggesting a robust explanatory power of the model (Stanton, 2001; Fernando, 2023).

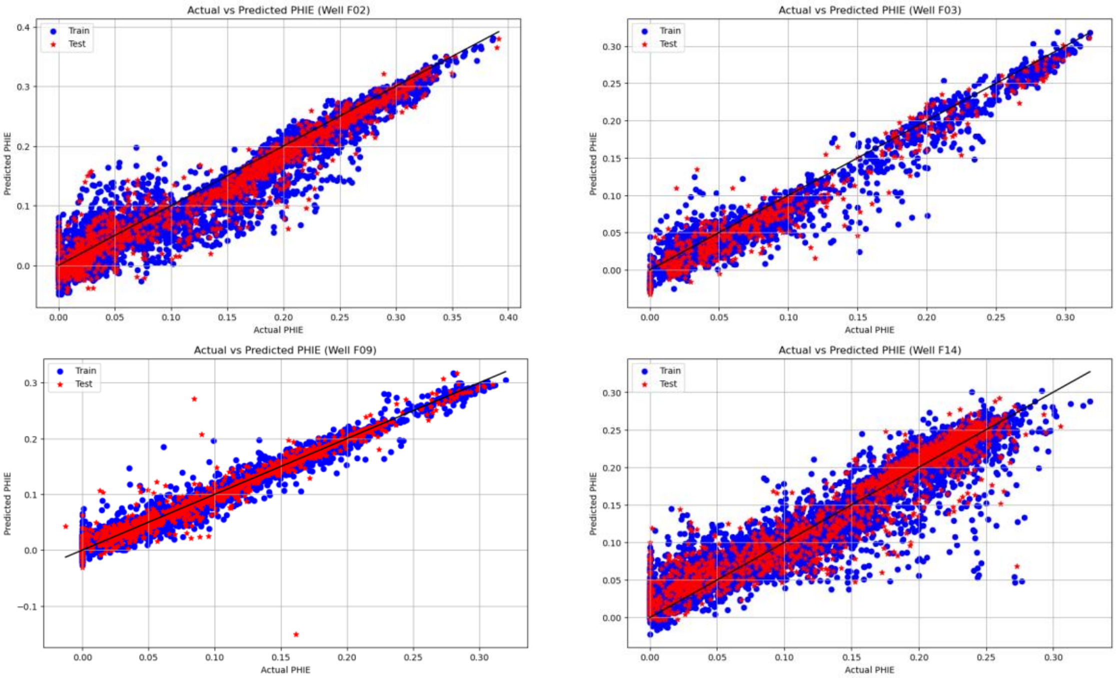


Figure 8. The cross plot of LSTM modeling prediction of Real PHIE versus Predicted PHIE in all wells.

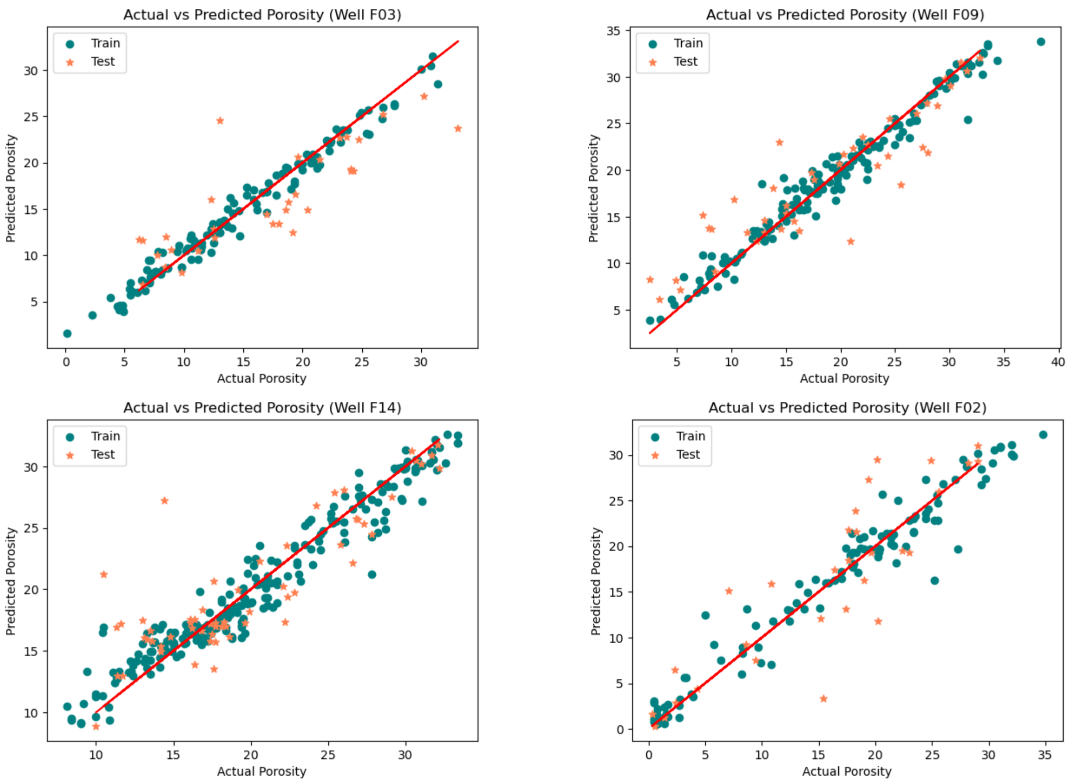


Figure 9. The cross plot of GBR modeling prediction of Real Porosity versus Predicted Porosity in all wells.

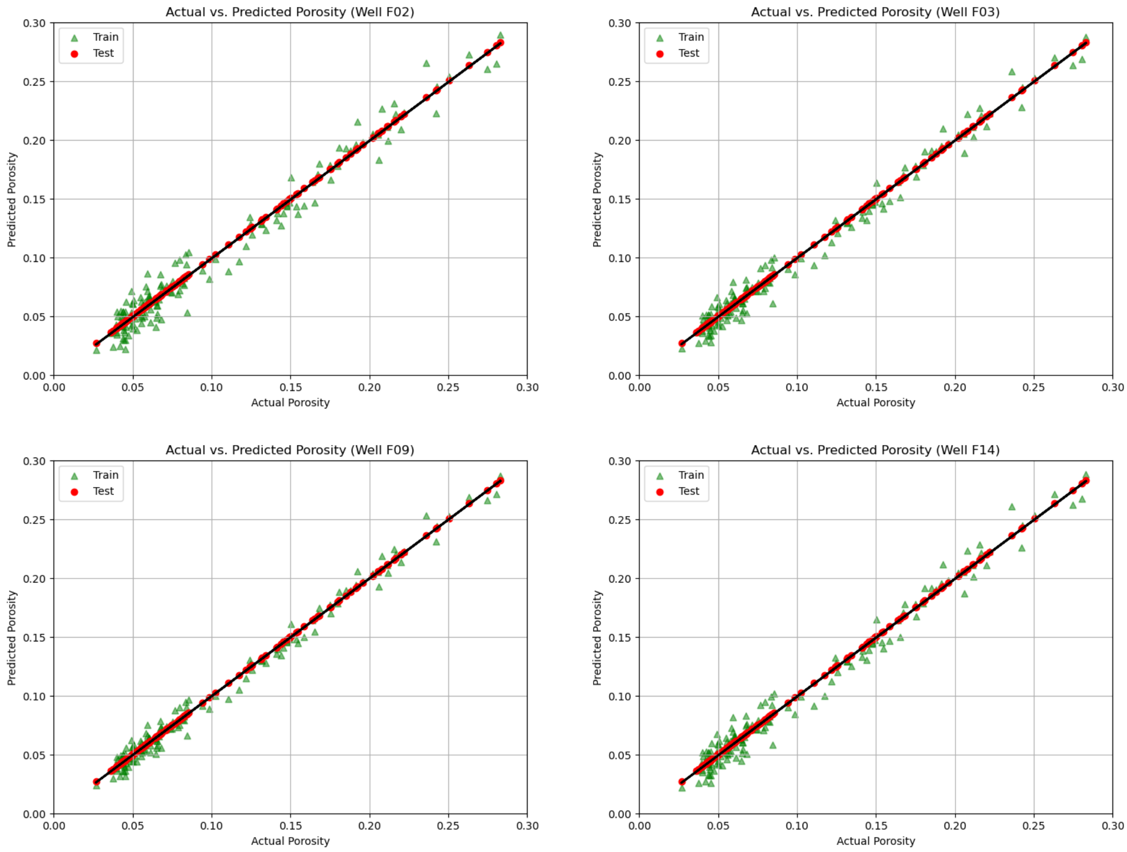


Figure 10. The cross plot of hybrid modeling prediction of Real Porosity versus Predicted Porosity in all wells based on core data.

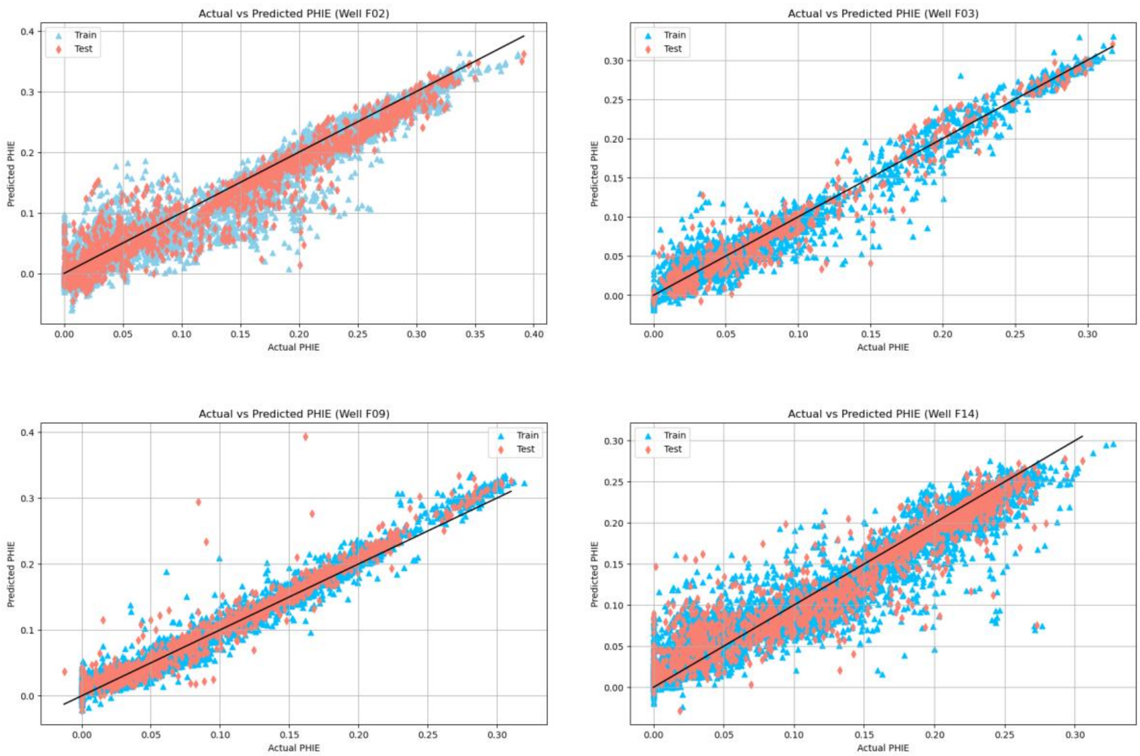


Figure 11. The cross plot of hybrid modeling prediction of Real PHIE versus Predicted PHIE in all wells based on log data.

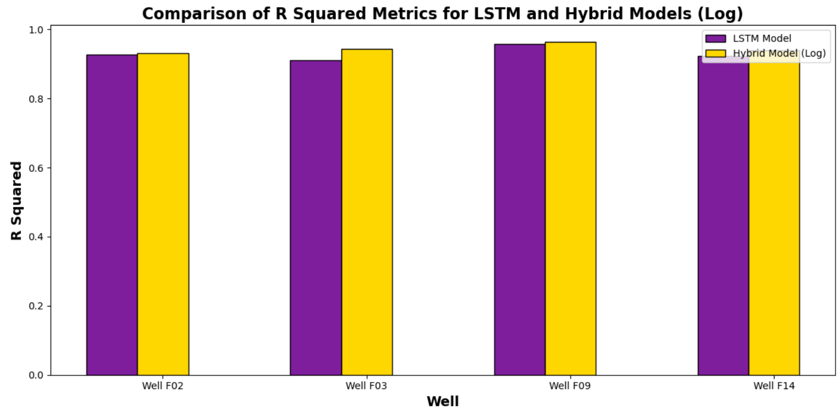


Figure 12. Visualization and comparison of R^2 metric of all wells (Log).

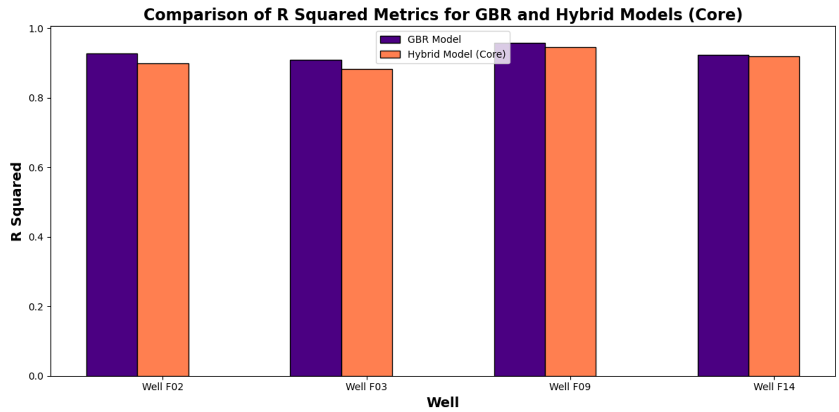


Figure 13. Visualization and comparison of R^2 metric of all wells (Log).

Regarding the prediction of PHIE by the well log-based LSTM model, Well F-9 and Well F-02 achieved the highest accuracy, with R^2 values of 0.958 and 0.927. Conversely, Well F-03 exhibited the lowest accuracy, with an R^2 value of 0.954. In this study, R^2 values remained above 0.90, suggesting that over 90% of the variability in porosity could be predicted from the models, which is a strong indicator of excellent model performance, given the complexity of the phenomena being modeled.

Additional evaluation metrics, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), were employed to assess the performance of the models. RMSE measures the average magnitude of errors with a bias towards large errors, while MAE provides the average error size, treating all errors equally. The corresponding results are illustrated in Figure 14 and 15.

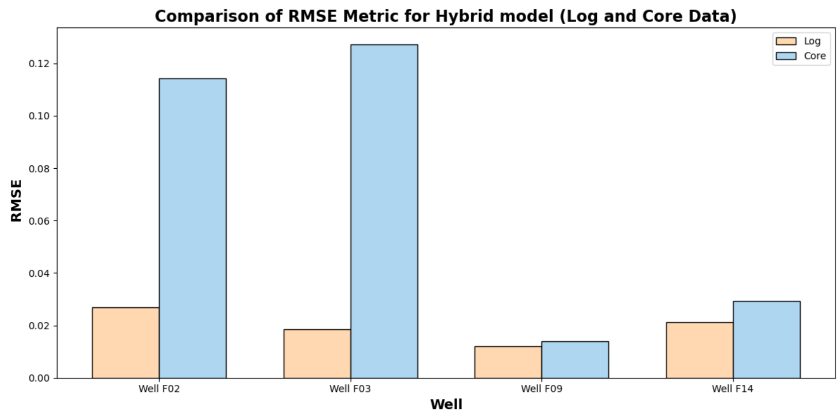


Figure 14. Visualization and comparison of RMSE metric of Hybrid model of all wells (Log and Core).

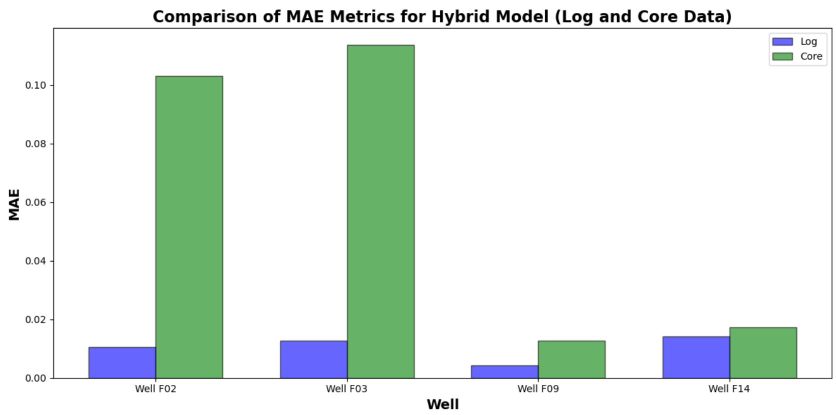


Figure 15. Visualization and comparison of MAE metric of Hybrid model of all wells (Log and Core).

When assessing predictive models, mean absolute error (MAE) and root mean square error (RMSE) emerge as pivotal measures. These metrics assess the typical size of errors by contrasting predicted values with observed outcomes. Effective models exhibit low MAE and RMSE values, indicating precise predictions and minimal discrepancies. Conversely, high MAE and RMSE values indicate inadequate performance, marked by inaccurate predictions and significant errors.

In the GBR model, due to the high number of missing data and lack of sufficient information, the precision is not as high compared to the results from log data. Well F-09 and Well F-14 achieved the highest accuracy, with R^2 values of 0.814 and 0.735. Conversely, Well F-02 exhibited the lowest accuracy, with an R^2 value of 0.716.

More details on R^2 values of the hybrid model from both datasets are depicted in earlier figures, showing that the combination of the models into a hybrid model has made the models more accurate and reliable in terms of porosity prediction.

Some statistical indices were also reported in Table 4 and Table 5 for further analysis of the models separately. These tables show the performance of the proposed models for the prediction of porosity using different metrics, including: root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2) and test loss.

Table 4. Statistical indices used for describing the performance of LSTM and GBR models for porosity estimation data

Model	Evaluation Metric	Well F02	Well F03	Well F09	Well F14
LSTM	Val_loss	0.001110	0.001179	0.000114	0.0003898
	R- Squared	0.92738	0.91038	0.95849	0.92314
	RMSE	0.02751	0.02317	0.01228	0.02156
	MAE	0.01952	0.01743	0.00516	0.01430
Gradient Boosting	R- Squared	0.72051	0.71632	0.81401	0.73521
	RMSE	3.16721	3.17940	2.14666	2.870223
	MAE	2.32964	2.26941	1.68402	1.26861

Table 5. Statistical indices used for describing the performance of Final Hybrid model for porosity estimation based on log and core data

Model	Evaluation Metric	Well F02	Well F03	Well F09	Well F14
Hybrid Model (Log Prediction)	R- Squared	0.93048	0.94292	0.96390	0.93892
	RMSE	0.02678	0.01849	0.01215	0.02115
	MAE	0.01053	0.01276	0.00436	0.01417
Hybrid Model (Core Prediction)	R- Squared	0.89958	0.88349	0.94567	0.91984
	RMSE	0.11430	0.12723	0.01393	0.02933
	MAE	0.10297	0.11367	0.01281	0.01721

Almost all deployed models exhibit encouraging outcomes individually, effectively fulfilling their designated tasks. The remarkable performance of the hybrid model can be credited to notable progress in amalgamating model techniques, resulting in enhancements across various critical facets. This hybrid approach has showcased exceptional learning capabilities and rapid training speeds.The proposed hybrid model demonstrates significant advantages in porosity prediction, offering enhanced accuracy and improved reliability by leveraging both well log and core data. Through the integration of ensemble techniques, such as the Gradient Boosting Regressor (GBR), and deep learning models designed for sequential data, such as LSTM, the model achieves superior accuracy compared to standalone models or the use of individual data sources. This enhanced accuracy stems from the hybrid approach's ability to effectively combine the strengths of diverse methodologies and datasets, ensuring comprehensive and precise predictions.

A key advantage of the hybrid model lies in its synergistic use of diverse data sources. Well log data, which provides continuous measurements along the well and captures critical spatial and sequential patterns, is adeptly combined with core data, which offers direct but spatially limited porosity measurements. This integration allows the model to exploit the complementary strengths of the two data sources, resulting in a more holistic representation of reservoir properties. Additionally, the LSTM component automates feature extraction from sequential well log data, effectively capturing significant spatial and temporal patterns without requiring extensive manual feature engineering.

The hybrid model also exhibits robustness against data anomalies, such as missing values and outliers, which are often inherent in well log and core datasets. The ensemble nature of the GBR component and the hybrid framework's capacity to handle irregularities enhances its resilience, ensuring reliable predictions even in the presence of challenging data conditions. Moreover, the incorporation of ensemble methodologies contributes to improved generalization and regularization, mitigating overfitting and enhancing the model's ability to perform accurately on unseen data. Beyond accuracy and robustness, the hybrid model enriches the interpretability of its predictions. By integrating diverse models and datasets, it provides a deeper understanding of the factors influencing porosity estimates, making the results more transparent and actionable. Furthermore, the hybrid architecture presents opportunities for discovering novel insights and relationships within the data that might remain hidden when using individual models or data sources in isolation. These insights not only enhance the current understanding of reservoir properties but also open avenues for further exploration and refinement of predictive methodologies.

While the proposed hybrid model for porosity prediction offers numerous advantages, it is important to acknowledge its potential limitations and challenges. One notable limitation is the heightened complexity introduced by the incorporation of various machine learning techniques, such as ensemble methods like GBR and deep learning models like LSTM. This increased complexity can pose challenges related to interpretability, computational demands, and model training optimization, making it more difficult to understand and deploy effectively. The model’s efficacy is also highly dependent on the quality and availability of data. Both well log and core data are critical to the model’s success, and inadequate, noisy, or biased data from either source can compromise its accuracy and robustness. This dependency highlights the necessity of high-quality and diverse datasets to ensure reliable performance. Additionally, the hybrid model’s computational requirements are considerable. Training and deploying the model, especially when working with large datasets or complex neural network architectures, can lead to significant computational overhead. This may limit its applicability in resource-constrained environments, where computational resources are scarce.

Another challenge lies in the complexity of parameter tuning and optimization. The hybrid model involves multiple components, each with its own set of hyperparameters, making the fine-tuning process intricate and time-intensive. Achieving the optimal configuration for all components, while ensuring their harmonious integration, requires extensive experimentation and domain expertise. Furthermore, despite efforts to enhance interpretability by integrating diverse models and data sources, the complexity of the hybrid architecture may still present difficulties in fully understanding and explaining the model’s predictions.

Finally, the successful implementation and refinement of the hybrid model demand substantial domain knowledge in areas such as petrophysics, well logging, and core analysis. This reliance on specialized expertise may limit the model’s accessibility and widespread adoption, particularly in contexts where domain knowledge is lacking. Addressing these limitations will be critical in refining the hybrid model and ensuring its broader applicability.

Future research directions could focus on mitigating these challenges by exploring methods to simplify the model architecture without compromising performance, improving data preprocessing techniques to handle noise and bias more effectively, and optimizing computational efficiency to reduce overhead. Additionally, developing automated tools for parameter tuning and enhancing interpretability through visualization techniques could help make the hybrid model more accessible to a wider audience. Efforts to incorporate domain knowledge into the model through expert-driven feature engineering or transfer learning approaches may also enhance its utility in practical applications.

4. Conclusions

In summary, this study introduces a comprehensive method for porosity prediction by integrating various machine learning techniques and exploiting complementary data sources. The hybrid model architecture proposed herein effectively amalgamates ensemble methods, deep learning models tailored for sequential data, and specialized approaches, culminating in a robust framework for porosity estimation. By combining a well log-based Long Short-Term Memory (LSTM) model with a Gradient Boosting Regression (GBR) model trained on core data, the hybrid architecture capitalizes on the complementary attributes of these constituents. The well log-based model captures spatial and temporal patterns from continuous well log measurements, while the core data model utilizes direct but limited porosity measurements. This harmonized fusion of diverse data modalities and modeling strategies has the potential to enhance predictive accuracy and resilience compared to conventional methodologies. Furthermore, the incorporation of LSTM networks in the well log-based component facilitates automatic feature learning and representation, obviating the necessity for explicit feature engineering and potentially unveiling significant patterns in the sequential well log data. The ensemble nature of the GBR components provides regularization, mitigating overfitting and enhancing generalization performance. The proposed hybrid model introduces several innovative aspects, including the integration of multiple machine learning techniques, the utilization of complementary data sources, the hybrid modeling paradigm, automatic feature learning and representation, and augmented robustness and generalization capabilities. These innovations differentiate the proposed approach from traditional methods and contribute to its potential for improved accuracy, adaptability, and resilience in porosity prediction tasks. Despite the considerable advantages offered by the hybrid model, it is imperative to acknowledge potential limitations, such as heightened model complexity, issues related to data quality and availability, computational demands, challenges associated with parameter tuning, interpretability issues, and dependencies on domain-specific knowledge. Addressing these limitations through continued research, development, and collaboration with domain experts is indispensable for the successful implementation and adoption of this approach in practical settings.

Data availability

The data will be available upon request. The corresponding author should be contacted for this purpose.

References

Abbas, M. A., Al-Mudhafar, W. J., & Wood, D. A. (2023). Improving permeability prediction in carbonate reservoirs through gradient boosting hyperparameter tuning. *Earth Science Informatics*, 16, 3417–3432. https://doi.org/10.1007/s12145-023-01099-0.

Ahmadi, M. A., Ahmadi, M. R., Hosseini, S. M., & Ebadi, M. (2014). Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: Application of artificial intelligence. *Journal of Petroleum Science and Engineering*, 123, 183–200.

Alyaev, S., & Elsheikh, A. H. (2022). Direct multi-modal inversion of geophysical logs using deep learning. *Earth and Space Science*, 9, e2021EA002186. https://doi.org/10.1029/2021EA002186.

Al-Khafaji, H., Qingbang, M., Wahib, Y., Samer, W., Wakeel, H., Ahmed, K. A., Fayez, H., & Ghamdan A. (2024). Advanced porosity prediction in heterogeneous oil reservoirs: Using novel machine learning and deep learning techniques. In J. Lin (Ed.), Proceedings of the International Field Exploration and Development Conference 2023. IFEDC 2023. Springer Series in Geomechanics and Geoengineering (pp. 123-138). Springer, Singapore. https://doi.org/10.1007/978-981-97-0479-8_45.

Al Shalabi, L., & Shaaban, Z. (Eds.). (2006). Normalization as a preprocessing engine for data mining and the approach of preference matrix. In 2006 International Conference on Dependability of Computer Systems. IEEE.

Bittar, M., Wang, S., Wu, X., & Chen, J. (2021). Multiple Well-Log Depth Matching Using Deep Q-Learning. Petrophysics, 62, 353–361. https://doi.org/10.30632/PJV62N4-2021a1.

Carey, C., Boucher, T., Mahadevan, S., Bartholomew, P., & Dyar, M. (2015). Machine learning tools for mineral recognition and classification from Raman spectroscopy. *Journal of Raman Spectroscopy*, 46(10), 894–903.

Dargi, M., Khamehchi, E., & Mahdavi, K. J. (2023). Optimizing acidizing design and effectiveness assessment with machine learning for predicting post-acidizing permeability. *Scientific Reports*, 13(1), 11851. https://doi.org/10.1038/s41598-023-38395-0.

Esteghlal, S., Samadi, S. H., Hosseini, S. M. H., & Moosavi-Movahedi, A. A. (2023). Identification of machine learning neural-network techniques for prediction of interfacial tension reduction by zein based colloidal particles. *Industrial & Engineering Chemistry Research*, 36, 106546. https://doi.org/10.1021/acs.iecr.3c00956.

Fernando, J. (2023, April 08). R-Squared. Investopedia. https://www.investopedia.com/terms/r/r-squared.asp.

Fu, L.-Y. (2003). An Information Integrated Approach for Reservoir Characterization. In Reservoir Characterization (pp. 177-192). Springer. https://doi.org/10.1007/978-94-017-0271-3_11.

Fu, L., Lin, T., Li, W., & Ma, S. (2022). Machine learning in petrophysics: Advantages and limitations. *Artificial Intelligence in Geosciences*, 3, 157-161.

Gamal, H., & Elkatatny, S. (2022). Prediction model based on an artificial neural network for rock porosity. *Arabian Journal for*

Science and Engineering, 47(9), 11211-11221.
Gupta, A., Pandey, A., Kesarwani, H., Sharma, S., & Saxena, A. (2022). Automated determination of interfacial tension and contact angle using computer vision for oil field applications. *Journal of Petroleum Exploration and Production Technology*, 12(5), 1453-1461. https://doi.org/10.1007/s13202-021-01313-0.
Hadavimoghaddam, F., Ostadhassan, M., Sadri, M. A., Bondarenko, T., Chebyshev, I., & Semnani, A. (2021). Prediction of Water Saturation from Well Log Data by Machine Learning Algorithms: Boosting and Super Learner. *Journal of Marine Science and Engineering*, 9, 666. https://doi.org/10.3390/jmse9060666.
Hassanzadeh, P., Hezarkhani, A., Rabbani, A. R., & Khajooie, S. (2019). Integrating geochemical and reservoir engineering approach to evaluate reservoir continuity: A case study from Foroozan field, Offshore Iran. *Journal of Earth System Science*, 128, 83. https://doi.org/10.1007/s12040-019-1105-x.
Hussain, W., Miao. L., Muhammad. A., Syed. M. H., Sajid. A., Sartaj. H., Asim. F. N., & Saddam. H. (2023). Machine learning - a novel approach to predict the porosity curve using geophysical logs data: An example from the Lower Goru sand reservoir in the Southern Indus Basin, Pakistan. *Journal of Applied Geophysics*, 214, 105067. https://doi.org/10.1016/j.jappgeo.2023.105067.
Hussain, W., Ehsan, M., Pan, L., Wang, X., Ali, M., Din, S. U., Hussain, H., Jawad, A., Chen, S., Liang, H., & 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.
Jo, H., Cho, Y., Pyrcz, M. J., Tang, H., & Fu, P. (2021). Machine learning-based porosity estimation from spectral decomposed seismic data.
Jo, J.-M. (2019). Effectiveness of normalization pre-processing of big data to the machine learning performance. *Journal of Korea Institute of Electronic Communication Sciences*, 14(3), 547–552.
Kirch, A., Celaschi, Y. M., de Almeida, J. M., & Miranda, C. R. (2020). Brine–oil interfacial tension modeling: Assessment of machine learning techniques combined with molecular dynamics. *ACS Applied Materials & Interfaces*, 12(13), 15837-15843. https://doi.org/10.1021/acsami.0c02972.
Li, Z., Xie, Y., Li, X., & Zhao, W. (2021). Prediction and application of porosity based on support vector regression model optimized by adaptive dragonfly algorithm. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 43(9), 1073-1086.
Mahzad, M., & Bagheri, M. (2025). Predictive reconstruction of missing geological events and patterns in real-life 3D post-stack seismic images: A novel U-Net-based deep learning approach. *Carbonates and Evaporites*, 40(12). https://doi.org/10.1007/s13146-024-01050-4.
Mahzad, M., & Riahi, M. A. (2024). Reservoir characterization reimagined: A hybrid neural network approach for direct three-dimensional petrophysical property characterization. *Carbonates and Evaporites*, 39(67). https://doi.org/10.1007/s13146-024-00975-0.
Maniscalco, R., Fazio, E., Punturo, R., Cirrincione, R., Di Stefano, A., Distefano, S., Forzese, M., Lanzafame, G., Leonardi, G. S., Montalbano, S., Pellegrino, A. G., Raele, A., & Palmeri, G. (2022). The porosity in heterogeneous carbonate reservoir rocks: Tectonic versus diagenetic imprint—A multi-scale study from the Hyblean Plateau (SE Sicily, Italy). *Geosciences*, 12(4), 149.
Moosavi, N., Bagheri, M., Nabi-Bidhendi, M., & Heidari, R. (2023). Porosity prediction using Fuzzy SVR and FCM SVR from well logs of an oil field in south of Iran. *Acta Geophysica*, 71(2), 769-782.
Munir, M. N., 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. *Arabian Journal for Science and Engineering*, 48,

7867–7882. https://doi.org/10.1007/s13369-022-07523-9.

Nasseri, A., & Mohammadzadeh, M. (2017). Evaluating distribution pattern of petrophysical properties and their monitoring under a hybrid intelligent-based method in southwest oil field of Iran. *Arabian Journal of Geosciences*, 10(10). https://doi.org/10.1007/s12517-016-2766-2.

Pan, J., Zhuang, Y., & Fong, S. (Eds.). (2016). The impact of data normalization on stock market prediction: using SVM and technical indicators. In Soft Computing in Data Science: Second International Conference, SCDS 2016, Kuala Lumpur, Malaysia, September 21–22, 2016, Proceedings 2. Springer.

Shiri, Y., Moradzadeh, A., Shiri, A., & Chehrazi, A. (2011). Application of adaptive Neuro-Fuzzy inference system for prediction of porosity from seismic attributes; case study, Farour.A oil field, Persian Gulf, Iran. *Journal of Seismic Exploration*, 20, 177-192.

Stanton, J. M. (2001). Galton, Pearson, and the Peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3). https://doi.org/10.1080/10691898.2001.11910531.

Subasi, A., El-Amin, M. F., Darwich, T., & Mubarak. D. (2022). Permeability prediction of petroleum reservoirs using stochastic gradient boosting regression. *Journal of Ambient Intelligence and Humanized Computing*, 13, 3555–3564. https://doi.org/10.1007/s12652-020-01986-0.

Sun, J., Zhang, R., Chen, M., Chen. B., Wang. X., Li. Q., & Ren. L. (2021). Identification of Porosity and Permeability While Drilling Based on Machine Learning. *Arabian Journal for Science and Engineering*, 46, 7031–7045. https://doi.org/10.1007/s13369-021-05432-x.

Sun, Y., Pang, S., Zhang, J., & Zhang, Y. (2024). Porosity prediction through well logging data: A combined approach of convolutional neural network and transformer model (CNN-transformer). *Physics of Fluids*, 36(2).

Tam, T. N. T., & Thanh, D. H. T. (2023). Estimate Petrophysical Properties by Using Machine Learning Methods. In Vo, P. L., Tran, D. A., Pham, T. L., Le Thi Thu, H., & Nguyen Viet, N. (Eds.), Advances in Research on Water Resources and Environmental Systems. GTER 2022. Environmental Science and Engineering. Springer, Cham. https://doi.org/10.1007/978-3-031-17808-5_29.

Talebkeikhah, M., Sadeghtabaghi, Z., & Shabani, M. (2021). A comparison of machine learning approaches for prediction of permeability using well log data in the hydrocarbon reservoirs. *Journal of Human Earth Future*, 2(2), 82–99.

Tiab, D., & Donaldson, E. C. (2016). Chapter 3 - Porosity and Permeability. In D. Tiab & E. C. Donaldson (Eds.), Petrophysics (4th ed., pp. 67-186). Gulf Professional Publishing.

Wu, L., Dong, Z., Li, W., Jing, C., & Qu, B. (2021). Well-Logging Prediction Based on Hybrid Neural Network Model. Energies, 14, 8583. https://doi.org/10.3390/en14248583.

Yousefmarzi, F., Haratian, A., Mahdavi Kalatehno, J., & Keihani Kamal, M. (2024). Machine learning approaches for estimating interfacial tension between oil/gas and oil/water systems: a performance analysis. Scientific Reports, 14, 858. https://doi.org/10.1038/s41598-024-51597-4.

Yu, S., Zhu, K., & Diao, F. (2008). A dynamic all parameters adaptive BP neural networks model and its application on oil reservoir prediction. Applied Mathematics and Computation, 195, 66-75. https://doi.org/10.1016/j.amc.2007.04.088.

Zamani, M. G., Nikoo, M. R., Rastad, D., & Nematollahi, B. (2023). A comparative study of data-driven models for runoff, sediment, and nitrate forecasting. *Journal of Environmental Management*, 341, 118006. https://doi.org/10.1016/j.jenvman.2023.118006.

Zou, Y., Chen, Y., & Deng, H. (2021). Gradient Boosting Decision Tree for Lithology Identification with Well Logs: A Case Study of Zhaoxian Gold Deposit, Shandong Peninsula, China. *Natural Resources Research*, 30, 3197–3217.

https://doi.org/10.1007/s11053-021-09894-6.