Research Article / Review Article FUPRE JOURNAL 8(3): 322-332(2024)

FUPRE Journal

of

Scientific and Industrial Research

ISSN: 2579-1184(Print) ISSN: 2578-1129 (Online)

http://fupre.edu.ng/journal

Machine Learning-Based Rock Facies Classification for Improved Reservoir Characterization in Niger Delta

IBOYITIE, O. [1,*], ***OKOLOGUME, C. W.*** [2], ***ONWUCHEKWA, C.*** [3],

OMO-IRABOR, O. O. [4]

[1,2] Department of Petroleum Engineering, Federal University of Petroleum Resources Effurun, Nigeria
[3] First Hydrocarbons Nigeria Ltd, Lagos State, Nigeria
[4]Department of Earth Sciences, Federal University of Petroleum Resources Effurun, Nigeria

ARTICLE INFO

Received: 02/02/2024
Accepted: 23/05/2024

Keywords

Characterization, Facies, Shoreface, Machine-learning, Reservoir, Rock

ABSTRACT

The Niger Delta, a cornerstone of Nigeria's oil and gas sector, plays a significant role in the Nation's energy landscape. This research concentrates on enhancing reservoir characterization, specifically emphasizing advancing rock facies classification. Employing advanced machine-learning methodologies and a dataset from 12 wells containing crucial well log parameters, such as Gamma Ray, Resistivity Micro-Spherical, Volume of Shale, Resistivity Deep, Resistivity Medium, Density, and Porosity, we conducted a rigorous evaluation of various classification models. The Random Forest algorithm emerged as the optimal choice, achieving an impressive F1 score of 0.93 and an accuracy of 0.93 on the cross-validation set. A meticulous analysis of identified facies classes, including Shale, Lower Shoreface, Middle Shoreface, Upper Shoreface, Transition Shoreface, Over Bank, and Channel, through confusion matrices, offered profound insights into the Model's efficiency. Feature importance analysis underscored the critical role of variables such as volume of shale, gamma ray, porosity, and bulk density in driving accurate predictions. This research significantly advances subsurface exploration in the Niger Delta, highlighting the effectiveness of machine learning for geologic characterization within the region's intricate geological landscape.

1. INTRODUCTION

In the energy exploration and production landscape, achieving operational excellence and optimizing resource utilisation are paramount goals, propelling the ongoing evolution of reservoir characterization methodologies. At the core of this process lies rock facies classification, where precise attribution of specific rock types to samples based on measured properties plays a foundational role (Dubois *et al.*, 2007). Accurate facies classification is crucial in seismic interpretation, as different rocks exhibit varying permeability and fluid saturation for a given porosity. In regions like the Niger Delta, achieving precise classification of rock facies poses formidable challenges due to complex geological

*Corresponding author, e-mail:ovieiboyitie@gmail.com
DIO

features (Opafunso, 2007). Traditionally, this endeavour has relied on manual interpretation by expert geologists, a process susceptible to subjectivity and time constraints.

The Niger Delta, spanning the Gulf of Guinea, holds a pivotal role in Nigeria's oil and gas sector, covering approximately 29,900 square kilometres of diverse landscapes (Opafunso, 2007). As the largest wetland in Africa, it supports abundant biodiversity and plays a crucial economic role. However, the region's distinct geological features, characterised by expansive wetlands and flat terrains, contribute to the inherent complexity of accurately predicting lithological formations (Opafunso, 2007).

In response to these challenges, machine-learning techniques have shown an increased prospect of addressing specific lithological issues in the Niger Delta. Machine learning's exceptional capacity to discern complex patterns, adapt to intricate data structures, and expedite decision-making presents an unprecedented opportunity to revolutionise rock facies classification (L'Heureux *et al*., 2017). Encouraged by the growth of big data and increased computational power, recent years have witnessed a renewed interest in machine-learning techniques within the geophysical community (Smith & Treitel, 2010; Zhang *et al*., 2014; Zhao *et al*., 2015; Kobrunov & Priezzhev, 2016).

In this study, a machine-learning model tailored for rock facies classification in the Niger Delta was developed. The study includes data collection, algorithm exploration and implementation, model optimisation, performance evaluation, and results analysis. By leveraging machine learning, this research aims to overcome the challenges posed by the region's geological complexities, ultimately contributing to enhanced reservoir characterisation and sustainable hydrocarbon recovery. The significance of this study is twofold. Firstly, it will provide a more objective and consistent approach to rock facies classification, reducing subjectivity and improving the reliability of reservoir characterisation. Secondly, it will enable better decision-making in petroleum engineering by facilitating accurate identification and understanding of subsurface geological formations. The successful implementation of machine learning techniques in rock facies classification will have practical implications for hydrocarbon recovery, operational efficiency, and reservoir management strategies.

2. MATERIALS AND METHODS

Our dataset comprises wireline log measurements extracted from 12 wells scattered across various fields within the Niger Delta region. These wells offer diverse geological conditions and formations, presenting a robust dataset for our analysis. The dataset encompasses wireline log measurements are Gamma ray (GR), Microresistivity (RES_MIC), Shale Volume (VSH), Deep Resistivity (RES_DEP), Medium Resistivity (RES_MED), Bulk Density (RHOB) and Porosity (PHIE).

These logs were vital indicators of lithological characteristics crucial for facies classification. Each well log was accompanied by depth intervals of half a foot, facilitating precise localization of measurements. Facie labels corresponding to these intervals were provided. The facies comprise seven distinct lithological facies prevalent in the Niger Delta, these include; Shale, Lower shoreface, Middle Shoreface, Upper shoreface, Transition shoreface, Overbank and Channel.

This study employed the petrolib library to process the dataset to load well-log data into a structured data frame, enabling efficient integration from multiple wells. Using a loop to iterate through well names, we loaded logs with 'pl.file_reader.load_las', aggregating data into a master dataframe. This streamlined approach ensured smooth data handling, setting the stage for comprehensive analysis.

```
# loading the required libraries
>>> import numpy as np
>>> import pandas as pd
>>> import seaborn as sns
>>> import lasio as las
>>> import petrolib as pl

# Creating a list that contains the
name of all the wells
>>> wells = ["Pake1", "Pake5",
"Pake6", "Pake9", "Pake23", "Pake11",
"Pake12", "Pake13","Pake14",
"Pake15","Pake16","Pake18"]

#Loading the well logs for all the
wells and saving them into one
dataframe
>>> df1, log_las =
pl.file_reader.load_las(wells[0],
return_csv= True)
>>> df1.reset_index(inplace = True)
>>> for well in wells[1:]:
    df2, log_las =
pl.file_reader.load_las(well,
return_csv= True)
    df2.reset_index(inplace = True)
    df1 = df2.append(df1)
>>> df1.reset_index(drop=True,
inplace = True)
>>> df1.head()
```

2.1 Data Pre-processing and Exploratory Data Analysis

After loading the well log data into a dateframe, Duplicate removal was conducted to ensure data integrity, outlier identification, and removal to enhance dataset robustness, and meticulous handling of missing values using sckit-learn iterative imputer. These steps collectively contributed to a clean, accurate, and complete dataset suitable for machine learning tasks. Exploratory data analysis was carried out on the data to gain insights into the data. Figure 1 shows the distribution of the rock facies in the data set. Upper shoreface emerges as the dominant rock facie, constituting approximately 59.08% of the dataset, while shale follows closely, representing around 21.50%. However, there is an imbalance in the representation of certain rock facies, Transition Shoreface, and Over Bank, comprising only 0.10% and 0.37% of the dataset respectively.

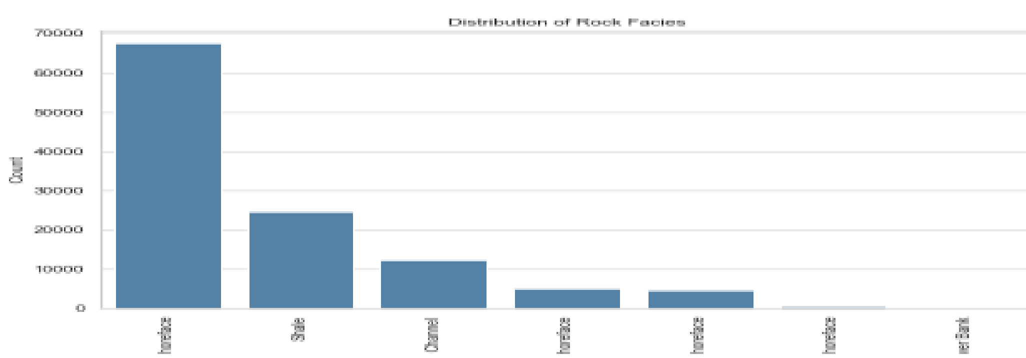

Figure 1: Distribution of Rock Facies

We can use `''data.describe()"` to provide a quick overview of the statistical distribution of the training data (Table 1). The count row in Table 1 indicates 114114 feature vectors in the dataset. Table 2 includes the descriptions linked to these classes. It should be noted that several of these facies do not have clear boundaries and instead transition gradually into one another. Misidentification of these adjacent facies is likely to happen. The Adjacent Facies column in Table 2 displays the associated classes.

Table 1: Dataset description and summary

	DEPT	GR	RES_MIC	VSH	FACIES	RES_DEP	RES_MED	RHOB	PHIE
count	114114	76858	76082	108174	114114	74596	90102	37285	55325
mean	5589.22	61.51	8.56	0.25	2.56	91.56	28.12	2.19	0.19
Std	2464.82	36.42	3.89	0.35	1.70	183.57	26.52	0.13	0.11
min	100.00	2.12	0.07	0.00	0.00	0.05	0.05	1.73	0.00
25%	3786.50	30.80	6.13	0.00	1.00	12.40	10.13	2.11	0.08
50%	5389.25	47.80	7.82	0.06	3.00	25.93	18.48	2.17	0.25
75%	6764.00	93.44	10.23	0.47	3.00	64.21	36.95	2.28	0.30
max	12419.5	160.60	35.36	1.00	6.00	1015.36	161.20	2.63	0.4

Table 2: Facies labels with their descriptions

Facie	Description	Adjacent Facies
0	Shale	-
1	Lower shoreface	4
2	Middle Shoreface	1,3
3	Upper shoreface	1,2
4	Transition shoreface	1,3
5	Overbank	-
6	Channel	5

The training data was standardized using Scikit-learn's 'StandardScaler class', ensuring consistency in quality. Additionally, a conventional approach of partitioning the dataset into training, testing, and cross-validation sets in a 48:12:40 ratio was adopted, employing the 'train test split' function from Scikit-learn to achieve this randomized division.

```
>>> from sklearn.preprocessing import StandardScaler

>>> scaler = StandardScaler()

>>> X = scaler. fit_transform (X)
```

*Corresponding author, e-mail:ovieiboyitie@gmail.com
DIO

```
>>> from sklearn.cross_validation import train_test_split

>>> X_train, X_CV, y_train, y_CV = train_test_split(X, y, test_size=0.4, random_state=42)

>>> X_train, X_test, y_train, y_test = train_test_split(X_train, y_train, test_size=0.2, random_state=42)
```

2.2 FACIES CLASSIFICATION ALGORITHM

The facies classification algorithms utilized in this work include common supervised machine learning algorithms, which are; Logistic Regression, Support Vector Machines (SVM), Decision Trees, K Nearest Neighbour (KNN), Random Forest, Single Layer Perceptron, and Artificial Neural Network (ANN),

Training involves using a collection of labelled data points from controlled wells to develop a Model, a function that links features to class labels. The classifier can predict class labels for new well logs based on unlabeled feature vectors after training.

3. RESULTS AND DISCUSSION

This section presents the results, which show the efficiency of the different models developed for classifying rock facies. The results are discussed, and possible remedial actions to improve the models' performance in classifying the facies are recommended.

3.1 Results

The results showing the performance of the models developed are shown in Tables 3 to 6 and Figures 2 to 4.

Table 3: Evaluation results for all the models - On the test set

Model	Test F1 Score	Test ROC AUC	Test Log Loss	Test Accuracy
Random Forest	0.935011	0.962579	0.257787	0.938841
Support Vector Machine	0.809340	0.902660	0.426820	0.845915
Logistic Regression	0.789094	0.855874	0.526079	0.822273
K-Nearest Neighbors	0.873271	0.836113	1.360678	0.879416
Decision Tree	0.911588	0.804393	3.201321	0.911182
Naive Bayes	0.828224	0.934143	0.752589	0.822547
Artificial Neural Network	0.893316	0.951886	0.273496	0.899133

Table 4: Evaluation results for all the models - On the Cross Validation (CV) set

Model	CV F1 Score	CV ROC AUC	CV Log Loss	CV Accuracy
Random Forest	0.926853	0.942965	0.273116	0.930991
Support Vector Machine	0.799597	0.910761	0.435781	0.837927
Logistic Regression	0.780818	0.867379	0.532323	0.815734
K-Nearest Neighbors	0.869217	0.845326	1.404324	0.876506
Decision Tree	0.904624	0.807707	3.434910	0.904701
Naive Bayes	0.823652	0.940975	0.774995	0.818319
Artificial Neural Network	0.888480	0.965677	0.278376	0.895018

Table 5: Accuracy metrics for the test set. – Random Forest classifier

Facies	precision	recall	f1-score	support
Shale	0.98	0.99	0.98	2343.00
Lower Shoreface	0.88	0.87	0.87	397.00
Middle Shoreface	0.89	0.86	0.88	484.00
Upper Shoreface	0.94	0.98	0.96	6523.00
Transition Shoreface	0.25	0.03	0.05	38.00
Overbank	0.00	0.00	0.00	14.00
Channel	0.84	0.71	0.77	1156.00
accuracy	0.94	0.94	0.94	0.94
macro avg	0.68	0.63	0.64	10955.00
weighted avg	0.93	0.94	0.93	10955.00

Table 6: Accuracy metrics for the Cross Validation (CV) set. – Random Forest classifier

Facies	precision	recall	f1-score	support
Shale	0.97	0.99	0.98	9751.00
Lower Shoreface	0.88	0.85	0.87	1852.00
Middle Shoreface	0.85	0.85	0.85	2020.00
Upper Shoreface	0.94	0.97	0.96	26835.00
Transition Shoreface	0.55	0.03	0.07	172.00
Overbank	0.33	0.02	0.05	41.00
Channel	0.84	0.69	0.76	4975.00
accuracy	0.93	0.93	0.93	0.93
macro avg	0.77	0.63	0.65	45646.00
weighted avg	0.93	0.93	0.93	45646.00

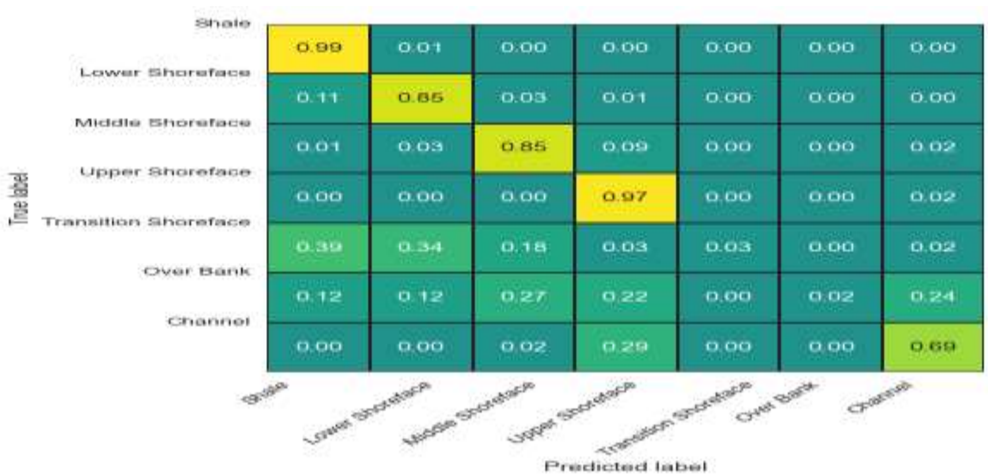

Figure 2: Confusion Matrix – Random Forest Classifier

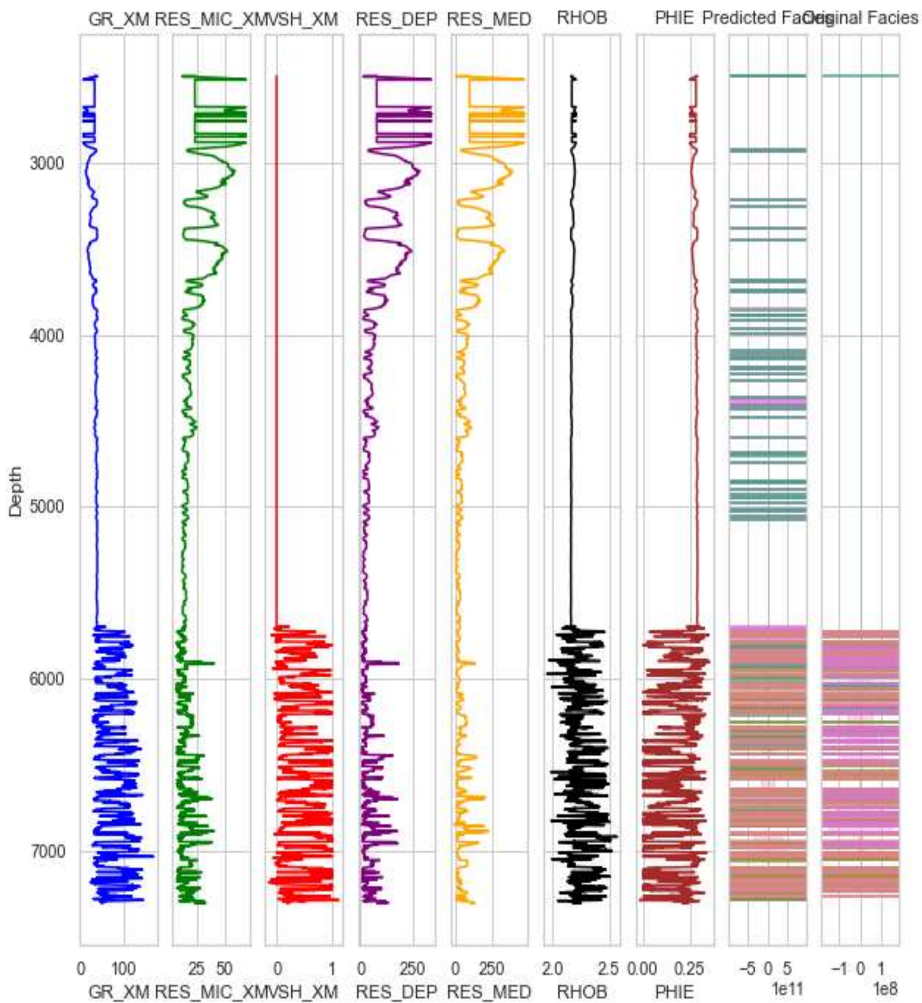

Figure 3: Well logs and facies classification results from a single well

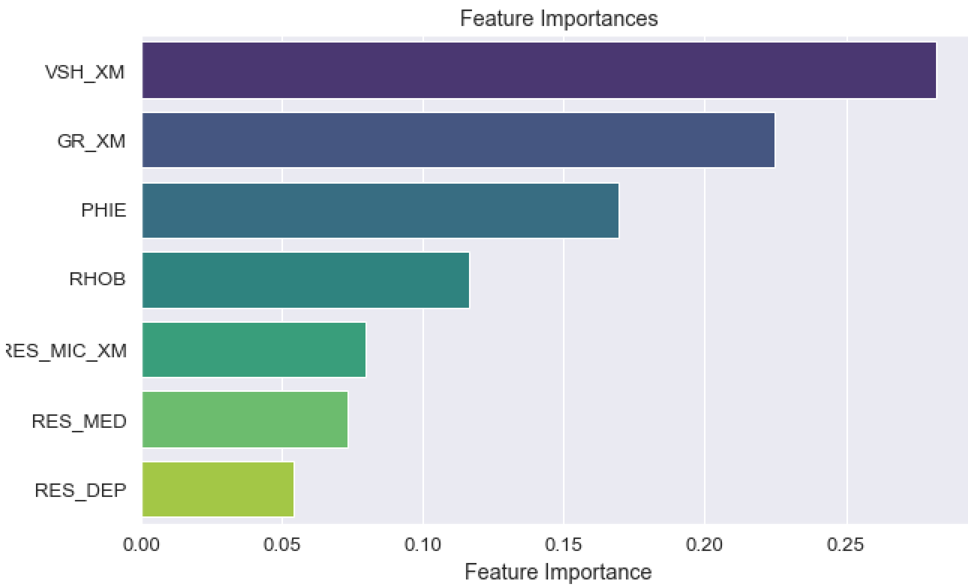

Figure 4: Feature importance

3.2 Discussion

The models were evaluated on the test and cross-validation sets using F1 score, test accuracy, log loss, and ROC AUC. In Table 3, a comprehensive overview of the performance metrics for various machine learning models employed in the facie classification is presented. The Random Forest algorithm demonstrates notable superiority across key evaluation criteria, with a Test F1 Score of 93.5% and a test accuracy of 93.9%. It showcases robust precision and recall capabilities. Additionally, the Model achieves a high-test ROC AUC of 0.963, indicating strong discrimination ability. Notably, the low-test log loss of 0.258 reflects the Model's confidence in its predictions. Furthermore, with a test accuracy of 93.9%, the Random Forest model emerges as the optimal choice for accurate facie classification.

In Table 5, a detailed assessment of the Random Forest classifier's efficiency in classifying various rock facies is presented. Shale and Upper Shoreface exhibited high precision and recall, signifying robust classification accuracy. However, Transition Shoreface and Overbank posed challenges, with lower precision and recall values indicating more frequent misclassifications. The confusion matrix, Figure 2, supported these results, with Shale instances predominantly accurately identified, contrasting with the struggles in classifying Transition Shoreface and Overbank. These observations underscored the complexity of facies classification, especially with adjacent facies where boundaries blur, as seen in the confusion matrix. For example, the Transition Shoreface's proximity to other facies, such as the Lower and Upper Shoreface, affected precision and recall scores.

Furthermore, the dataset's insufficient representation of Over Bank and Transition Shoreface poses a significant challenge to accurate classification. These facies represent only 0.10% and 0.37% of the entire dataset.

Such imbalances may hinder the Model's ability to discern distinct patterns and features associated with these facies, potentially leading to misclassifications or lower predictive performance. Addressing these challenges is significant for enhancing the Model's accuracy and ensuring more accurate geological interpretations. Visualising the Random Forest classifier results using log plots, Figure 3 is modeled after the descriptions found in Pake2 (Anonymous well) in the Niger Delta. It features the graphical representation of the seven logs used as features, displaying the actual (Manually determined) and predicted facies class logs.

The analysis of feature importance, shown in Figure 4, highlights key variables driving the Random Forest model's predictions for rock facies. VSH_XM (Volume of Shale) emerges as the most influential feature, with an importance value of 0.2823, followed closely by GR_XM (Gamma Ray) at 0.2246. Additionally, features like PHIE (Porosity) and RHOB (Bulk Density) contribute significantly, with importance values of 0.1696 and 0.1165, respectively. These findings underscore the Model's reliance on shale volume, gamma ray measurements, porosity, and bulk density information for accurate predictions, enhancing our understanding of the geological factors shaping the classification process.

3.2.1 Result Validation

The developed models were retested using the cross-validation dataset. The results are shown in Table 4 and Table 6. The Random classifier has an accuracy of 93.1% and an F1 score of 92.7%, as shown in Table 4. This closely matches the Model's performance on the test set.

4. CONCLUSION

The incorporation of machine learning algorithms presents a robust and effective approach for classifying rock facies in Niger Delta. The models are trained using the well logs to identify patterns within the dataset, including Gamma-ray, Microresistivity, Shale Volume, Deep Resistivity, Medium Resistivity, Bulk Density, and Porosity (PHIE). Models were developed using Logistic Regression, Support Vector Machines (SVM), Decision Trees, K Nearest Neighbour (KNN), Random Forest, Single Layer Perceptron, and Artificial Neural Network (ANN) algorithms, out of which Random Forest has the best results in classifying the rock facies. Despite its high accuracy in predicting the rock facies, it could have better classified the Overbank and Transition shoreface due to the imbalanced dataset and the Transition Shoreface's proximity to other facies, such as Lower Shoreface and Upper Shoreface. The volume of Shale, Gamma Ray, Porosity, and Bulk Density had the most influence on the random forest model's predictions.

The following recommendations are required to improve this research work;

1. More data, particularly for the underrepresented, should be obtained to improve the ability of the Model to classify them correctly.
2. Continuous Monitoring: Regularly recalibrate the Model using updated data to maintain relevance.

REFERENCES

Dubois, M. K., Bohling, G. C., & Chakrabarti, S. (2007). Comparison of four approaches to a rock facies classification problem. Computers &

Geosciences, 33(5), 599-617.Opafunso, Z. O. (2007). The Oil and Gas Industry and the Niger Delta: Implications for the Environment. J. Appl. Sci. Environ. Manage., 12(3), 29-37.

Ganguli, S. S., & Dimri, V. P. (2023). Reservoir characterisation: State-of-the-art, key challenges and ways forward. *Developments in Structural Geology and Tectonics*, *6*, 1-35. https://doi.org/10.1016/B978-0-323-99593-1.00015-X

Kobrunov, A., and I. Priezzhev, (2016). Hybrid combination genetic algorithm and controlled gradient method to train a neural network: Geophysics, 81, IM35–IM43

L'Heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. M. (2017). Machine Learning With Big Data: Challenges and Approaches. IEEE Access, 5, 7776-7797.

Smith, T., and S. Treitel, (2010). Self-organising artificial neural nets for automatic anomaly identification, in SEG Technical Program Expanded Abstracts 2010: Society of Exploration Geophysicists, 1403–1407.

Wolf, M., J. Pelissier-Combescure, et al., 1982, Faciolog automatic electrofacies determination: Presented at the SPWLA 23rd Annual Logging Symposium, Society of Petrophysicists and Well-Log Analysts.

Zhang, C., C., Frogner, M. Araya-Polo, and D. Hohl, (2014). Machine-learning based automated fault detection in seismic traces: Presented at the 76th EAGE Conference and Exhibition 2014.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, (2015). A comparison of classification techniques for seismic facies recognition: Interpretation, 3, SAE29–SAE58