

Performance Evaluation of Polymer Enhance Oil Recovery Using Machine Learning

Uchenna Alua, Anthony Ogbagbe Chikwe, Ifeanyi Oguamah, Matthew Chidubem Udechukwu, Anthony Kerunwa, Julian Ubanozie Obibuike, Chukwuebuka Francis Dike*, Federal University of Technology Owerri, Owerri, Nigeria

Abstract

Enhanced Oil Recovery (EOR) techniques utilizing environmentally-friendly and locally accessible biopolymers have attracted considerable attention as sustainable substitutes for synthetic polymers. This research conducts a comparative performance evaluation of four locally-sourced biopolymers, namely Okra (OK), Colocasia esculenta (CY), Xanthan Gum (XG), and Afzelia africana (AA), for EOR applications. Machine-learning models, particularly Random Forest (RF) and Extreme Gradient Boosting (XG-Boost), were utilized to forecast and optimize the displacement efficiency of the polymers. In all instances, XG-Boost outperformed RF, attaining regression accuracies (R^2) of 97.8%, 98.4%, 98.2%, and 98.9% for OK, CY, XG, and AA respectively. Feature importance analysis indicated that initial oil saturation, polymer viscosity, and polymer viscosity after adsorption were the most influential parameters, whereas salinity consistently exhibited minimal influence on EOR efficiency. The results demonstrate that the locally-sourced biopolymers can effectively compete with synthetic polymers in terms of displacement efficiency. Furthermore, the integration of machine-learning models in the evaluation process provides a robust framework for predicting and optimizing EOR performance. This approach not only reduces the need for extensive experimental trials but also offers insights into the key factors affecting the efficiency of biopolymer-based EOR techniques. The findings of this research have important implications for the oil and gas industry, particularly in regions where access to synthetic polymers is limited or where environmental regulations are stringent.

Introduction

There is always entrapped crude oil within the pore spaces of the reservoir rock after primary and secondary recovery (Gbadamosi et al. 2018), and this creates the opportunity for tertiary or enhanced oil recovery (Agi et al. 2022). Several enhanced oil recovery (EOR) techniques such as thermal, microbial, miscible and immiscible exist, but chemical EOR is the most utilized (Sedagahet et al. 2013). Chemical EOR deals with the introduction of alkali, surfactants, polymer and/or nanoparticles (Dike et al. 2024). In recent times polymer EOR have recorded global traction due to its effectiveness when utilized at field-scale levels (Agi et al. 2018). Polymer EOR deals with the introduction of polymer chemicals water to improve its viscosity and sweeping efficiency (Zerkalov et al. 2012). Conventional waterflood is plagued by viscous fingering” phenomenon due to the unfavorable mobility ratio with the entrapped crude oil (Abidin et al. 2012). To prevent this, polymer chemicals are introduced to boost the viscosity of the displacing fluid (Gbadamosi et al. 2019). Several polymer chemicals but xanthan gum (bio-based) and hydrolyzed polyacrylamide (synthetic-based) are the most utilized polymer (Ogolo et al. 2015). The polyacrylamide (PAM) has appreciable higher viscoelastic features when compared with xanthan gum (XG) in

Copyright © the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License.

Improved Oil and Gas Recovery

DOI: 10.14800/IOGR.1507

Received January 23, 2026; revised February 08, 2026; accepted February 24, 2026.

*Corresponding author: dike.chukwuebuka@futo.edu.ng

freshwater, but not in saline water (Sheng, 2013). Polymer CEOR is most suited for oil reservoir with viscosity (<100cp), temperature (<72°C) and low-moderate salt reservoir (Uzoho et al. 2015). The performance of the polymers can be evaluated in the laboratory using aqueous stability (Obuebite et al. 2018), polymer viscosity (Eiroboyi et al. 2019), polymer adsorption test (Dike et al. 2024) and polymer coreflood study (Ekene et al. 2024). Conventional polymers such as XG and PAM have recorded global traction in EOR but are not cost-friendly to the developing countries due to the importation cost. This has led to search for cheaper and suitable local options.

Abdulraheem et al. (2018) carried out a chemical EOR using natural gum arabic, modified gum arabic, xanthan gum and hengfloc. From the result of their evaluation modified gum arabic recorded the best performance with additional oil recovery of 41% while natural gum arabic, xanthan gum and hengfloc recorded 28.81%, 23.96% and 28.51% respectively. Obuebite et al. (2018) experimented on the EOR performance of ewedu, okro and editan. From the result of their study, Okro yielded the best EOR performance with additional recovery of 26.3%, while ewedu and editan recorded additional recoveries of 18.5% and 18.1% respectively. Fadairo et al. (2019) explored the EOR potentials of banana peel starch and gum arabic. From the result of their study, banana peel starch recorded maximum oil recovery of 64.5% while gum arabic recorded 62.6%. Uoho et al. (2019) explored the EOR prospect of *irvingia gabonensis*, *abelmoschus esculentus*, *brachystegia eurycoma*, *detarium microcarpium* and *mucuna flagellipes* respectively. From the result of their study, local reagent showed EOR prospects with maximum displacement efficiency of 88% and above. Ekene et al. (2024) carried out an EOR study using *afelia africana*, *colocasia* and hydroxyethyl cellulose. From the result of their experimental research, *afelia africana* and *colocasia esculenta* compared favorably with hydroxyethyl cellulose. Dike et al. (2024) conducted an EOR experimental study using cook pine gum, *terminalia mantaly* gum, xanthan gum and *afelia africana*. From the result of their experimental study, cook pine gum recorded the best EOR performance with 17.65% additional recovery while *terminalia mantaly*, xanthan gum and *afelia africana* recorded 16.67%, 15% and 14.5% additional oil recoveries respectively. However, the effectiveness of these local polymers in their deployment to ascertain the conditions that supports their utilization.

In this study, performance evaluation of polymers; *abelmoschus esculenta*, *colocasia esculenta*, xanthan gum and *afelia africana*, was carried out using machine learning algorithm. Machine learning models and feature importance were developed for each polymer flooding scenario.

Methodology

The dataset used in the study comprises of seven input parameters (denoted $X = X_1, X_2, \dots, X_7$), which represent the various parameters that impacts on the performance of four polymers: *abelmoschus esculenta* (OK), *colocasia esculenta* (CY), xanthan gum (XG) and *afelia africana* (AA), in polymer flood operations. The input parameters include the initial oil saturation, residual oil saturation after water-flooding, rock permeability, polymer viscosity, polymer concentration, polymer viscosity after adsorption and salinity, while the target variable is the displacement efficiency.

Data Preprocessing and Resampling. The data preprocessing and resample approach comprised of target imputation, feature scaling, re-sampling with Latin Hypercube Sampling (LHS) and data split and cross-validation.

Target Imputation (Resolving Missing Data). The missing data in the dataset were addressed using a supervised imputation approach. Specifically, a Gradient Boosting Regressor (GBR) model was utilized to predict the missing target values based on the features in the dataset using the process mathematically expressed as

$$\hat{Y}_i = f(X_i; \theta), \dots \dots \dots (1)$$

where $\hat{Y}_i$ is the imputed displacement efficiency for the i-th sample; X_i represents the feature values associated with the i-th sample; θ represents the parameters of the trained GBR model.

To train the GBR model, Mean Squared Error (MSE) was utilized as the loss function, which is minimized to find the best parameters,

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i; \theta))^2, \dots \dots \dots (2)$$

Once trained, the GBR model was used to predict the missing values of displacement efficiency, filling the gaps in the target variable.

Feature Scaling. After resolving the missing data, the features were standardized to ensure that they contributed equally to the model training process. Z-score normalization was conducted to standardize the features as depicted.

$$X'_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}, \dots \dots \dots (3)$$

where X'_{ij} is the standardized value of feature j for sample i ; X_{ij} is the original value of feature j ; μ_j and σ_j are the mean and standard deviation of feature j , respectively.

This normalization ensures that each feature has a mean of zero and a standard deviation of one, allowing the machine learning models to treat all features equally, avoiding any one feature from dominating due to its scale.

Resampling with Latin Hypercube Sampling (LHS). Latin Hypercube Sampling (LHS) was carried out to augment the dataset and mitigate class imbalance. LHS is a stratified sampling approach that ensures even distribution of sampled points across the feature space, minimizing clustering and maximizing the diversity of the generated sample. The goal of LHS is to sample n points $\{x_1, x_2, \dots, x_n\}$ such as,

$$\int_0^1 f(x) dx \approx \frac{1}{n} \sum_{i=1}^n f(x_i), \dots \dots \dots (4)$$

where $f(x)$ is the probability distribution function of the feature space. By sampling more evenly, LHS ensures that the augmented data points maintain the statistical properties of the original dataset.

Upon the generation of synthetic data points, the Gradient Boosting Regressor was employed to forecast displacement efficiency for the augmented rows. In this manner, the dataset was augmented while preserving the feature distributions.

Data-Split and Cross-Validation. After the data preprocessing and augmentation, the dataset was split into train set (model fitting) and test set (for evaluation). To ensure consistent performance of the model across several data partitions, 5-fold cross-validation was utilized. The process provided a more reliable estimate of model performance by averaging results across several partitions. It helps mitigate overfitting and ensures that the model is generalizable across unseen data.

Model Development. For the model development two machine learning algorithms, including Random Forest (RF) Regressor and Extreme Gradient Boost (XGB) Regressor, were utilized.

Random Forest Regression (RFR). The Random Forest Regression (RFR) model was developed using the RF algorithm class on the scikit-learn. The hyperparameters utilized for the RF model are as follows.

- (a) `n_estimators`: 600 (number of trees in the forest)
- (b) `max_depth`: None (trees are allowed to grow to full depth)
- (c) `min_samples_split`: 2 (minimum number of samples required to split an internal node)
- (d) `min_samples_leaf`: 1 (minimum number of samples required in leaf nodes)
- (e) `max_features`: sqrt (number of features to consider for each split)

The mean squared error (MSE) was used as the loss function during model fitting.

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i; \theta))^2, \dots \dots \dots (5)$$

where $f(X_i; \theta)$ represents the Random Forest model's prediction for the i th sample, and Y_i is the observed displacement efficiency.

Extreme Gradient Boost (XGB). The Extreme Gradient Boost (XGB) model was trained using the XGB algorithm class on the scikit-learn package. The hyper-parameters utilized for the XGB model are as follows

- (a) `n_estimators`: 600 (number of boosting rounds),
- (b) `max_depth`: 6 (maximum depth of each tree),
- (c) `learning_rate`: 0.1 (shrinkage parameter to prevent overfitting),
- (d) `subsample`: 0.8 (randomly sample 80% of the data for each tree),
- (e) `colsample_bytree`: 0.8 (sample 80% of features for each tree).

The objective function for XGBoost is expressed as

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i; \theta))^2 \dots\dots\dots (6)$$

XGBoost minimizes the loss function using gradient boosting, iteratively adding trees that correct the errors made by previous trees.

Model Evaluation. The developed model was rigorously evaluated by employing two widely recognized statistical metrics: the regression squared error and the mean absolute error. These error metrics serve as crucial indicators of the model's predictive accuracy, quantitatively assessing the extent to which the model's predictions deviate from the actual observed data. By analyzing both squared and absolute deviations, the evaluation provides a comprehensive understanding of the model's performance, ensuring a balanced perspective on both large and small prediction errors.

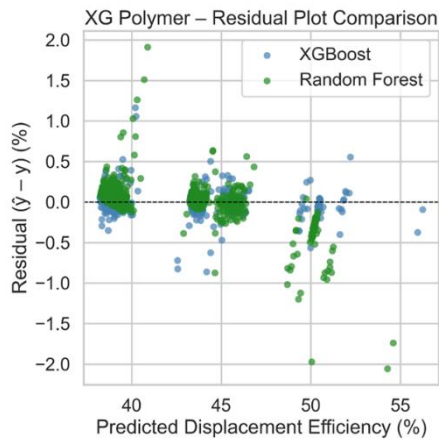
Feature Importance. The feature importance analysis was conducted to identify and evaluate which parameters exerted the greatest or most significant influence on the target variable. In the case of the Random Forest (RF) model, the assessment of feature contributions was performed using a percentage-based methodology, which quantifies the relative impact of each feature in terms of its proportional contribution to the model's predictive accuracy. Conversely, for the XGBoost (XGB) model, a weight-based approach was employed to determine feature importance, relying on the cumulative gains attributable to each feature across all splits in the boosted trees. This comparative framework ensures a comprehensive understanding of how different modeling techniques attribute importance to input variables, thereby offering deeper insights into the underlying drivers of the target outcome.

Results and Discussion

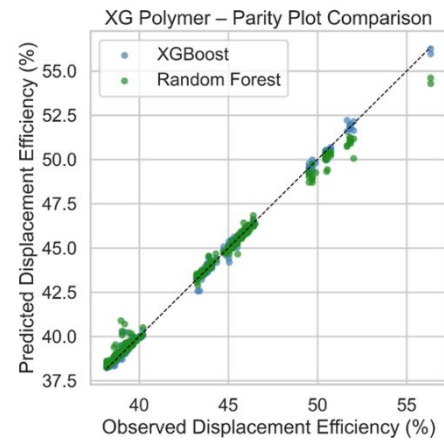
Model Development. Table 1 presents the goodness-of-fit statistics for the four polymers. The XG-Boost algorithm achieved optimal regression performance for all polymers, with regression rates of 97.8%, 98.4%, 98.2%, and 98.9% for OK, CY, XG, and AA, respectively. In contrast, the Random Forest (RF) algorithm recorded regression rates of 96.2%, 92.2%, 97.1%, and 96.3% for OK, CY, XG, and AA, respectively. This phenomenon can be observed through the lower Mean Absolute Error (MAE) in XG-Boost compared to Random Forest (RF) Machine Learning. Moreover, XG-Boost achieved the optimal regression performance at AA and the poorest regression performance at OK. RF attained the best regression for XG and the least favorable regression for CY. The performance of RF and XG-Boost can be clearly visualized in their residual plots and predicted-versus-actual plots. The CY polymer demonstrated a relatively high MAE yet still performed moderately well, with an R^2 value of 0.922. The XG polymer exhibited greater variability in MAE for RF Machine Learning, indicating a higher level of complexity in its prediction. The disparities in MAE and R^2 can be ascribed to the complexity of the polymers. The XG polymer shows a higher prediction variance, which might imply the necessity for additional data or more effective feature engineering for this specific polymer. **Figures 1 through 4** present detailed comparisons of the residual plots alongside the plots of predicted values against actual values for the four methods: XG, OK, CY, and AA. In the plots illustrating predicted values versus actual values, the proximity of data points to the trendline is a critical indicator of the model's predictive accuracy. A tighter clustering of points around this line demonstrates a stronger alignment between the predicted outcomes and the actual observed data, highlighting the reliability and precision of the modeling approach.

Table 1—Goodness-of-fit statistics for the various polymers.

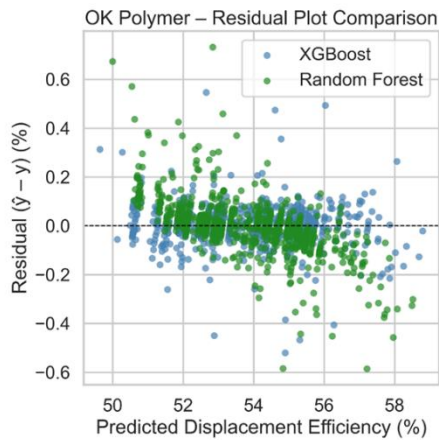
Polymer	Random Forest		XG-Boost	
	MAE	R ²	MAE	R ²
OK	0.184	0.962	0.119	0.978
CY	0.292	0.922	0.103	0.984
XG	0.325	0.971	0.153	0.982
AA	0.206	0.963	0.069	0.989



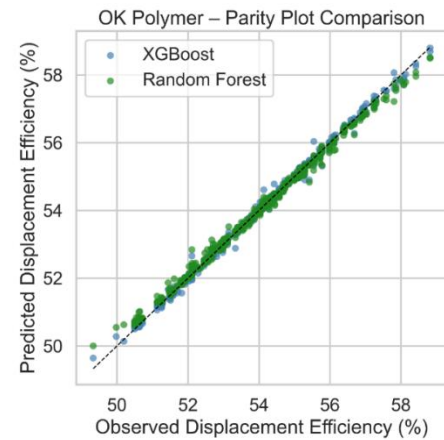
a) Residual Plot



b) Predicted vs Actual Plot

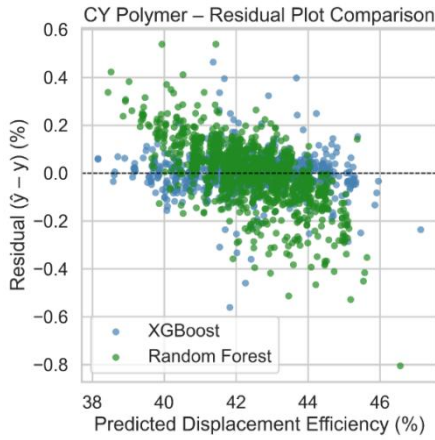
Figure 1—Model performance analysis for XG.

a) Residual Plot

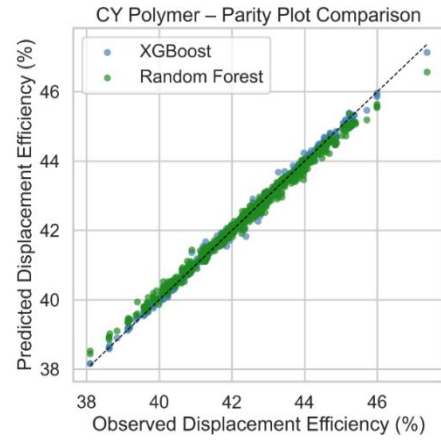


b) Predicted vs Actual Plot

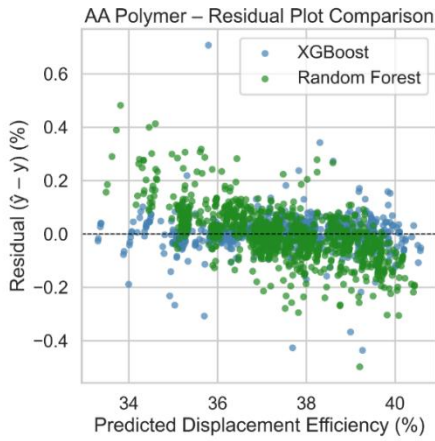
Figure 2—Model performance analysis for OK.



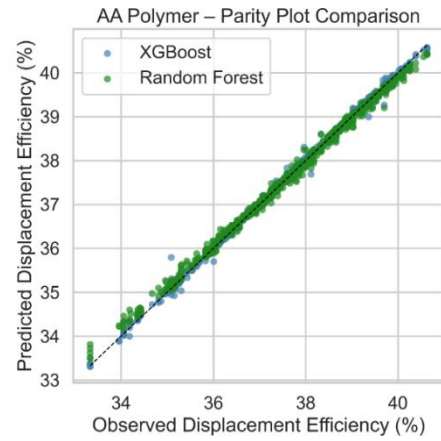
a) Residual Plot



b) Predicted vs Actual Plot

Figure 3—Model performance analysis for CY.

a) Residual Plot



b) Predicted vs Actual Plot

Figure 4—Model performance analysis for AA.

Feature Importance. Table 2 shows the percentage-based RF Feature Importance of the parameters that contribute to the displacement efficiency of the polymers. As observed, S_{oi} recorded the highest contribution to displacement efficiency with 55%, while salinity recorded the least contribution with 1.65% for OK. For CY polymer, polymer viscosity recorded the highest contribution to displacement efficiency with 37.46%, while salinity recorded the least contribution with 2.78%. For AA polymer, S_{oi} recorded the highest contribution to displacement efficiency with 40.8%, while salinity recorded the least contribution with 1.99%. For XG polymer, polymer viscosity after adsorption recorded the highest contribution to displacement efficiency with 56.12%, while residual S_w after water-flooding. Table 3 shows the weight-based XG-Boost ML Feature Importance of the parameters that contributed to the displacement efficiency of the polymer. As shown in Table 3, S_{oi} recorded the highest weight contribution to displacement efficiency with value of 20.0310, while salinity recorded the least weight contribution to displacement efficiency with value of 0.3162 for OK. For CY, polymer viscosity recorded the highest contribution with value of 12.004, while salinity recorded the lowest contribution with value of 0.06. For AA, S_{oi} recorded the highest contribution with value of 28.7082, while salinity recorded the least with value of 0.0319. For XG, polymer viscosity after adsorption recorded the highest contribution with value of 259.8826, while salinity recorded the least contribution with value of 1.3048. As a result, major emphasis must be focused on initial oil saturation, polymer viscosity and polymer viscosity after adsorption for optimal displacement

efficiency when utilizing OK polymer for EOR. For CY, emphasis on S_{oi} , polymer viscosity and core permeability to optimize displacement efficiency and EOR. For AA, emphasis on S_{oi} , polymer viscosity and polymer concentration to optimize displacement efficiency and EOR. For XG, emphasis on polymer viscosity, polymer viscosity after adsorption and polymer concentration to optimize displacement efficiency and EOR.

Table 2—Percentage-based feature importance for random forest ML.

Feature	Percentage-Based Feature Importance			
	OK	CY	AA	XG
S_{oi}	55.00	24.31	40.80	1.83
Polymer Viscosity	16.92	37.46	30.68	13.52
Polymer Viscosity After Adsorption	9.43	3.13	2.64	56.12
Polymer Concentration	7.63	8.68	9.46	23.12
Permeability	6.16	15.74	8.34	1.92
Residual S_w After Water Flooding	3.21	7.89	6.09	1.71
Salinity	1.65	2.78	1.99	1.78

Table 3—Feature importance for XG-Boost ML.

Feature	Weight-Based Feature Importance			
	OK	CY	AA	XG
S_{oi}	20.0310	6.3332	28.7082	1.5147
Polymer Viscosity	5.1845	12.0040	15.7706	45.7264
Polymer Viscosity After Adsorption	2.6679	0.7188	0.6595	259.8826
Polymer Concentration	1.2914	2.2856	3.7835	65.0359
Permeability	1.1502	4.7567	3.6012	1.6326
Residual S_w After Water Flooding	0.4575	2.2899	2.3460	2.0509
Salinity	0.3162	0.0600	0.0319	1.3048

Conclusion

Based on the machine learning research employing the random forest regression and extreme gradient boost algorithms, the following conclusions can be derived:

- (1) Extreme Gradient Boost demonstrated a superior alignment with the actual values, achieving a regression rate ranging from 97.8% to 98.9%, whereas Random Forest exhibited a regression rate spanning from 92.2% to 97.1%.
- (2) Based on the feature importance analysis conducted through random forest regression and extreme gradient boost algorithms, the initial oil saturation was found to be a significant determinant of the enhanced oil recovery (EOR) performance of *Abelmoschus esculentus* and *Afzelia africana*. Polymer viscosity emerged as a crucial factor for the EOR performance of *Colocasia esculenta*, whereas the polymer viscosity after adsorption played a vital role in the EOR performance of xanthan gum.
- (3) The viscosity of polymers exerts a substantial influence on the enhanced oil recovery (EOR) performance of *Abelmoschus esculentus* and *Afzelia africana*. The initial oil saturation and rock permeability have an impact on the EOR performance of *Colocasia esculenta*. Meanwhile, the viscosity and concentration of polymers affect the EOR performance of xanthan gum.

Conflict of Interest

The author(s) declare that they have no conflicting interests.

Reference

- Abidin, A.Z., Puspasari, T., and Nugroho, W.A. 2012. Polymers for Enhanced Oil Recovery Technology. *Procedia Chemistry* **1**(4): 11-26.
- Abdulraheem, M., Hamisu, T., Abdullahi, G., et al. 2018. Comparative Analysis on Rate Dependent Polymer Flooding Using Bio and Synthetic Polymers. Paper presented at the SPE Nigeria Annual International Conference and Exhibition, 6-8 August. SPE-193529-MS.
- Agi, A., Junin, R., Gbonhinbor, J., et al. 2018. Natural Polymer Flow Behaviour in Porous Media for Enhanced Oil Recovery Applications: A Review. *Journal of Petroleum Exploration and Production Technology* **8**(4): 1349-1362.
- Agi, A., Junin, R., Jaafar, M.Z., et al. 2022. Laboratory Evaluation to Field Application of Ultrasound: A State-of-the-Art Review on the Effect of Ultrasonication on Enhanced Oil Recovery Mechanisms. *Journal of Industrial and Engineering Chemistry* **110**(1): 100-19.
- Dike, C.F., Izuwa, N.C., Kerunwa, A., et al. 2024. Parametric Evaluation on the Interfacial Tension Response of Agro-Surfactant. *Improved Oil and Gas Recovery Journal* **8**(1): 1-12.
- Dike, C.F., Izuwa, N.C., Kerunwa, A., et al. 2024. An Investigation on the Enhanced Oil Recovery Performance of Local Biopolymers. *Journal of Chemical and Petroleum Engineering* **58**(2): 311-324.
- Ekene, D.A., Nkechi, N.A., Uchechukwu, O.N., et al. 2024. Experimental Investigation of the Suitability of Afzelia africana and Colocasian esculenta as Alternative to Hydroxyethyl Cellulose in Enhanced Oil Recovery. *Petroleum Science and Engineering* **8**(1): 16-26.
- Eiroboyi, I., Ikiensikimama, S.S., Oriji, B.A., et al. 2019. The Effect of Monovalent and Divalent Ions on Biodegradable Polymers in Enhanced Oil Recovery. Paper presented at SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 5-7 August. SPE-198788-MS.
- Fadairo, A., Adeyemi, G., Onyema, O., et al. 2019. Formulation of Bio-Waste Derived Polymer and Its Application in Enhanced Oil Recovery. Paper presented at SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 5-7 August. SPE-198750-MS.
- Gbadamosi, A.O., Kiwalabye, J., Junin, R., and Augustine, A. 2018. A Review of Gas Enhanced Oil Recovery Schemes Used in the North Sea. *Journal of Petroleum Exploration and Production Technology* **8**(1): 1373-87.
- Gbadamosi, A.O., Radzuan, J., Muhammad, A.M., et al. 2019. An Overview of Chemical Enhanced Oil Recovery: Recent Advances and Prospect. *International Nano Letters* **9**(1): 171-202.
- Obuebite, A.A., Onyekonwu, M.O., Akaranta, O., et al. 2018. Effect of Salinity and Divalent Ions on Local Bio Polymers. Paper presented at the SPE Nigeria Annual International Conference and Exhibition, 6-8 August. SPE-193450-MS.
- Ogolo, N.A., Ogiriki, S.O., Onyiri, V.I., et al. 2015. Performance of Foreign and Local Agents for Enhanced Oil Recovery of Nigerian Crude. Paper presented at the SPE Nigeria Annual International Conference and Exhibition, 4-6 August. SPE-178305.
- Sedaghat, M.H., Ahadi, A., Kordnejad, M., et al. 2013. Aspect of Alkaline Flooding: Oil Recovery Improvement and Displacement Mechanisms. *Middle East Journal of Scientific Research* **18**(1): 258-263.
- Sheng, J.J. 2013. A Comprehensive Review of Alkaline-Surfactant-Polymer (ASP) Flooding. Paper presented at SPE Western Regional Meeting, Bakersfield, USA, 19-25 April. SPE-165358-MS.
- Uzoho, C.U., Onyekonwu, M., and Akaranta, O. 2015. Formulation of Local Alkaline-Surfactant-Polymer (ASP) for Enhanced Oil Recovery in Niger Delta: A Review. Paper presented at SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 4-6 August. SPE-178300-MS.
- Uzoho, C.U., Onyekonwu, M., and Akaranta, O. 2019. Chemical Flooding Enhanced Oil Recovery Using Local Alkali-Surfactant-Polymer. *World Journal of Innovative Research* **7**(1): 16-24.
- Zerkalov, G. 2012. Polymer Flooding for Enhanced Oil Recovery. <http://large.stanford.edu/courses/2015/ph240/zerkalov1/> (accessed 7 August 2025).