ZIBELINE INTERNATIONAL™ PUBLISHING

Engineering Heritage Journal (GWK)

DOI: http://doi.org/10.26480/gwk.01.2025.01.06

ISSN: 2521-0904 (Print)
ISSN: 2521-0440 (Online)
CODEN: EHJNA9

CrossMark

RESEARCH ARTICLE

MACHINE LEARNING APPLICATION FOR PREDICTION OF POROSITY AND PERMEABILITY LOGS: A CASE STUDY OF O-W FIELD NIGER DELTA

Osisanya Olajuwon Wasiu[a], Eze Uchechukwu Stanley[b], Ogugu Augustine Abiodun[c], Uti Lawrence Oghenebrume[c]

[a] Department of Physics, University of Benin, Benin City, Edo state, Nigeria.
[b] Department of Applied Geology, Olusegun Agagu University of Science and Technology, Okitipupa, Ondo State 350002, Nigeria.
[c] Department of Petroleum Engineering and Geosciences, Petroleum Training Institute, Effurun, Nigeria.
Corresponding author email: wasiu.osisanya@uniben.edu

ARTICLE DETAILS

Article History:

Received 07 January 2025
Revised 15 February 2025
Accepted 10 March 2025
Available online 21 April 2025

ABSTRACT

Predicting the porosity and permeability of hydrocarbon reservoirs is a key part of figuring out how much fluid is retained and how much moves through them. However, the absence of conventional porosity logs often complicates these predictions due to factors like borehole instability and logging challenges. This research looks at how well three machine learning algorithms—Linear Regression (LR), Random Forest (RF), and Gradient Boosting (XGBoost)—can predict porosity and permeability from well-log data in the Niger Delta basin. The well-log data includes gamma ray, caliper, density, and compressional sonic logs. The goal of the study was to create and improve machine learning models that could guess the properties of a reservoir without using traditional porosity logs. To get better results, hyperparameters for RF and XGBoost were tweaked, and the models' work was checked using the coefficient of determination (R^2) on training, validation, and blind testing datasets. Results indicated that XGBoost and RF outperformed LR in both porosity and permeability predictions, with R^2 values reaching 0.94–0.95 for porosity and 0.98–0.99 for permeability in test data. Blind testing further confirmed the robustness of the models, achieving R^2 values of 0.99 for porosity and 0.999 for permeability. This study adds to what is known in the oil industry by showing how machine learning techniques can be used to accurately predict key reservoir properties. These techniques can be used as a reliable alternative to traditional log data when they are not available, like in the Niger Delta basin. Furthermore, accurate prediction of reservoir properties can optimize operations and reduce uncertainties.

KEYWORDS

Hydrocarbon formation, Hydrocarbon production, Well logging, Reservoirs, Niger Delta Terrain.

1. INTRODUCTION

Basic log analysis of well-log data is critical for predicting reservoir properties, especially when full-wave sonic data are available. Sonic velocities, derived from recorded sonic interval travel times, provide accurate, reliable, and continuous indicators of the formation's fluid retention capability and reservoir rock properties through various correlations. However, the absence of sonic logs, along with other essential log data, can significantly complicate the prediction of vital parameters such as porosity and permeability. These parameters are crucial for assessing fluid retention and transmissibility in hydrocarbon reservoirs. In the absence of such data, it can lead to the neglect of promising reservoirs in the field due to the difficulty of accurately estimating these properties.

Traditionally, porosity and permeability estimation relies on the availability of porosity logs. However, these logs are sometimes unavailable due to inherent logging problems or cost constraints. As a result, alternative methods have been sought, with machine learning techniques emerging as a promising solution. Several studies have explored the use of machine learning to predict porosity and permeability logs when conventional logging data are not available. For instance, some researchers proposed a workflow utilizing depth-normalized log and core data, trained with an Artificial Neural Network (ANN), and successfully predicted permeability in uncored wells (Eriavbe and Uzoamaka, 2019). A group researchers employed various machine learning methods—such as Multi-Layer Perceptron Neural Network (MLP), Radial Basis Function Neural Network (RBF), Support Vector Regression (SVR), decision tree (DT), and random forest (RF)—to predict permeability from well log data in hydrocarbon reservoirs (Talebkeikhah et al., 2021). Their findings revealed that artificial intelligence models outperform traditional methods for permeability estimation. Similarly, a group researcher utilized a Bayesian framework with two multi-layer perceptron neural networks to predict porosity and permeability, specifically addressing the challenge of the absence of density and neutron logs (Job et al., 2020). Their research showed that machine learning techniques provided significantly more accurate permeability estimations compared to conventional methods.

This study builds on these advancements by further investigating the potential of machine learning algorithms for predicting porosity and permeability logs without relying on traditional porosity logs. Specifically, we explore three machine learning models—linear regression, random forest, and gradient boosting (XGBoost)—trained using well-log data, including caliper, density, gamma ray, and compressional sonic data. The specific objectives of this research are: (1) to develop and evaluate machine learning models for predicting porosity and permeability logs, (2) to optimize the hyperparameters of the random forest and gradient boosting algorithms for more accurate predictions, (3) to assess model performance using the coefficient of determination (R^2) on training, validation, and blind testing datasets, and (4) to evaluate model

Quick Response Code	Access this article online	
	Website: www.gwk.com.my	**DOI:** 10.26480/gwk.01.2025.01.06

Cite the Article: Osisanya Olajuwon Wasiu, Eze Uchechukwu Stanley, Ogugu Augustine Abiodun, Uti Lawrence Oghenebrume (2025). Machine Learning Application for Prediction of Porosity and Permeability Logs: A Case Study of O-W Field Niger Delta. *Engineering Heritage Journal (GWK), 9(1): 01-06.*

generalization through blind testing on two additional wells for validation.

The results of this study are expected to demonstrate the effectiveness of the XGBoost model in predicting porosity and permeability logs with high accuracy, as indicated by the R^2 values. These parameters are essential for assessing the commercial viability of a reservoir and optimizing its development and management strategies. This research suggests that machine learning techniques can serve as a reliable proxy for estimating key reservoir parameters, particularly in regions like the Niger Delta basin, where the lack of log data is a common challenge due to borehole instability.

2. Location of The Study Area

The Niger Delta Basin is located in the southernmost tip of Nigeria, bordering the Atlantic Ocean, and stretches from approximately Longitude 300 00′E to 90 00′E and Latitude 40 3′ N to 50 20′ N. The 'OW' field is the research area. It is mostly an offshore field located within the Niger Delta (Lambert, 1981; Osisanya et al., 2023). The field comprises six wells: OW-1, OW-2, OW-4, OW-9, OW-10 and OW-11 with aligned distribution orientations enabling geological inference (Figure 1).

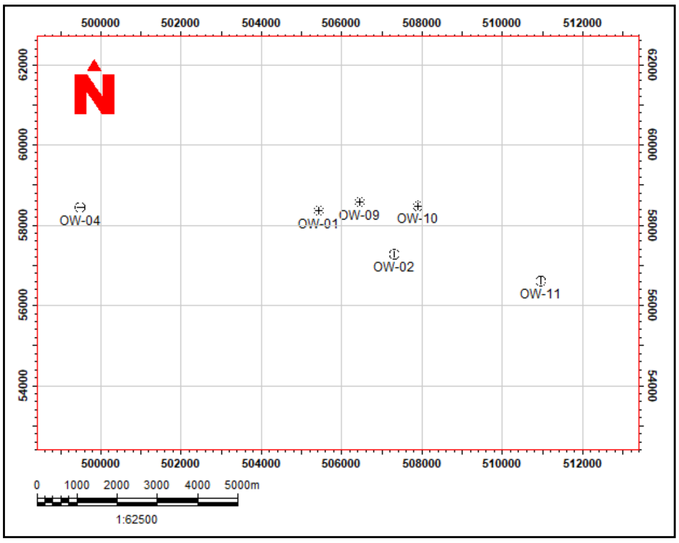

Figure 1: The base map of the Study area

The Niger Delta Basin is a sedimentary basin located in Nigeria and is known for its extensive hydrocarbon reserves (Onwuka et al., 2019). The basin is characterized by a thick sequence of sedimentary rocks, including sandstones, shales, and siltstones, which were deposited in a deltaic environment (Siren et al., 2018). The sedimentary rocks in the Niger Delta Basin have been found to contain high• quality source rocks, reservoir rocks, and seals, making it a prolific area for oil and gas production (Onwuka et al., 2019).

3. Lithostratigraphy of The Niger Delta

The Niger Delta Basin is a complex depositional environment that has undergone multiple phases of sedimentation, tectonic activity, and erosion (Figure 2). The lithostratigraphy of the Niger Delta Basin can be divided into several units based on their Akata Formation was deposited in a deep marine environment during the late Eocene to early Oligocene epoch. Between the Agbada Formation and the Akata Formation lies the Paralic and Deltaic Sandstones, which consist of sandstones, shales, and siltstones that were deposited in a transition zone between the deltaic and marine environments (Short and Stauble, 1967). The lithostratigraphy of the Niger Delta Basin has been studied extensively due to its importance for hydrocarbon exploration and production. Understanding the distribution and characteristics of the different lithostratigraphic units can help identify potential reservoir rocks and seals, as well as provide insights into the depositional history of the basin.. The Geology of the Niger Delta and its surroundings is a complex mixture of sedimentary rocks that have been deposited in a variety of environments over millions of years. This has resulted in a unique geological setting that is rich in petroleum resources and has significant economic and environmental importance (Maju and Osayande 2023).

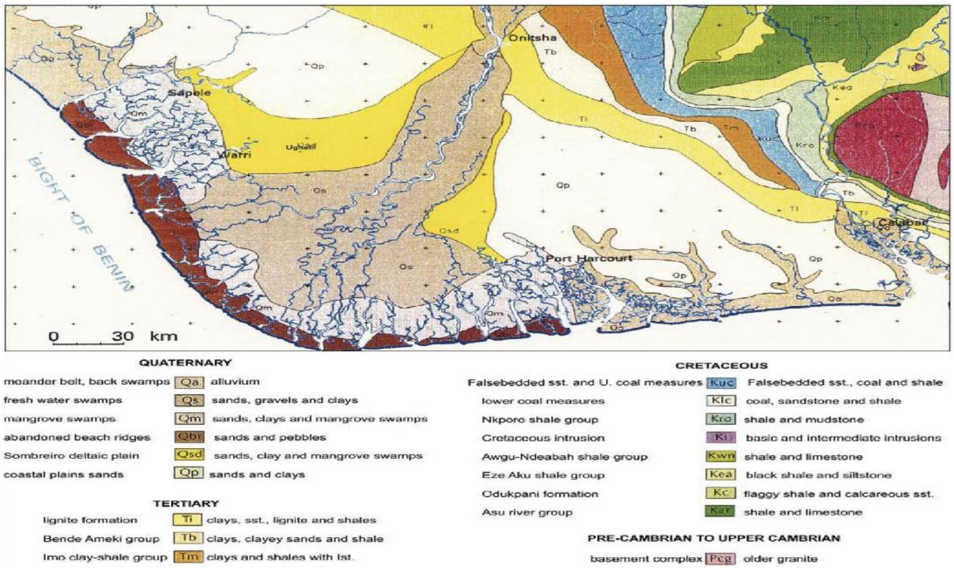

Figure 2: Geological map of Niger delta and its environs (modified from Reijers, 2011)

4. Materials and Methods

4.1 Dataset

Dataset used for this study includes well-log data such as caliper, density, gamma ray, deep resistivity, neuron, porosity, permeability, and compressional sonic logs (Table 1). These logs were procured from six (6) boreholes as shown in the base map of the study area. The dataset was pre-processed to remove any outliers, missing values, and noise to ensure a good data quality (Figure 3)

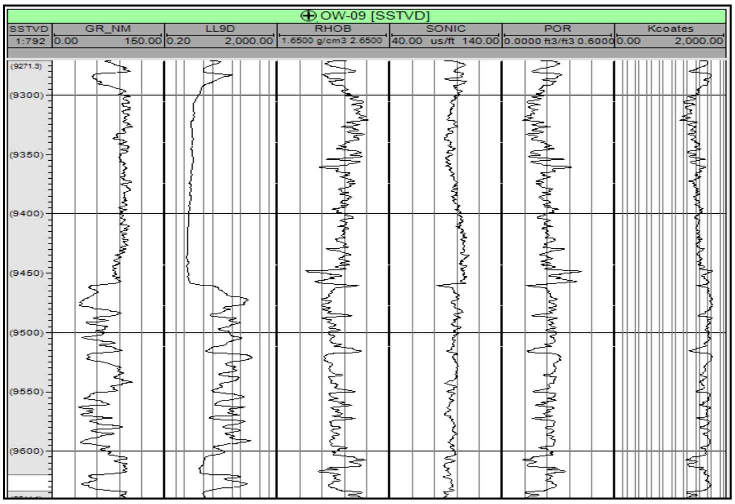

Figure 3: Typical well log display prior for conditioning Process prior to ML modeling

4.2 Data Splitting

The dataset was split into three subsets: a training set, a validation set, and a testing set. The training set was used for model development, the validation set for hyperparameter tuning, and the testing set for evaluating the final model's performance.

4.3 Feature Engineering

Feature engineering was performed to enhance the predictive power of the machine learning models. This involved transforming and scaling the input variables (well-logging curves) to ensure that the data was suitable for the modeling process. Feature selection techniques were also applied to identify the most relevant features for prediction.

4.4 Machine Learning Models

Three machine learning algorithms were selected for this study: linear regression, random forest, and gradient boosting (XGBoost). These models were chosen for their ability to handle both regression tasks (porosity and permeability prediction) and their capacity to capture complex relationships in the data.

4.5 Training and Hyperparameter Tuning

The selected machine learning models were independently trained using the training dataset. Hyperparameters of the random forest and gradient boosting models were fine-tuned to optimize their performance. Techniques like grid search and cross-validation were employed to find the best hyperparameters.

4.6 Model Evaluation

The performance of each model was assessed using the validation set. Metrics such as the coefficient of determination (R^2) were calculated to measure the models' ability to predict porosity and permeability accurately. The model that exhibited the best performance on the validation set was selected as the final model for both property predictions.

4.7 Blind Testing

To evaluate the generalization capability of the selected model, blind testing was conducted on two additional wells where porosity and permeability logs were not available. The model's performance was assessed on this unseen data using the same evaluation metrics.

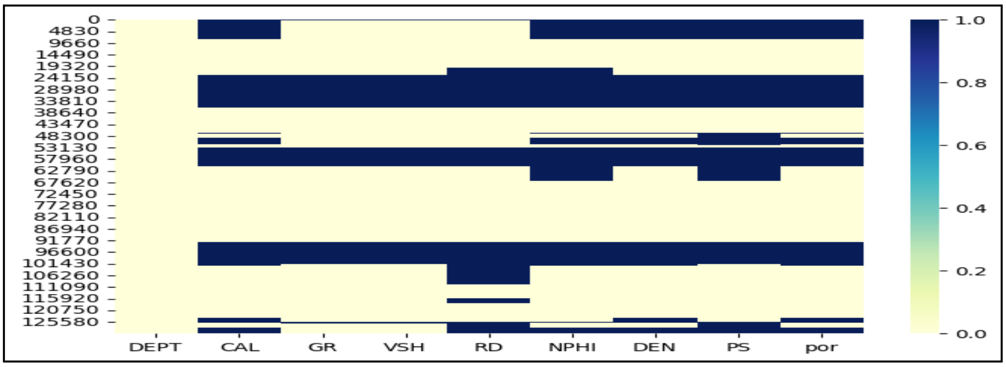

Figure 3a: Heat map showing before Data Cleaning

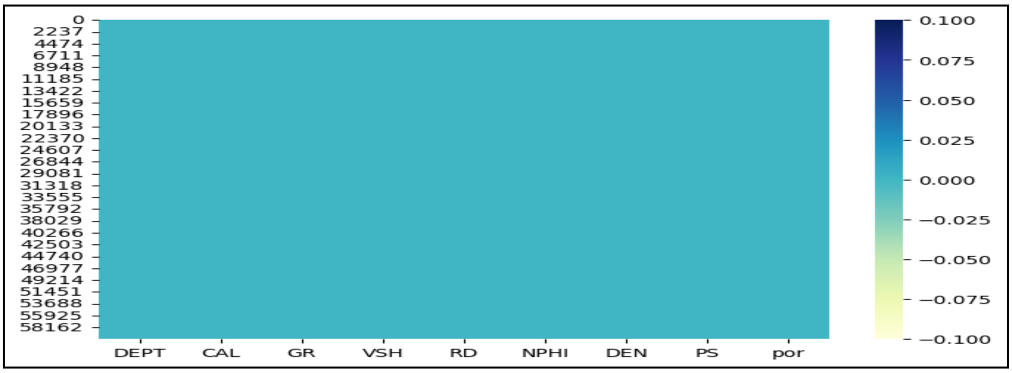

Figure 3b: Heat map showing after Data Cleaning

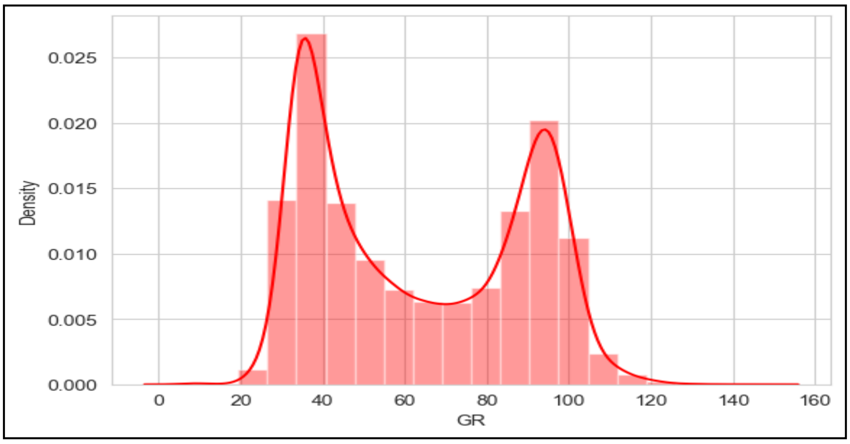

Figure 4a: Distribution plot of Gamma ray Log before transformation

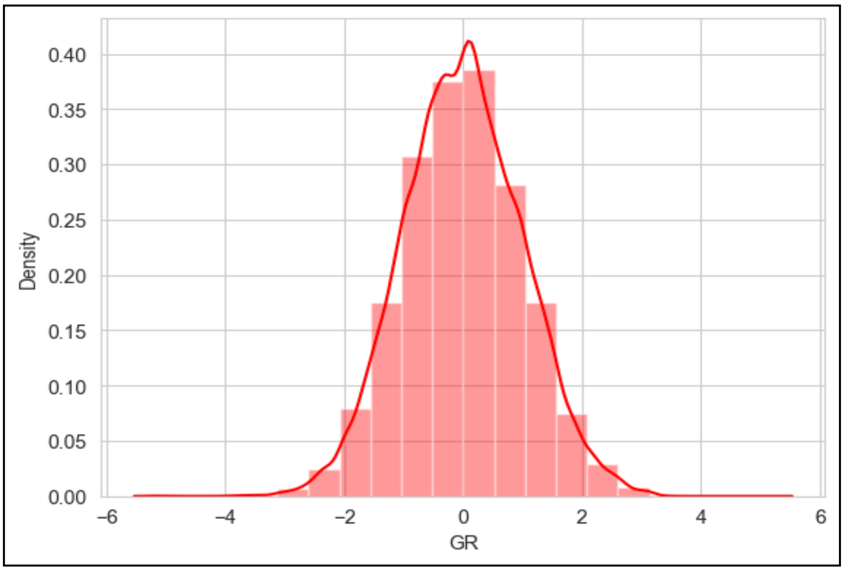

Figure 4b: Distribution plot of Gamma ray Log After transformation

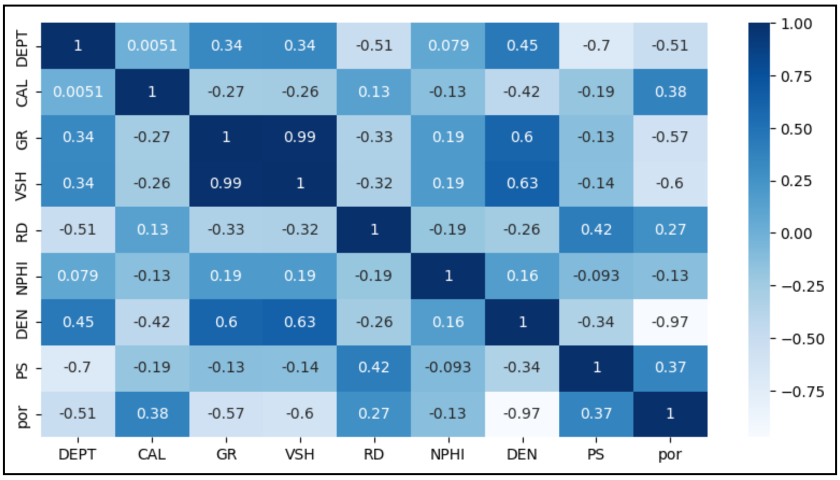

Figure 5: Typical Correlation Heat map showing the relationship between the features and targe

4.8 Presentation of Results

The machine learning (ML) models used for porosity and permeability predictions for training and testing data sets and their performance metrics are shown in Table 1 and 2 respectively.

Table 1: Performance Metrics of the ML Models used for Porosity

Target	Training Model	Training/Validation Set				Test Set			
		R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Porosity	**XGB**	0.9457	0.01006	0.00026	0.0164	0.9443	0.01011	0.00028	0.0167
	Random Forest	0.9392	0.011257	0.000301	0.0173	0.93702	0.01122	0.000317	0.0178
	Linear Regression	0.503	0.0349	0.00246	0.0496	0.5215	0.0353	0.00240	0.04908

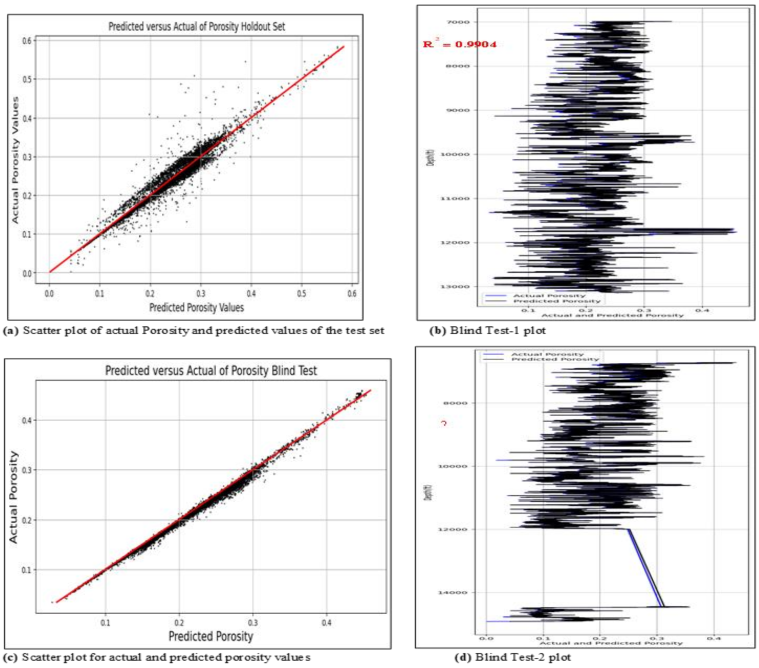

Figure 6: (a) Scatter plot of actual Porosity and predicted values of the test set. (b) Blind test-1 plot (c) Scattered plot for actual and predicted values (d) Blind test-2 plot.

Table 2: Performance Metrics of the ML Models used for Permeability prediction									
Target	**Training Model**	**Training/Validation Set**				**Test Set**			
		R^2	**MAE**	**MSE**	**RMSE**	**R^2**	**MAE**	**MSE**	**RMSE**
Permeability	**XGB**	0.990	1.316	32.473	5.6985	0.985	1.358	39.992	6.3239
	Random Forest	0.986	1.99	45.35	6.73	0.983	2.014	46.029	6.78
	Linear Regression	0.3463	20.019	2135.454	46.210	0.3805	19.064	1716.314	41.428

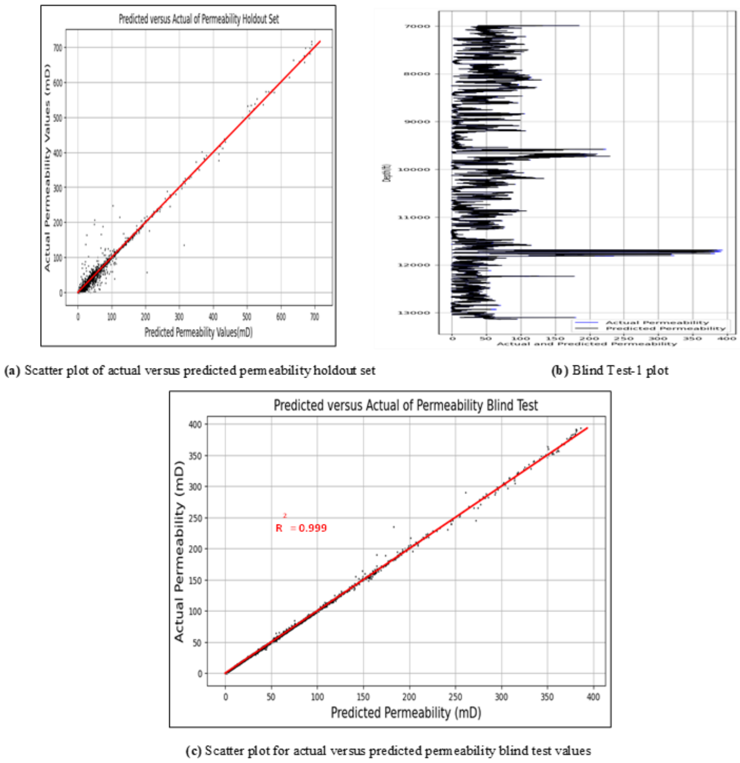

Figure 7: (a) Scatter plot of actual Permeability versus predicted holdout values (b) Blind test-1 plot (c) Scattered plot for actual versus predicted permeability blind test values.

5. Discussion of Results

In this present study, we aimed at incorporating different machine learning models, including linear regression (LR), random forest (RFR), and gradient boost (XGB), to predict the porosity and permeability values of a field within the Agbada Formation of the Niger Delta basin, where missing wireline logs have been a major issue. A comparative analysis of the three (3) models applied for these predictions was carried out using the optimal hyperparameters to ascertain the best predicting model and make recommendations for future studies.

Tables 1 and 2 show the results of the performance metrics calculated for the predictions of each model on the training, validation, and test sets for porosity and permeability values. The result suggests that the coefficient of determination (R^2) values of the LR, RFR, and XGB are 0.50, 0.94, and 0.95 in the porosity training data set, respectively (Table 1). At the same time, the R^2 values for LR, RFR, and XGB are 0.52, 0.94, and 0.94 in the porosity testing data set, respectively (Table 1).

Also, in Table 2, the R^2 values for LR, RFR, and XGB are 0.35, 0.99, and 0.99 in the permeability training data set, respectively, while the R^2 values for LR, RFR, and XGB are 0.38, 0.98, and 0.98 in the permeability testing data set, respectively. In both training and testing predictions of porosity, gradient boosting (XGR) and random forest (RFR) were observed to produce the best results in terms of accuracy (with the highest R^2 and lowest root mean square error-RMSE) in Table 1. It was also observed that linear regression (LR) performed poorly in both predictions with the lowest R^2 values of 0.503 and 0.5215 and the highest RMSE values of 0.0496 and 0.04908 training and testing data, respectively (Table 1).

Similarly, in both training and testing predictions of permeability, gradient boosting (XGR) and random forest (RFR) were observed to produce the best results in terms of accuracy (with the highest R^2 and lowest root mean square error-RMSE) in Table 2. It was also observed that linear regression (LR) performed poorly in both predictions with the lowest R^2 values of 0.3463 and 0.3805 and the highest RMSE values of 46.210 and 41.428 for the training and testing data sets, respectively (Table 2).

The results of the blind testing for the the porosity log are shown in Figure 6 a-d. It was observed that there was an almost perfect match between actual porosity and predicted porosity values from a shallow to a deeper depth. The scatter plots of actual porosity against the predicted porosity are shown in Figures 6a and 6c. The proposed ML model for porosity prediction is seen to have better improvement accuracy with a coefficient of determination ranging from 99.04 (R^2 = 0.9904) to 99.182% (R^2 = 0.99182), as shown in Figure 6.

The scatter plot of actual permeability against the predicted values is shown in Figures 7a-c. The result of the blind testing for permeability is shown in Figures 7 b and c. Similarly, the proposed ML model for porosity prediction is seen to have better improvement accuracy with a coefficient of determination value of 99.9% (R^2 = 0.999). This observation is an indication that the proposed model for the prediction of porosity and permeability values is not biased in its predictions. The results suggest that our proposed model can be accurately used for predicting the porosity and permeability log for the other wells where they are not available with a high level of confidence.

6. Conclusion

This study successfully demonstrates the application of machine learning models, specifically linear regression (LR), random forest (RFR), and gradient boosting (XGBoost), for the prediction of porosity and permeability logs in the Agbada Formation of the Niger Delta Basin, where missing well-log data has posed significant challenges. The results reveal that XGBoost and RFR outperformed LR in both training and testing phases, achieving high accuracy with coefficient of determination (R^2) values exceeding 0.94 for porosity and permeability predictions, as well as low root mean square error (RMSE) values. These findings suggest that machine learning approaches can effectively compensate for missing data in regions like the Niger Delta, where borehole instability often leads to absent logging curves. The XGBoost model, in particular, demonstrated its potential to produce highly accurate predictions, with R^2 values approaching 0.99 during blind testing. This highlights the generalization ability of the model and confirms its robustness in predicting reservoir properties in the absence of conventional porosity and permeability logs. The successful application of the proposed methodology across different wells further supports its potential for widespread use in hydrocarbon reservoir evaluation, especially in fields with limited log data. In conclusion, machine learning models such as XGBoost and random forest can serve as valuable tools for accurate porosity and permeability prediction, addressing data gaps in reservoir characterization. This study provides a reliable framework for future research and development in machine learning applications for subsurface reservoir analysis, particularly in complex geological settings like the Niger Delta. Future work may explore the integration of additional well-log types, refine model calibration techniques, and expand the scope of blind testing to enhance the accuracy and applicability of these models in diverse reservoir environments.

Recommendations

Further Expansion of Machine Learning Algorithms: While this study demonstrated the efficacy of XGBoost, Random Forest, and Linear Regression, it is recommended that future work explore the application of other machine learning models, such as Deep Learning (e.g., Convolutional Neural Networks or Recurrent Neural Networks), to determine if they offer additional improvements in predictive accuracy, especially for complex subsurface conditions.

Incorporating Additional Data Types: The current study utilized a specific set of well log data, but incorporating other types of geological, seismic, or petrophysical data (such as formation pressure or temperature) could enhance the robustness of the models. These additional data sets may provide further insights into the relationship between well log data and reservoir properties.

Long-Term Validation with More Wells: Although blind testing with two additional wells provided promising results, future studies could benefit from testing the model across a broader range of wells from different geographical regions, especially those with varying geological formations. This would assess the generalization capability of the models under diverse reservoir conditions.

Integration with Real-Time Data: Incorporating real-time well-log data from ongoing drilling operations into machine learning models can be an effective way to monitor and predict reservoir characteristics in real-time. This would enable operators to make more informed decisions about reservoir management.

Optimization of Hyperparameter Tuning Methods: Further exploration of advanced hyperparameter optimization techniques, such as Bayesian optimization or genetic algorithms, could be conducted to refine the model parameters further and potentially increase prediction accuracy for porosity and permeability.

References

Egbe, T., and Emudianughe J., 2024. Permeability and pore pressure prediction from well logs using machine learning: A study in the Niger delta. International Journal of Science and Research Archive, 13 (02), Pp. 2590–2611.

Eriavbe, F.E., and Uzoamaka, O.O., 2019. Machine Learning Application to Permeability Prediction Using Log and Core Measurements: A Realistic Workflow Application for Reservoir Characterization. Paper presented at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria. doi: https://doi.org/10.2118/198874-MS

Job, G.U., Ebong, D.E., Anthony, E.A., Emmanuel, I.A., 2020. A new approach for porosity and permeability prediction from well logs using artificial neural network and curve fitting techniques: A case study of Niger Delta, Nigeria. Journal of Applied Geophysics, Volume 183, 2020, 104207, https://doi.org/10.1016/j.jappgeo.2020.104207.

Maju-Oyovwikowhe, E.G., and Osayande, A.D., 2023. Hydrocarbon Evaluation and Distribution in well-X and well-Y in the Niger Delta basin: Findings and Validation through Porosity Comparison. Scientia Africana, 22 (1), Pp 255-278. 8 https://dx.doi.org/10.4314/sa.v22i1.22

Onwuka, S.U., Okolie, C.N., and Oshodi, O.O., 2019. The Niger Delta Basin: A Tectonic Overview. In: A. Akujieze (ed.) Advances in Petroleum Exploration and Production, IntechOpen.

Short, K.C., and Stauble, A.J., 1967. Outline of geology of Niger delta. AAPG bulletin, 51 (5), Pp. 761-779.

Sircar, A., Mukhopadhyay, S., and Mandal, A., 2021. Machine learning-based prediction of formation permeability from well log data. Journal of Petroleum Exploration and Production, 12 (4), Pp. 1677–1688.

Siren, G., Atakpo, E., and Ikwuetoghu, O.O., 2018. Sedimentary facies and depositional environments of the Niger Delta. In: H.C. Oyelami and O.O. Ikwuetoghu (eds.) The Niger Delta Region, Springer.

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, and Future, 2 (2), Pp. 82.