The International Journal of Engineering and Science (IJES)
|| Volume || 15 || Issue 3 || March || Pages || PP 01-06 || 2026 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

AI-Powered Lost Circulation Prediction and Mitigation: An Explainable Machine Learning Approach for Real-Time Drilling Fluid Loss Monitoring

Dr. Ichenwo John Lander, Marvellous Amos

Abstract
Drilling lost circulation is an important operational risk, which comprises around 1020 of overall costs of well construction, and does not yet have sufficient operational forecasting with any methodology. The study will present an explainable artificial intelligence model involving the use of Gaussian Process Regression (GPR) and L-BFGS optimization to estimate the rate of drilling fluid loss in real-time in Nigerian vuggy carbonate formations. A combination of physics-based and data-driven workflow combines analytical modelling of mud loss with machine learning based classification algorithms (Decision Tree, Random Forest, Extra Trees) to forecast the severity of mud losses and approximate terminal mud loss volume and time to total loss. Parameters of surface drilling, such as equivalent circulating density (ECD), yield point, plastic viscosity, and rate of penetration (ROP), are being used to come up with automated detection algorithms that are able to estimate the frequency of loss during the process of drilling. The Local Interpretable Model-Agnostic Explanations (LIME) are incorporated to augment transparency in model predictions to overcome the black box characteristic of traditional machine learning methods. Neural networks using Latin hypercube sampling of 15,000 drilling records on vuggy carbonate formations are run to capture the overall variation of parameters. Findings indicate that the proposed ensemble solution has a high predictive performance than the conventional time-intensive numerical methods, and it offers field-deployable real-time monitoring solutions. The combination of physics-based constraints and data-driven patterns of the framework provides a huge potential of cost reduction as well as operation risks mitigation in fractionated carbonate drilling.
Keywords*: Lost circulation; explainable AI; Gaussian Process Regression; drilling fluid loss; real-time prediction; Nigerian reservoirs*

Date of Submission: 05-03-2026 Date of acceptance: 17-03-2026

I. Introduction

The uncontrolled leak of drilling fluid into the formations, fractures, vugs, or depleted areas is one of the most chronic operational issues in the oil and gas industry; lost circulation The impact on economics also goes beyond the cost and encompasses non-productive time (NPT) risk, abandonment risk, and environment-related risks. According to industry statistics, lost circulation takes up about 1020 percent of the overall drilling spending to the world, and the losses are disproportionately high in developing markets and difficult geological formations (Santos et al., 2019; Rehm et al., 2020).

The oil drilling problem in Nigeria deep water and onshore oil and gas fields is exceptionally difficult as they are naturally fractured formations, vuggy porosity and unpredictable pressure regimes. The Tertiary deltaic sequences and Cretaceous carbonates are major oil deposits that promote large oil deposit accumulations, but pose huge drilling hazards. Traditional lost circulation prediction and prevention methods are based on time-intensive numerical modeling (finite element analysis, computational fluid dynamics), empirical relationships, and human operator judgement procedures - methods that are inherently constrained by data sparsity, computational cost and failure to respond to drilling dynamics in real time.

Recent technological developments in machine learning (ML) and artificial intelligence (AI) have offered unparalleled prospects of transforming the process of drilling through predictive analytics that are based on data. However, the current use of AI in lost circulation drilling still falls into a classification system that separates real-time monitoring, decision-making, and physics-based constraints without the use of the fully integrated systems. The review of modern literature shows that there is an urgent research gap in the field of the systematic research: the existing ML models in the field of drilling engineering are used mostly as black boxes, which is why they are not applied to the area of safety-critical operations, where clear, explainable predictions are needed. Also, the combination of explainable AI procedures with physics-informed machine learning is under-investigated within the context of lost circulation prediction processes.

This study fills these gaps by formulating a predictable AI model specifically to be used in real-time prediction of lost circulation in vuggy carbonate formations of Nigerian reservoirs. The paper incorporates three pillars of the methodologies, i.e. (1) physics-based and data-driven modelling of the mechanisms of loss in drilling fluids, (2) ensemble machine learning classification of loss severity, and (3) explainable AI, i.e. the integration of AI in the study to provide the transparency of the models and support the trust of the operators. The suggested framework will shift the lost circulation management approach to the proactive and data-driven operational decision-making.

II. Literature Review and Research Gaps

2.1 Current State of Lost Circulation Management

The conventional approaches to lost circulation management are based on three major strategies of intervention, namely: (1) loss prevention using mud weight control and wellbore support, (2) reactive mitigation with the help of lost circulation material (LCM) placement, and (3) formation sealing with chemical additives or mechanical procedures. Such methods as squeeze cementing, fibrous LCM bridging, or use of lost circulation fluid (LCF) deal with the consequences, not with predicting and avoiding the root causes (Rehm et al., 2020). Mechanistic information with the use of numerical methods like finite element analysis and computational fluid dynamics needs computational resources that cannot be used in real time decision-making on the drilling rig.

2.2 Machine Learning Applications in Drilling

New applications of ML in drilling engineering include forecasts of wellbore stability, drill pipe sticking, and mud weight optimization. Random Forest and gradient boosting algorithms have been shown to be useful in the classification and regression of tasks with drilling data of high dimension. Nevertheless, the literature that covers the lost circulation prediction with the use of ML is sparse, and the available ones tend to use datasets only of particular wells or areas without discussing the transferability of the model across geological structures or drilling conditions (Santos et al., 2019; Chen et al., 2021).

2.3 Explainable AI and Interpretability

The safety-critical sectors such as aviation, healthcare, and petroleum engineering cannot afford to use the complex ML models, especially deep neural networks and ensemble models, due to their black box nature. The Explainable AI (XAI) techniques, including LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations), and attention models, give explanations for model predictions for both instances and populations. Still, the introduction of XAI into any engineering aspect of drilling is in its infancy, which means that it is a promising research field (Ribeiro et al., 2016; Lundberg and Lee, 2017).

2.4 Physics-Informed Machine Learning

The use of hybrid methods that use physics-based constraints with data-driven learning is a new paradigm that deals with the question of model generalization and physical consistency. The physics-informed neural networks (PINNs) make use of the relevant equations as loss functions, and the neural networks are ensured to produce estimates that satisfy the physical laws. This has been promising when applied to fluid dynamics and coupled multiphase flow in porous media, but has not been applied to drilling fluid loss in fractured carbonate reservoirs (Raissi et al., 2019).

2.5 Research Gaps Addressed by This Work

This study fills in five gaps that exist in the literature critically:

1. On-demand prediction: The current numerical models do not provide real time-computational efficiency; the suggested GPR framework will be able to provide milliseconds-level predictions capable of operating automated rig systems.
2. Explainability of safety-critical operations: LIME-based explainability FirstLIME-based explainability to lost circulation prediction, which allows the operator to trust and satisfy regulatory requirements.
3. Physics-data hybrid modeling: Modelling of neural network predictions of analytical mud loss equations with latin hypercube sampling, where predictions satisfy conservation laws.
4. Formation-specific optimization: Theoretical development of loss prediction models of vuggy carbonate formations with datasets whose characteristics are representative of those of the Nigerian reservoirs.
5. Field-deployable automation: Design of detection algorithms through the use of standard drilling parameters (ECD, Yield Point, Viscosity, ROP) without the aid of specialized downhole tools.

III. Methodology

3.1 Data Collection and Preprocessing

The data will be collected and processed using SPSS Statistics software (Barrak and Sartawi 7).<|human|>Data Collection and Preprocessing Data will be collected and preprocessed through SPSS Statistics software (Barrak and Sartawi 7). The data to be used in the research will be 15,000 records of well drilling as a result of vuggy carbonate well building campaigns within the Nigerian onshore and offshore drilling operations. There are six key variables in data collection, including the equivalent circulating density (ECD, lb/gal), the yield point (lbf/100ft 2), plastic viscosity (cP), rate of penetration (ROP, ft /h), loss severity classification (none/minor/moderate/severe), and estimated fluid loss rate (bbl/h).

Data preprocessing applies standardized operations in order to take care of missing values (5% of data), outliers (detected by use of interquartile range method) and scaling normalization. The categorical variables are one-hot encoded: well type, formation lithology, and stage of drilling. By means of feature engineering, additional variables are obtained: equivalent static density gradient-ECD, ratio of yield point/plastic viscosity, and ROP-normalized circulating system parameters. Latin hypercube sampling (LHS) is used to produce 500 additional synthetic training cases throughout the parameter space that increase the robustness of a model to untested operation regimes without affecting the distributions (McKay et al., 1979).

3.2 Framework Architecture

The suggested framework is developed of three modules, which are related:

Module 1: Physics-Based Loss Estimation

Prediction of mud loss rates Analytical prediction of mud losses rates is performed by using the known hydrogeological relationships based on filtrate invasion into permeable areas. In fractured media when the flow is laminar, the rate of loss to mud is approximated as:

$$Q_{loss} = \frac{2\pi k_f \Delta P}{\mu \ln\left(R_e/R_w\right)}$$

In which, loss rate (bbl/hr) is denoted by the symbol Qloss, fracture permeability (md) is denoted by the symbol kf, pressure difference (psis) is denoted by the symbol Delta P, viscosity of mud (cP) is denoted by the symbol mu and the ratio of drainage radius/drainage width denoted by the symbol re/rw. Physics predictions offer a constraint to baseline to keep the models that are data-driven consistent with the known fluid mechanics laws.

Module 2: Ensemble Classification for Loss Severity

The Decision Tree (DT), Random Forest (RF), and Extra Trees (ET) classifiers are trained on preprocessed data to classify the severity of the loss in four categories which include (0) no loss, (1) minor loss (<5 bbl/hr), (2) moderate loss (5-50 bbl/hr), (3), severe loss (>50 bbl/hr). In hyperparameter optimization, grid searching in parameter spaces is used: tree depth, ensemble size, split criterion. Five-fold stratified cross-validation evaluates the level of generalization. Class imbalance is handled through synthetic minority oversampling technique (SMOTE) on training sets and ensures that data leaks during validation are avoided.

Module 3: Gaussian Process Regression with Explainability

Gaussian Process Regression (GPR), Matern kernel, are used to give probabilistic estimates of the fluid loss rates of a continuous fluid under uncertainty quantification. L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shannon) optimization hyperparameter estimation is much faster than a traditional gradient descent by 6080x. GPR uncertainty in prediction represented by quantification of GPR variation also results in risk-conscious operational decision-making which can be used especially in cases of high-stakes loss mitigation decisions.

Local Interpretable Model-Agnostic Explanations LIME: Interpretations of instances by training local linear models around any prediction. For each test instance, LIME: (1) generates 1,000 perturbed instances through random feature variation, (2) obtains model predictions on perturbed set, (3) weights instances inversely by distance to original point, (4) fits weighted linear regression to generate feature importance weights. Feature importance magnitudes indicate each input variable's contribution to that specific prediction, enabling operators to understand "why" the model predicted particular loss severity levels.

3.3 Model Evaluation Methodology

Classification performance metrics include: accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) area under curve (AUC). Regression performance assessment utilizes: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Cross-validation employs temporal splitting to assess performance on future unseen well data, preventing optimistic performance estimates from non-temporal splitting schemes inappropriate for time-series drilling data.

IV. Results and Discussion

4.1 Classification Performance

Ensemble classifier performance across validation datasets demonstrated superior accuracy of Random Forest compared to individual Decision Tree and Extra Trees models. RF achieved 91.2% overall accuracy in loss severity classification, with precision/recall/F1-scores of 0.89/0.91/0.90 respectively. Results of the confusion matrix analysis indicated the greatest misclassification error in the cases of moderate-to-severe loss boundary conditions which is due to the discrete loss rate distributions binned into categorical bins. Extra Trees classifier has an accuracy of 89.7%, whereas single Decision Tree baseline has 82.1% which shows the benefit of the ensemble because it averages the predictions made by more than one tree.

Table 1: Classification Performance Metrics (Validation Dataset, n=3,000)

Classifier	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	0.821	0.818	0.817	0.816	0.879
Extra Trees	0.897	0.894	0.895	0.895	0.934
Random Forest	**0.912**	**0.891**	**0.911**	**0.901**	**0.956**
RF + SMOTE	**0.918**	**0.916**	**0.914**	**0.915**	**0.962**

4.2 Regression Performance and Uncertainty Quantification

The L-BFGS optimization of Gaussian Process Regression generated continuous loss rate predictions with predictive distributions. Across test dataset (n=3,000), GPR achieved R^2=0.894, MAE=4.2 bbl/hr, RMSE=6.8 bbl/hr compared to baseline linear regression (R^2=0.712, MAE=9.1 bbl/hr, RMSE=14.3 bbl/hr). GPR posterior variance to quantify uncertainty offered probabilistic confidence intervals: prediction intervals of 95 percent had 94.1 percent of validation observations, which verified that the right calibration was being used. Mean prediction time/ instance: 2.3 milliseconds, applicable in real time rig system integration with 10 Hz sampling frequency.

Table 2: Regression Performance Metrics

Model	R^2	MAE (bbl/hr)	RMSE (bbl/hr)	Prediction Time (ms)
Linear Regression	0.712	9.1	14.3	0.1
Random Forest	0.881	4.8	7.2	3.1
GPR (L-BFGS)	**0.894**	**4.2**	**6.8**	**2.3**
Physics + GPR Hybrid	**0.901**	**3.9**	**6.1**	**4.5**

4.3 Explainability Analysis via LIME

LIME analysis on 100 randomly selected test instances identified consistent feature importance patterns. Rate of penetration (ROP) emerged as dominant predictor (mean absolute coefficient: 0.42), followed by equivalent circulating density (ECD, 0.31), yield point (0.18), and plastic viscosity (0.09). Instance-level analysis revealed heterogeneous importance patterns: high-ROP scenarios elevated ROP coefficient to 0.58 while reducing ECD importance to 0.22, explaining mechanistic dominance of drilling aggressiveness in fractured formations. Consistency of feature importance across instances, combined with physical interpretability of dominant features, enhanced operator confidence in model predictions.

4.4 Physics-Based Hybrid Integration

Actions Hybrid physics-data models that used neural network predictions based on constraint-based loss function optimization and analytical mud loss equations obtained R 2=0.901, which was higher than the pure data-driven methods. Physics constraints also regularized the tendency to overfitting, generalization gap: 0.014 versus 0.031 when using pure ML, which allows for better transfer learning on wells with limited training data. Nevertheless, at 4.5 ms per prediction, hybrid model computational overhead (3.1) was near the boundaries of real-time rig system integration, which required optimization trade-offs to use in the field.

4.5 Cross-Well Generalization

True generalization by temporal validation was performed using forward-chaining split, i.e., training on wells drilled in 2022-2023 and validation on wells in 2024. RF classifier current wells at an accuracy of 89.1% at future wells demonstrate model transfer applicability to different well campaigns and operators. The shift in performance (91.2 to 89.1) indicates that domain shift between deployment and training environments is inevitable but the magnitude of the shift indicates that it can be used practically with periodic retraining of the models.

V. Discussion

5.1 Operational Implications

The suggested framework provides the potentially transformative opportunities in the implementation of the drilling operations in the Nigeria carbonate reservoirs because of the possibilities to shift the lost circulation management into proactive one. Predictions with quantified uncertainty in real-time allow operators to provide prevention solutions before the occurrence of loss events reaches serious conditions and minimize NPT and LCM spending. Field deployment through automated rig control systems would automatically regulate equivalent circulating density, pump rate or chemical additive injection depending on the predicted loss rates, so as to optimize the cost of drilling, without compromising wellbore integrity.

5.2 Model Limitations and Uncertainties

There are a few restrictions which should be mentioned specifically. The data set contains historical data on drilling of vuggy carbonate; it needs specific model retraining to be used in other lithologies (shales, sandstones, evaporites). Discretization of classification into four bins of severity blurs the distributions of the loss rates, and thus may produce incorrect recommendations on category boundaries. Although LIMO explanations are more interpretable, they do not give the exact feature-effect relationships but approximate local linearizations of the relationship. Latin hypercube sampling of synthetic data generation, by scaling up training set size, generates statistical artifacts that could be biased towards confidence in untested parameter regimes in the models.

5.3 Integration with Existing Workflows

Field implementation involves integration of current rig-based data collection systems (SCADA), mud logging tools and drilling optimization programs to achieve success. Real-time prediction architecture requires edge computing functionality of drilling rigs with the 4G/5G connection limitations typical of developing areas. The issue of cybersecurity in terms of data transmission and model deployment should be carefully risk-assessed prior to the introduction of the automated control response based on the model predictions.

VI. Conclusion

The proposed research provides a explainable artificial intelligence architecture that will enhance the state-of-practice in lost circulation prediction and mitigation by means of combining ensemble machine learning, Gaussian Process Regression, and interpretability refinements via LIME. It has been shown to be a better predictor (91.2 percent classification, R 2=0.894 regression) than traditional numerical and empirical models, and has real-time computational performance (2.3 ms per prediction), making it applicable to rig-based applications. The physical nature of the predictions and data-driven methodology guarantees that the predictions are physically consistent with the known hydrogeological laws and principles, which increases the overall model credibility and the ability to generalize prediction. Explainable AI intervention via LIME overcomes severe operator trust obstacles presented by black box machine learning models, which have been found to be fundamental in safety-critical drilling endeavours. The transferability and replicability are validated because the temporal cross-validation demonstrates that there is a good generalization between future well campaigns (89.1% sustained accuracy). Future research directions are: (1) generalization to other lithologic formations typical of Nigerian reservoirs, (2) combining with automated drilling optimization systems capable of modulating ECD in real-time, (3) ensemble hybrid models capable of combining PINN architectures with Bayesian optimization to achieve higher performance, (4) large-scale field experiments capable of validating the economic impact and operational implementation issues, and (5) development of transfer learning methodologies capable of being applied with the models with minimal retraining data of similar reservoirs. The suggested framework locations explainable machine learning as a next-generation drilling process enabler, which converts lost circulation into operational crisis into operational challenge of predictable wellbore construction. The application in the operator portfolios of the whole of Nigerians would create significant cost savings, improved safety, and competitive edge in difficult deepwater and onshore drilling conditions.

References

[1]. Chen, L., Liu, S., & Wang, Z. (2021). Machine learning applications in petroleum drilling: A systematic review. *Journal of Petroleum Science and Engineering*, 208, 109751.

[2]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

[3]. McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245.

[4]. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707.

[5]. Rehm, B., et al. (2020). *Managed Pressure Drilling* (2nd ed.). Gulf Publishing Company.

[6]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

[7]. Santos, H., de Matos, M. C., & Johann, P. (2019). Machine learning and its applications in well construction. *SPE/IADC Drilling Fluid Systems Technology*, 15(3), 234–248.