Nigerian Journal of Physics (NJP)
ISSN online: 3027-0936
ISSN print: 1595-0611
DOI: https://doi.org/10.62292/njp.v35(s).2026.656
Volume 35(S), 2026

Comparative Evaluation of Machine Learning Algorithms for Permeability Prediction from Wireline Well Log Data

***[1]Babalola Oludare Olukayode, [2]Ogunleye Odunayo Christiana, [1]Adedokun Damilola Rukayat and [1]Oloruntola Racheal Foluke**

[1]Department of Physics, Lead City University, Ibadan, Nigeria.
[2]Department of Geology, Obafemi Awolowo University, Ile-Ife. Osun State, Nigeria.

*Corresponding Author's Email: darekay35@gmail.com

Keywords:
Permeability prediction,
Machine learning,
XGBoost,
Random Forest,
Leave-one-well-out validation.

Submitted: 20 Jul. 2026

Accepted: 3 Aug. 2026

Published: 10 Aug. 2026

ABSTRACT
Accurate permeability prediction from wireline well logs remain a persistent challenge in reservoir characterization, as conventional empirical correlations often fail to capture the heterogeneous nature of reservoirs, while core-derived permeability data are costly and rarely available across all wells. Machine learning offers a data-driven alternative capable of modelling these complex relationships, but most existing studies rely on random train-test splitting, which ignores the spatial autocorrelation inherent in well log data and can overstate model generalizability. In this research, five different models of machine learning: random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost), one-dimensional convolutional neural network (1D-CNN), and CNN–LSTM network were compared in terms of their ability to predict formation permeability based on well log data collected from three wells (Jay_01, Jay_02, Jay_03) in the Jay field, Niger Delta Basin, Nigeria. Four log curves, namely gamma ray (GR), neutron porosity (NPHI), deep resistivity (RT), and bulk density (RHOB), were used as inputs. The methodology used in this paper compares model performance through two validation procedures: random train-test splits (80% and 20%), and Leave-One-Well-Out (LOWO) cross-validation where every well in turn was left out as test data to take into account spatial autocorrelation in sequential in-depth logging data. R^2 values estimated through random splits (0.83-0.97) overestimate generalization capacity of the models in comparison with LOWO R^2 values (0.63-0.73). Among the five models, XGBoost performs the best under LOWO procedure ($R^2 = 0.734 \pm 0.163$, MAE = 0.654 log-mD), followed by Random Forest ($R^2 = 0.725 \pm 0.169$).

INTRODUCTION

Permeability is an important parameter governing fluid flow through porous media, one of the most critical petrophysical parameters for reservoir characterization, simulation model initialization, and well performance evaluation (Okoro et al., 2025). Direct measurement using core analysis procedures is costly, time-consuming, and not applicable in most wells drilled (Ottesen & Hjelmeland, 2008). Well logging data on the other hand can be collected continuously in almost all the drilled wells, and several empirical equations have utilized the data in predicting permeability (Abdulwahab et al., 2026; Timur, 1968; Coates & Denoo, 1981). However, these empirical equations assume mathematical forms which do not account for the heterogeneous nature of clastic reservoirs such as those of the Niger Delta Basin (Adamolekun, 2023). Machine Learning (ML) and Deep Learning (DL) models excel at modeling complex, multi-dimensional permeability behaviors from well logs relative to traditional correlation methods (Mohaghegh, 2000; Al-Anazi & Gates, 2010; Wood, 2020; Gu et al., 2021). However, the majority of the research suffer from a fundamental limitation in their methodology: the process of randomly splitting data into train/test sets ignores the fact that well log data is inherently spatially auto-correlated (Okunlola et al., 2021). Nearby points at depths of several feet away from each other are highly correlated to each other in a single well; hence, a random split allows the model to learn from the nearest neighbors of its own test points (Bichri et al.,2024). This paper bridges the gap by evaluating two validation techniques, random splitting and Leave-One-Well-Out (LOWO), through the application of five machine learning algorithms on three wells within the Jay field in Niger Delta. This paper makes the following contributions: a spatio-temporal LOWO validation methodology; the first comprehensive comparison of five algorithms under the two validation techniques on a Niger Delta data set; an explicit quantification of the random splitting inflation factor for all the five algorithms; and recommendations for algorithm selection in Niger Delta reservoirs.

MATERIAL AND METHODS

Niger Delta Basin

The Niger Delta Basin is situated along the passive continental margin of West Africa and is one of the world's most productive basins with proven reserves over 34 billion barrels (Tobiloba, 2024). The stratigraphic layers include the marine Akata which is predominantly a shaly formation, the paralic Agbada formation which is sandstones with the intercalation of shales (which is the reservoir), and the overlying Benin formation which is sandstones (which is the continental) (Short & Stäuble, 1967.). Sandstone of the Agbada formation forms distributary channels, barrier bars, and tidal flats with pore volumes between 15% and 30%, and permeabilities of 0.1 mD to 2,000 mD, with grain size of the formations being the most important factor influencing both the permeability and porosity (Egbokhare et al., 2026).

Well Log Dataset

The dataset comprises wireline logs for three wells, namely Jay_01, Jay_02, and Jay_03. After quality check and elimination of null values, 4,109 valid depth measurements at an interval of 0.5 feet remained. There were four input features for ML that included: gamma ray (GR, API); bulk density (RHOB, g/cc); neutron porosity (NPHI, v/v); and deep resistivity (RT, Ω.m). Dataset properties are shown in Table 1 while Figure 1 displays stratigraphic correlation of reservoir sands in the three wells using Gamma Ray log and Deep Resistivity log. Two reservoir Sand A and Sand B was picked using the GR and Resistivity logs signatures and correlated across the three wells in Jay field; regions with low GR values and high resistivity are picked as the reservoir [sand].

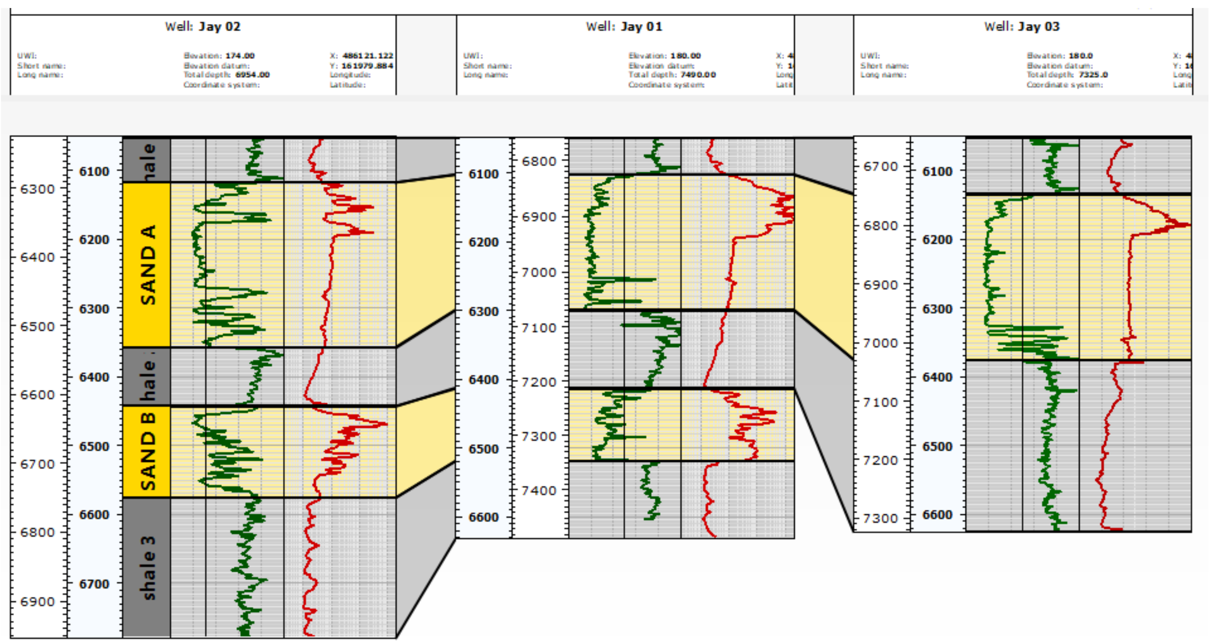

Figure 1: Stratigraphic Correlation of three wells (Techlog 2015)

Table 1: Well log dataset characteristics for the Jay field, Niger Delta

Well	Depth Range (ft)	N Samples	GR (API)	RHOB (g/cc)	RT (Ω·m)	NPHI (v/v)
Jay_01	6760–7450	1,380	20–145	2.10–2.65	0.8–120	0.08–0.38
Jay_02	6640–7370	1,461	18–138	2.08–2.63	1.0–110	0.07–0.40
Jay_03	6580–7210	1,268	22–152	2.12–2.67	0.9–105	0.09–0.36

Target Variable

Permeability was determined using the Timur equation (Timur, 1968):

$$k = \frac{8581\, x \emptyset_E^{\;4.4}}{S_w^{\;2}} \qquad (1)$$

where $\emptyset_E$ is effective porosity and S_w is water saturation; all predictions are in $\log_{10}(k)$. The use of a synthetic target variable instead of core data is a common technique in cases when core data are not available.

Data Preprocessing

The input curves were normalized with Z-scores that were estimated only on the training set to avoid data leakage (Schipper et al., 2025). All samples having even one missing value were removed from the dataset (Palanivinayagam & Damaševičius, 2023).

Algorithm Descriptions

Random Forest (RF)

Ensemble consisting of 300 uncorrelated decision trees based on bootstrap samples (Breiman,2001). Parameters: $max\_depth = 12$, $min\_samples\_split = 4$, $min\_samples\_leaf = 2$, $max\_features = 0.8$.

Support Vector Machine (SVM)

Support Vector Regression with RBF kernel (C=100, ε=0.1, γ=scale). Standardization of inputs is mandatory (Zhuo et al., 2026).

XGBoost

Gradient boosting with L1 (α=0.1) and L2 (λ=1.0) penalties, row/column subsampling (0.8/0.8), and learning rate η=0.05

1D-CNN

3-Layer 1D-CNN (filters 16/32/64; kernel=3; ReLU) on 11 sample sliding window of depth. Head: 128→64→1 with 30% dropout. Optimized with Adam (lr=5×10^{-4}), with gradient clipping.

Hybrid CNN+LSTM

2 CNN layers (32/64 filters) + two-layer LSTM (64 units) on a 15-sample window capturing deep range sequential dependence (Hochreiter and Schmidhuber, 1997). Optimized with Adam (lr=2×10^{-4}).

Validation Methodology

Random 80/20 Split

Standard random split (80% training / 20% testing; seed=42) used for all depth samples (Tripathi et al.,2025). Useful for comparing to existing work but known to exaggerate generalizability metrics (Huang et al., 2026).

Leave-One-Well-Out (LOWO) Cross-Validation

Leave one well out (LOWO): Each of the three wells completely left out as the test set, while model trained on the other two wells (Ahdritz et al., 2024). Feature scaler is refit per fold on training wells. LOWO mimics deploying on a completely novel well and is the primary reported metric (Soccodato et al., 2024).

Performance Metrics

Model performance metrics are R^2, MAE (log-mD), and RMSE (log-mD) in $\log_{10}(k)$ (Hlaing et al.,2024)

RESULTS AND DISCUSSION

Random-Split Performance

All five models achieved high R^2 values under random split, with RF (0.9652) and XGBoost (0.9639) leading the field. Figure 2 to figure 4 shows actual-vs-predicted scatter plots and residual distributions for the three scikit-learn models; RF produced the tightest residual spread (MAE = 0.219 log-mD).

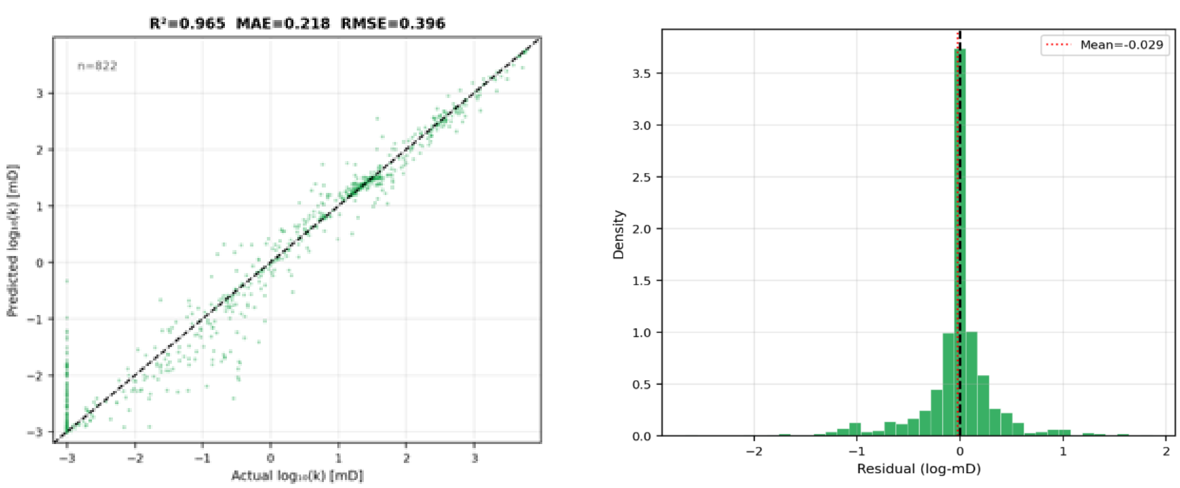

Figure 2: Actual vs predicted $\log_{10}(k)$ scatter plots and residual distributions under random 80/20 split for Random Forest

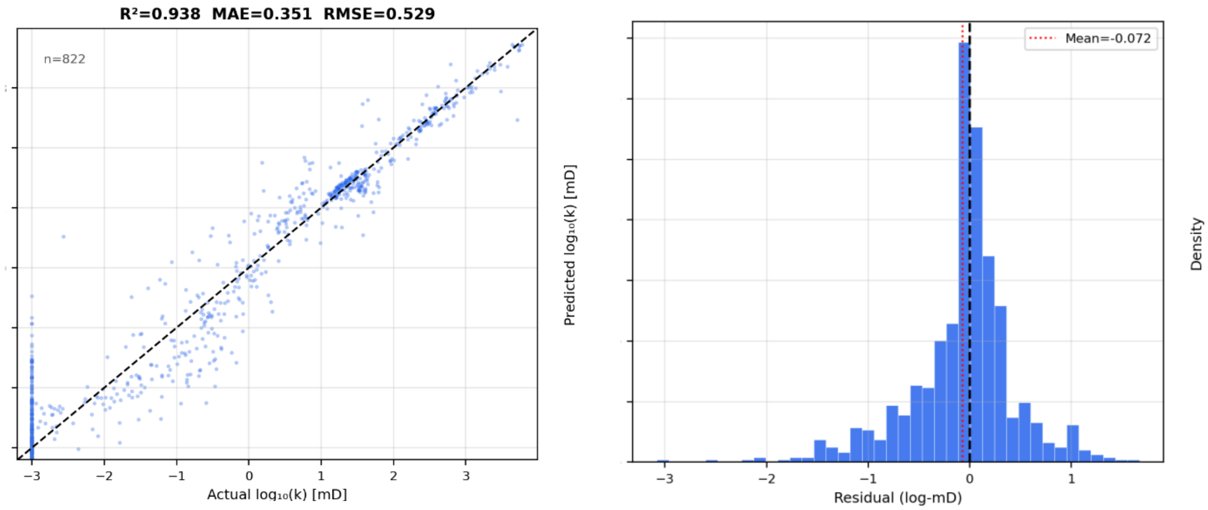

Figure 3: Actual vs predicted $\log_{10}(k)$ scatter plots and residual distributions under random 80/20 split for Support Vector Machine

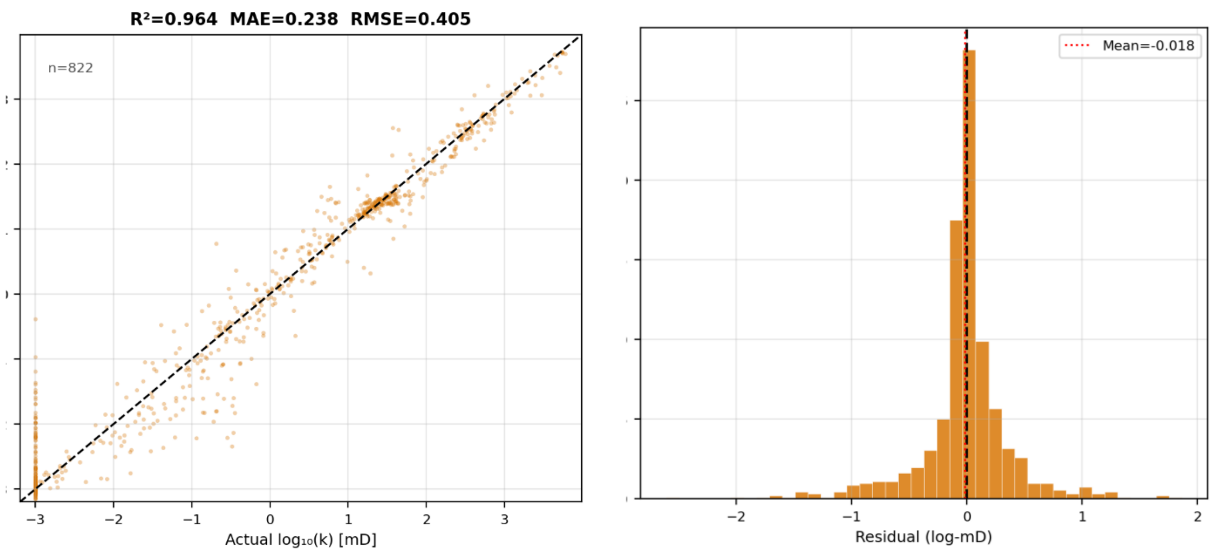

Figure 4: Actual vs predicted $\log_{10}(k)$ scatter plots and residual distributions under random 80/20 split for XGBoost

Feature Importance

Figure 5 sets out feature importance for RF and XGBoost alongside the Pearson correlation matrix. Both models rank RT and RHOB as the most influential predictors — a result consistent with their petrophysical roles as proxies for fluid saturation/connectivity and porosity, the two primary controls on permeability in clastic reservoirs.

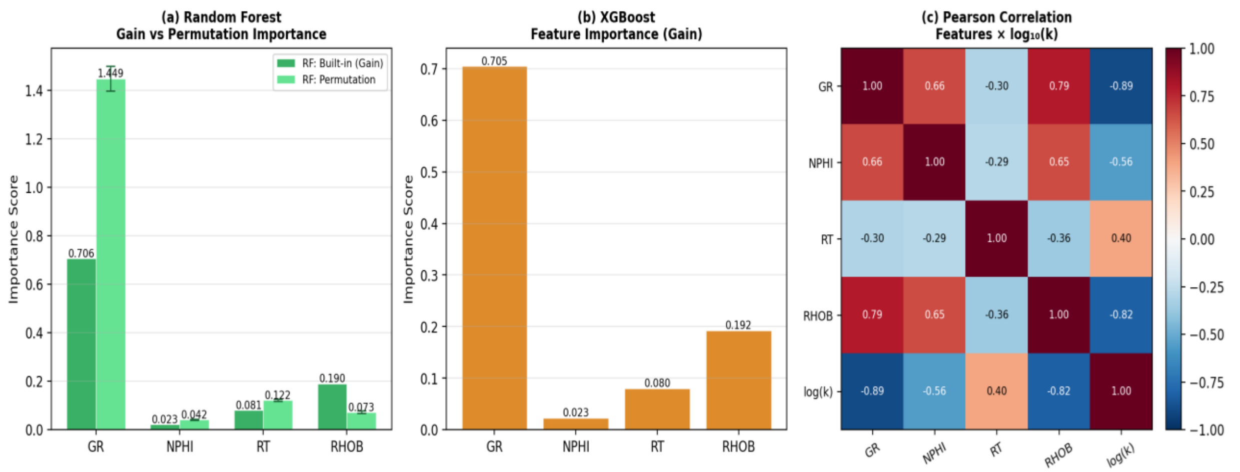

Figure 5: Feature importance. (a) RF built-in gain vs permutation importance (error bars = ±1 std, 30 permutations). (b) XGBoost gain importance. (c) Pearson correlation matrix

Leave-One-Well-Out Performance

All models show substantial R^2 reductions under LOWO validation relative to the random-split results (Table 2; Figure 4), confirming spatial autocorrelation leakage. XGBoost leads with mean LOWO R^2 = 0.734 ±0.163, followed by RF (0.725 ±0.169), CNN+LSTM (0.695 ±0.175), 1D-CNN (0.673 ±0.171), and SVM (0.633 ±0.133). Jay_01 proved most predictable when held out (R^2: 0.777–0.964), while Jay_02 posed the greatest generalisation challenge (R^2: 0.456–0.607) — a contrast that points to lithofacies differences between wells.

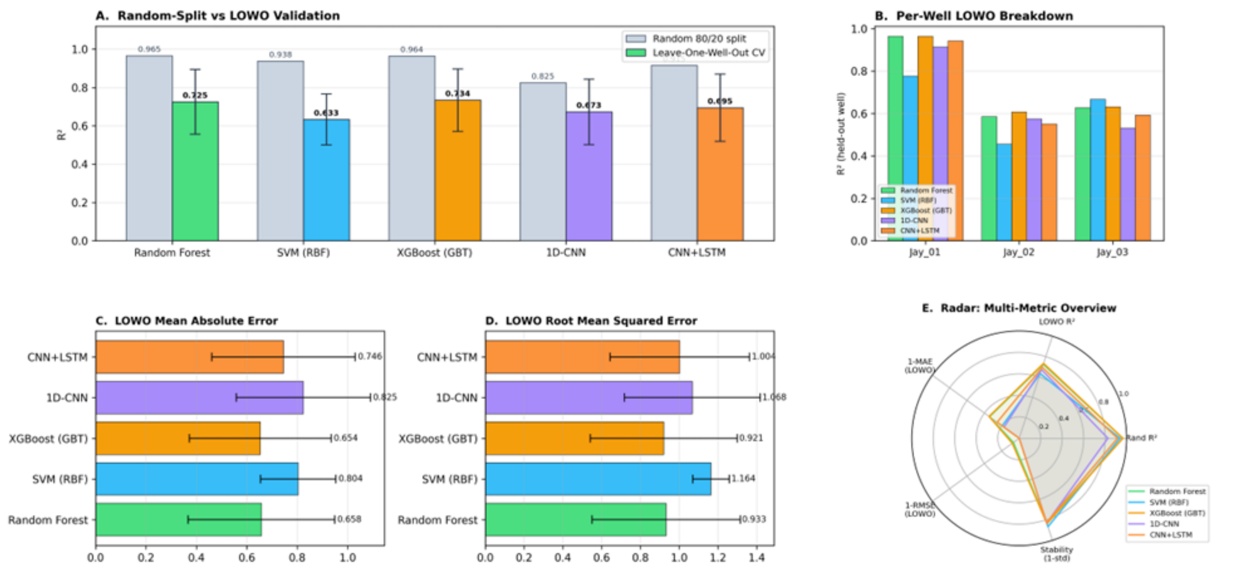

Figure 6: Five-model comparison. (A) Random-split vs LOWO R^2. (B) Per-well LOWO R^2 breakdown. (C–E) Scatter plots for RF, SVM, XGBoost

Table 2: Performance metrics — all five models, both validation strategies

Model	Rand R^2	Rand MAE	Rand RMSE	LOWO R^2 (mean±std)	LOWO MAE	LOWO RMSE	Jay_01	Jay_02	Jay_03
Random Forest	0.9652	0.219	0.397	0.725±0.169	0.658	0.933	0.963	0.585	0.627
SVM (RBF)	0.9382	0.351	0.529	0.633±0.133	0.804	1.164	0.777	0.456	0.667
XGBoost	0.9639	0.237	0.404	0.734±0.163	0.654	0.921	0.964	0.607	0.631
1D-CNN	0.8254	0.706	0.889	0.673±0.171	0.825	1.068	0.913	0.575	0.531
CNN+LSTM	0.9154	0.519	0.619	0.695±0.175	0.746	1.004	0.942	0.550	0.593

Validation Strategy Gap Analysis

Table 3 quantifies the R^2 gap between random-split and LOWO results. Every model loses more than 0.15 R^2 units, and SVM suffers the largest relative drop at 32.5%. Even XGBoost declines by 23.9% this show that no algorithm can totally escape spatial leakage inflation.

Table 3: Validation strategy gap: ΔR^2 = Random-Split R^2 minus LOWO R^2, and relative performance drop

Model	Random-Split R^2	LOWO R^2 (mean)	ΔR^2 (gap)	Relative drop (%)
Random Forest	0.9652	0.7253	0.2399	24.8%
SVM (RBF)	0.9382	0.6333	0.3049	32.5%
XGBoost	0.9639	0.7339	0.2300	23.9%
1D-CNN	0.8254	0.6727	0.1527	18.5%
CNN+LSTM	0.9154	0.6949	0.2205	24.1%

Permeability Depth Tracks

Figure 7 plots XGBoost depth-track predictions against the Timur–Coates reference for all three wells. The predicted profiles closely track the reference through high-permeability intervals (k > 100 mD), diverging more noticeably in tight zones. Orange shading at intervals where reference permeability exceeds 100 mD.

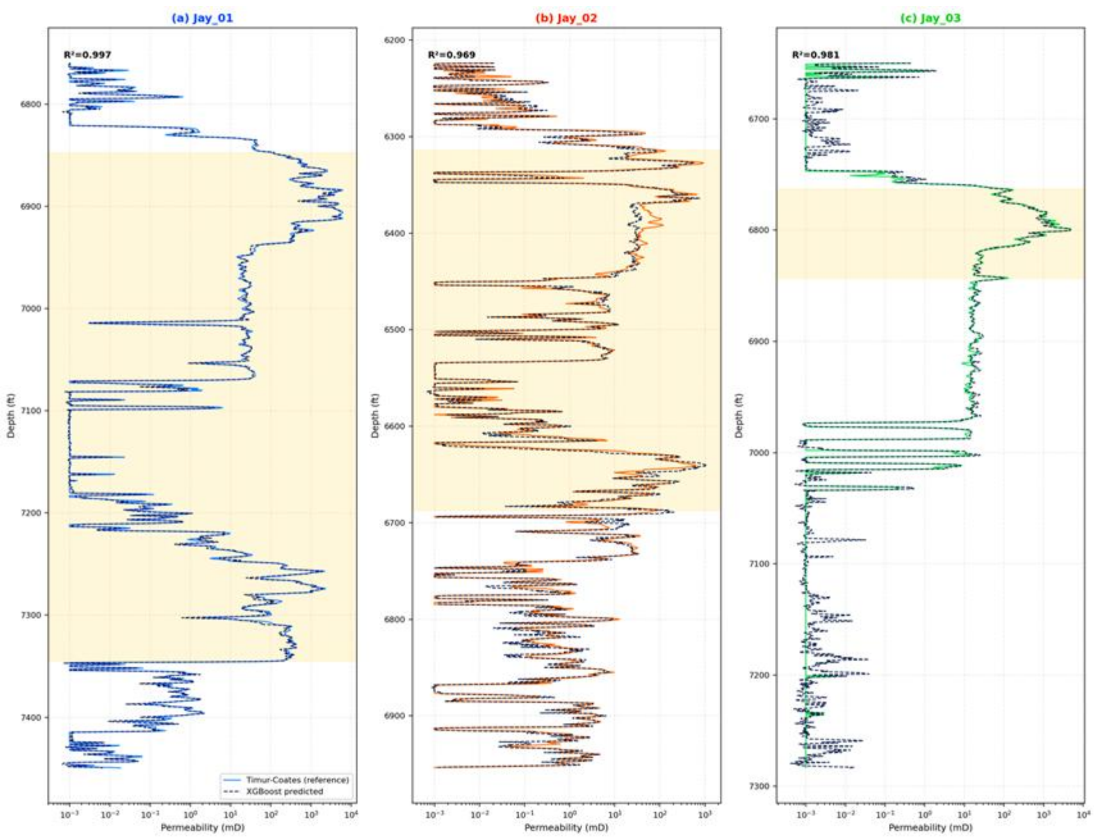

Figure 7: XGBoost permeability depth tracks vs Timur–Coates reference: (a) Jay_01, (b) Jay_02, (c) Jay_03

Discussion

Ensemble Methods vs Deep Learning in Low-Data Regimes

The XGBoost and RF models performing better than the CNN models under the LOWO validation is a trend consistent with the general pattern seen in the ML literature involving small and medium-sized tabular datasets (Grinsztajn et al., 2022). In this case, there are approximately 2,600-2,800 samples for each LOWO fold and four input features. Gradient Boosted Trees and Random Forests fit the dimensionality of the task better than deep learning models that require a lot of data to account for the complexity introduced by the extra parameters. The results also show that the CNN + LSTM model performed better than the 1D-CNN model (LOWO R^2 = 0.695 vs. 0.673). This shows that the sequence context information provided by the LSTM is helpful.

Spatial Autocorrelation and the LOWO Imperative

In this study, one of the main methodological findings is that there is always a big difference between the performance values obtained through random split and LOWO approaches for the five algorithms used. Spatial autocorrelation exists in the data taken from the wells with the interval of 0.5 feet, and in the random-split approach, spatially autocorrelated data points get distributed between the training and test samples, thus increasing the performance value. The same failure mode has been recognized by remote sensing and environmental machine learning (Roberts et al., 2017); however, in the aspect of petrophysical machine learning, this issue has not been properly addressed.

Per-Well Variability and Implications for Field Application

Variations of LOWO R^2 values are found between wells (Jay_01: 0.777–0.964; Jay_02: 0.456–0.607; Jay_03: 0.531–0.667), but these variations indicate that the prediction ability depends on the geological resemblance between training wells and tested wells. Indeed, Jay_02 proved to be difficult to predict, and this implies that there might be slight variation in the geological rock formation and structures of the Jay field.

Limitations and Future Directions

There are four main limitations to the results obtained. The first is that the artificial Timur-Coates target is an equivalent of the core-based data, so the validation using core-based data is very important prior to the use of the models in practice. Secondly, there are only three wells in the LOWO dataset that can be used in the analysis. Thirdly, the models were not conditioned by the facies types, and facies stratification of the models could show different permeability characteristics depending on the lithotype.

CONCLUSION

This study presents a comparative machine learning framework for well-log-based permeability prediction in the Jay Field, Niger Delta Basin, and shows that the choice of validation strategy strongly affects the reliability of reported model performance.

XGBoost and Random Forest achieved the best cross-well generalizability under LOWO validation (mean R^2 = 0.734 and 0.725, respectively), implying that tree-based ensemble methods are the more reliable choice for permeability prediction.

R^2 values fell by 18–33% when validation moved from random splitting to LOWO, showing that random-split metrics substantially overstate real-world predictive performance. LOWO (or an equivalent spatially aware scheme) should therefore be adopted as the minimum acceptable validation standard for petrophysical machine learning studies, to avoid overoptimistic reserve and productivity estimates.

CNN+LSTM outperformed the 1D-CNN under LOWO validation (R^2 = 0.695 vs. 0.673), indicating that where deep learning is used, architectures that explicitly capture depth-sequence context should be favoured over purely convolutional designs.

Per-well LOWO R^2 ranged from 0.456 to 0.964, showing that predictive accuracy depends on how geologically similar a new well is to the wells used for training. This implies that models should be recalibrated, or supplemented with local well data, before being applied to structurally or lithologically dissimilar parts of a field.

The target variable was derived from the Timur equation rather than measured core permeability, model outputs should be validated against core data before operational use; doing so is essential to avoid propagating systematic bias into well completion, stimulation, and reserve-estimation decisions.

Collectively, these findings establish LOWO validation as a stricter and more field-realistic benchmark for petrophysical ML, identify tree-based ensembles as the currently most robust and cost-effective tool for permeability prediction in Niger Delta reservoirs.

ACKNOWLEDGEMENTS

The authors acknowledge Dharmattan Nigeria Limited for well log data. All ML modelling was performed using open-source Python libraries. The authors declare no conflicts of interest.

REFERENCES

Abdulwahab, A. A., Al-Mudhafar, W. J., & Abbas, M. A. (2026). Hybrid physics-guided discrepancy modeling and explainable machine learning for permeability prediction in heterogeneous carbonate reservoirs using noisy legacy datasets. Computational Geosciences, 30(3), 49. https://doi.org/10.1007/s10596-026-10445-2

Adamolekun, O. J. (2023). Reservoir quality controls and reservoir modelling of continental to transitional shallow marine sedimentary basins of Southern Nigeria. KIT Scientific Publishing. https://doi.org/10.5445/KSP/1000156506

Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., & AlQuraishi, M. (2024). OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nature Methods, 21(8), 1514-1524. https://doi.org/10.1038/s41592-024-02272-z

Al-Anazi, A.F., Gates, I.D., 2010. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. Engineering Geology 114, 267–277. https://doi.org/10.1016/j.enggeo.2010.05.005

Breiman, L.,(2001) Random forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD, 785–794. https://doi.org/10.1145/2939672.2939785

Coates, G., Denoo, S.,(1981)The producibility answer product. The Technical Review 29, 54–63.

Egbokhare, O. J., Amigun, J. O., Gebretsadik, H. T., Alao, O. A., & Olutoki, J. O. (2026). Seismic sequence stratigraphy and spectral decomposition analysis for subtle reservoir delineation of "Turu" field, offshore Niger Delta. Arabian Journal for Science and Engineering, 1-21. https://doi.org/10.1007/s13369-026-11221-1

Grinsztajn, L., Oyallon, E., Varoquaux, G. (2022) Why tree-based models still outperform deep learning on tabular data. NeurIPS 35, 507–520. https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf

Gu, Y., Zhang, D., Bao, Z., (2021) A new data-space inversion for efficient uncertainty quantification in subsurface flow problems. Mathematical Geosciences 53, 679–707.. https://doi.org/10.1007/s11004-016-9672-8

Hlaing, P. T., Humphries, U. W., & Waqas, M. (2024). Hydrological model parameter regionalization: Runoff estimation using machine learning techniques in the Tha Chin River Basin, Thailand. MethodsX, 13, 102792. https://doi.org/10.1016/j.mex.2024.102792

Hadikurniawati, W., Hartomo, K. D., & Sembiring, I. (2026). Lightweight hybrid feature engineering and fusion with variational autoencoder and XGBoost for nutritional data classification. Engineering, Technology & Applied Science Research, 16(1), 30859-30868. https://doi.org/10.48084/etasr.15310

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, S., Zhang, Y., Rezaei, M., & Zhou, J. (2026). Hybrid RandWPSO–XGBoost with SHAP-based interpretability for prediction of caving-fracturing zone height in longwall mining. Rock Mechanics and Rock Engineering, 1-32. https://doi.org/10.1007/s00603-026-05481-4

Mohaghegh, S., (2000). Virtual-intelligence applications in petroleum engineering. Journal of Petroleum Technology 52, 64–73. https://doi.org/10.2118/58046-JPT

Okoro, E. E., Ogali, O. I., Sanni, S. E., Bariza, N., Mormah, P. C., John, I. T., & John, C. B. (2025). Optimizing hydrocarbon recovery in mature fields: A comparative simulation study of infill well placement and performance using integrated reservoir modeling techniques. Engineered Science, 34, 1399. https://doi.org/10.30919/es1399

Okunlola, O.A., Alobid, M., Olubusoye, O.E., Ayinde, K., Lukman, A.F., Szűcs, I. (2021). Spatial regression and geostatistics discourse with empirical application to precipitation data in Nigeria. Scientific Reports, 11(1), 16848. https://doi.org/10.1038/s41598-021-96124

Ottesen, B., & Hjelmeland, O. (2008). The value added from proper core analysis. In International Symposium of the Society of Core Analysts, Abu Dhabi, UAE, Weatherfordlabs (Vol. 29). https://jgmaas.com/SCA/2008/SCA2008-04.pdf

Palanivinayagam, A., & Damaševičius, R. (2023). Effective handling of missing values in datasets for classification using machine learning methods. Information, 14(2), 92. https://doi.org/10.3390/info14020092

Roberts, D.R., Bahn, V., Ciuti, S., et al., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881

Schipper, M., de Leeuw, C. A., Maciel, B. A., Wightman, D. P., Hubers, N., Boomsma, D. I., & Posthuma, D. (2025). Prioritizing effector genes at trait-associated loci using multimodal evidence. Nature Genetics, 57(2), 323-333. https://doi.org/10.1038/s41588-025-02084-7

Short, K. C., & Stäuble, A. J. (1967). Outline of geology of Niger Delta. AAPG bulletin, 51(5), 761-779. https://doi.org/10.1306/5D25C0CF-16C1-11D7-8645000102C1865D

Soccodato, D., Penazzi, G., Pecchia, A., Phan, A. L., & Auf der Maur, M. (2024). Machine learned environment-dependent corrections for aspds empirical tight-binding basis. Machine Learning: Science and Technology, 5(2), 025034. https://doi.org/10.1088/2632-2153/ad4510

Timur, A.(1968). An investigation of permeability, porosity and residual water saturation relationships for sandstone reservoirs. The Log Analyst 9, 3–17. www.onepetro.org/journal-paper/SPWLA-1968-vIXn4a2

Tobiloba, A. I. (2024). Integrated well log analysis for reservoir characterization of the Tomboy Field, Niger Delta using two wells (Doctoral dissertation, Department of Geology, Faculty of Science, Obafemi Awolowo University, Ile-Ife).

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. Springer, New York. https://doi.org/10.1007/978-1-4757-3264-1

Wood, D. A. (2020). Predicting porosity, permeability and water saturation applying an optimised nearest-neighbour ML workflow to well-log data. Journal of Petroleum Science and Engineering 184, 106587. https://doi.org/10.1016/j.petrol.2019.106587

Zhuo, H., Yang, M., Wu, B., Xiao, Y., Ma, J., Chen, Y., ... & Shi, P. (2026). Detection and precision application path planning for cotton spider mite based on UAV multispectral remote sensing. Agriculture, 16(4), 424. https://doi.org/10.3390/agriculture16040424