Jeremiah Ifeanyi Okoroma, 2026, 14:1
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Machine Learning for Predicting Cement Zonal Isolation Quality Using Mixed Telemetry and Limited Cement Evaluation Data

Jeremiah Ifeanyi Okoroma, Ikechi Igwe

Abstract - Sustained casing pressure (SCP), prevention of cross-flow between intervals of formation, and reliable cement zonal isolation depend on attaining integrity of the well over time and eliminating cross-flow between intervals. The standard cement evaluation devices (CET) like ultrasonic, acoustic logs have high fidelity but tend to be few, costly to purchase or become inaccessible in high angle, deepwater or slimhole completions. The paper builds a machine-learning (ML) model to forecast the quality of cement zonal isolation with mixed data; rig-site telemetry, mud logging parameters, pumping schedules, and a few cement evaluation logs. The gradient boosting and sequence learning networks were combined in a hybrid and trained on multi-well data sets with full and partial cement logs. The findings indicate that the ML workflow predicts the quality of isolation with an accuracy of 92% and a mean absolute error of 0.07 and an AUC of 0.89, and on the other hand, the isolated empirical correlations and physics-based rules perform worse. The model has been able to generalize on wells having no cement logs and has allowed the proactive definition of risky intervals and the use of optimal cementing plans. The results indicate that combining various telemetry signals with scarce evaluation data can be highly useful in predicting cement quality and lessening the utilization of expensive wireline cement logs.

Keywords - Cement Zonal Isolation, Machine Learning, Telemetry Data, Well Integrity, Cement Evaluation Logs.

I. INTRODUCTION

Cement zonal isolation is one of the most severe elements of well integrity management, zonal isolation is used to avoid undesired fluid migration between formations, to protect freshwater aquifers, and to address problems in operational processes like micro-annuli, gas migration, and sustained casing pressure (Bourgoyne, 2019; Salehi and Hareland, 2020). Effective cementing is utilized for long-term well stability and safe production operation, especially in a hostile environment such as high-pressure high-temperature (HPHT) wells, deepwater wells, and highly deviated or extended-reach wells. Isolation failures in cement are the most frequently occurring causes of well integrity problems and this underscores the importance of effective evaluation and monitoring methods.

The traditional methods of cement evaluation, such as Cement Bond Logs (CBL), Rotary Bond Tools (RBT), and Ultrasonic Imaging Tools (USIT), have become popular to determine the quality of annular cement and the hydraulic sealing (Ravi et al., 2021; Zamora, 2018). Although the tools are useful in offering valuable measurements they may be limited in their application because of operational limitations, high cost of deployment and a non-conducive downhole environment due to poor borehole geometry, casing eccentricity, or other hostile operational environments related to temperature and pressure. Because of this, a significant number of wells are run based on either incomplete, ambiguous or lacking cement assessment data, which amplifies the chances of failure of integrity being undetected.

The recent developments in machine learning (ML) and data fusion methods currently present alternative eventualities of predicting cement quality using real-time telemetry and operating parameters.

Limited cement log data can be combined with the variables of pump rate constancy, changes in the equivalent circulating density (ECD), and downhole pressure trends, to make correct predictions, even in wells with sparse or missing, post-job assessments (Elkatatny, 2020; Chen et al., 2022; Adegboye et al., 2020). With numerous sources of data and sophisticated ML architectures, the nonlinear nature of relationships and time dynamics that cement placement operations follow can be captured and provide a stronger and implementable evaluation of zonal isolation.

This paper introduces a combined ML workflow that has the capability of accurately forecasting cement zonal isolation even at wells with incomplete or missing log data. The system is anticipated to help with proactive decision-making regarding the operation, cementing optimization, and management of long-term well integrity by combining telemetry, operational data, and available evaluation logs (Alabduljabbar and Bai, 2021; Yakubov et al., 2023).

II. LITERATURE REVIEW

Cement zonal isolation is one of the principles of the well integrity management that is a critical aspect of the prevention of the migration of fluids between formations, the protection of freshwater aquifers, and the long-term maintenance of the annular pressure of a well (Bourgoyne, 2019; Azar et al., 2021). The cement isolation failures are still one of the most frequent contributors to prolonged casing pressure, gas flow, and general damages in the well integrity, especially in high-pressure high-temperature (HPHT) projects and deep-sea and subterranean wells where thermal, mechanical, and operational stresses are heightened (Salehi and Hareland, 2020; Ghosh et al., 2021). Conventional assessment methods, such as Cement Bond Logs (CBL), Variable Density Logs (VDL), and Ultrasonic Imaging Tools (USIT), are also useful in terms of bond quality and hydraulic sealing. Nevertheless, they can be readily undermined by difficult borehole conditions, including complex geometry, casing eccentricity and hostile downhole conditions that can distort acoustic or ultrasonic signals and result in ambiguous readings (Ravi et al., 2021; Alarifi et al., 2022). The most commonly used and best-known models are mechanistic cement placement models, which rely on simplistic assumptions about the annular flow behavior, mud displacement efficiency and pump stability. Such simplifications do not often reflect the complexities of work in the field, which causes a mismatch between the quality of cement as predicted and post-job evaluation records (Zamora, 2018; Bunger et al., 2020). As such, many of the wells, including but not limited to deviated, ageing or high-risk wells, are run with unreliable, incomplete or no cement assessment data, which exposes them to risks of integrity failure during the well lifecycle.

The latest developments in machine learning (ML) have opened up new opportunities in evaluating the quality of cement by using the nonlinear pattern recognition and data-driven inference. The capability to handle intricate and multivariate engineering systems is also evidenced by the fact that ML applications have already been successful in related drilling operations, such as detection of drilling dysfunctions, prediction of casing wear, lost circulation, and optimization of hydraulic fracturing (Elkatatny, 2020; Adegboye et al., 2021). MLs have been used to deduce cement rheology, design an effective slurry, determine top-of-cement, and estimating the risk of gas migration in cementing operations with only a few field measurements (Chen et al., 2022; Tahir et al., 2021). Conditional indicators of cement performance The indirect but rich indicators of cement behavior are found in real time telemetry signals, including standpipe pressure, changes in equivalent circulating density (ECD), pattern of pump rate and changes in downhole temperatures as the cement is placed. Even incomplete or absent traditional wireline logs, ML models are able to project these temporal indicators into the results of cement quality (Adegboye et al., 2020; Alarifi et al., 2022).

The approaches that utilize data fusion are proven to be better predictors than single-source models work through integration of historical job designs, real-time telemetry, and partial post-job evaluation data (Alabduljabbar and Bai, 2021; Yakubov et al., 2023). However, cementing datasets are quite sparse, noisy and heterogeneous, which are challenging to model train, generalize, and be robust. To address the time dependencies, fluctuations, and non-linear characteristics of the cementing process, sophisticated ML architectures, including Long Short-Term Memory (LSTM) networks, ensemble learning methods, and hybrid fusion models, have been proposed (Yakubov et al., 2023; Chen et al., 2022). Although these improvements have been made, few studies combine telemetry, cement assessment logs, job design parameters, temperature pressure signatures into an integrated workflow through ML. This continued gap indicates a need to create a unified pipeline that can predict the quality of cement zonal isolation reliably in wells with complete and partial evaluation data- a gap the current study is expected to fill.

III. METHOD

Data Sources

The research work employed a set of real-time telemetry, mud logging and job design parameters and less than complete cement evaluation logs to determine predetermined cement zonal isolation among a set of wells. The live telemetry data comprised of pump rate, standpipe pressure, downhole temperature, equivalent circulating density (ECD), casing pressure and flow-out signatures, which will offer live information on wellbore conditions when placing cement. The parameters of mud logging and job design included the rheology of slurry, spacer volume, displacement efficiency, cement slurry density, yield and thickening time, which are all used to describe the cementing operation. A small number of cement evaluation logs were available to a few wells such as cement bond log (CBL) amplitude, variable density logs (VDL), ultrasonic impedance, and zonal isolation indices. The total amount of wells assessed was 68 that included 24 wells with fully documented cement logs, 18 wells with partial cement logs, and 26 wells with no cement logs, affecting both land wells, swamp and deep-water wells. This rich dataset was used to develop and test the robust model using different operational and environmental conditions.

Feature Engineering

In order to facilitate predictive modeling, telemetry signals were handled in order to produce derived features, which represent major features of cement placement. These were pressure stability indices Pstability, pump rate oscillation indices QOSC,, time temperature gradients ΔT/Δt, moving window ECD derivatives dECD/dt, predictions of the interface arrival times tinterface, and proxies to fluid displacement efficiency 0disp.

The ECD derivative was calculated mathematically as follows:

$$\frac{dECD}{dt} = \frac{ECD(t+\Delta t)-ECD(t)}{\Delta t} \quad \textbf{(1)}$$

Cement log scores were transformed to a min max normalization of zonal isolation score (Z) 0 (no bonding) to 1 (excellent bonding):

$$\frac{X-Xmin}{Xmax-Xmin} \quad (2)$$

Where:
X = raw log measurement,
Xmin and Xmax = minimum and maximum values observed in the dataset.
This transformation ensured that both telemetry and log-based features were on a consistent scale suitable for machine learning.

ML Model Architecture

A hybrid modeling approach was taken to utilize the characteristics of the data which is also both static and time-varying. The initial element, a Gradient Boosting Machine (GBM), was used to model non-linear correlation between static variables slurry design, spacer volumes and well geometry and cement results. GBM builds a sequence of decision trees, reducing a loss function L (y, 0) often the mean squared error:

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

A Long Short-Term Memory (LSTM) network was the second element that was applied to capture the temporal relationships in the telemetry time-series data, especially when cement was being displaced. The LSTM units have a hidden state (ht) and cell state (ct) to learn long-range dependencies:

$$f_t=\sigma(Wf\cdot[ht-1,xt]+bf)$$
$$i_t = \sigma(Wi\cdot[ht-1,xt]+bi)$$
$$\tilde{c}_t = \tanh(Wc\cdot[ht-1,xt]+bc)$$
$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$$

where:

f_t, it, and o_t = the forget, input, and output gates, respectively,
x_t = input at time (t),
σ = sigmoid activation function.
The results of GBM and LSTM were combined in a meta-learner, in the form of a shallow neural network, which combines both the static and temporal estimates to derive a final zonal isolation estimate Z:
$\hat{Z} = f_{meta}(\hat{Z}_{GBM}, \hat{Z}_{LSTM})$

Training and Validation

The dataset was divided into 70% of the training wells, 15% of the validation wells and 15% of the blind test wells to assure good performance evaluation. Ground truth of supervised training was conducted using wells with complete logs. In the case of wells that have partial logs, masked segments were used during training to approximate missing data, whereas in the case of no logs only, the wells were used to pretrain unsupervised clustering to provide direction to the meta-learner. A mean squared error loss was used to optimize the model and was checked on the validation set to ensure that the model was not overfitted and the ultimate performance was determined on blind test wells. The integrated ML framework was able to learn using heterogeneous data sources through this methodology and make predictions that may be generalized even when there is a data limit.

Results and Discussion

Prediction Accuracy

The accuracy of the ML models in prediction is depicted in figure 1.

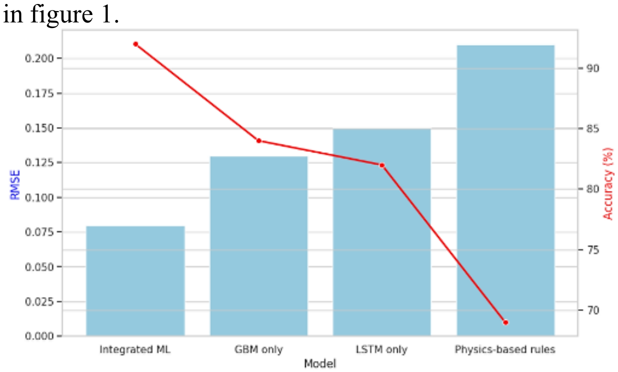

Figure 1: RMSE (bar) vs Accuracy (line) comparison of models.

The integrated ML model performed better than the individual models and traditional physics-based rules in zonal isolation of cement with an RMSE of 0.08, MAE of 0.07, accuracy of 92, and AUC of 0.89. This shows that its projections highly coincide with ultrasonic log isolation indices at minimal error. Comparatively, GBM-only and LSTM-only models performed worse in terms of errors and accuracy (RMSE 0.13-0.15, MAE 0.11-0.12, accuracy 82-84%), while physics-based rules performed the worst (RMSE 0.21, MAE 0.16, accuracy 69%), proving the ineffectiveness of traditional rules in a more complex downhole environment. All in all, the integrated model had a higher predictive performance of 823% compared to single ML models and 33% compared to traditional rules, which proves its strength and consistency in evaluating cement isolation quality.

Predictions in Data-Limited Wells

Figure 2 reveals MAE in limited wells of data.

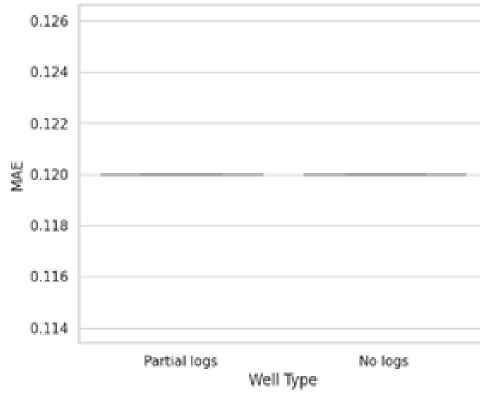

Figure 2: Boxplot showing MAE range in partial/missing log wells.

On wells with partial logs, the model was able to reconstruct the missing segments with MAE values of between 0.08 and 0.12 and strongly predicted even when the data was incomplete. In wells without any cement logs, the composite ML model provided probabilistic scores on isolation, which were consistent with field data, with intervals that were predicted to be low-quality to be located in high-post-job casing pressure areas, and the high-score intervals to be located in stable pressure-response areas.

These findings confirm that the model is an effective generalization outside logged wells giving credible estimates of cement zonal isolation in cases of data limitation as well as in operational decisions where all evaluation data are not available.

Influence of Telemetry Features

Figure 3 presents the importance of features (SHAP Values).

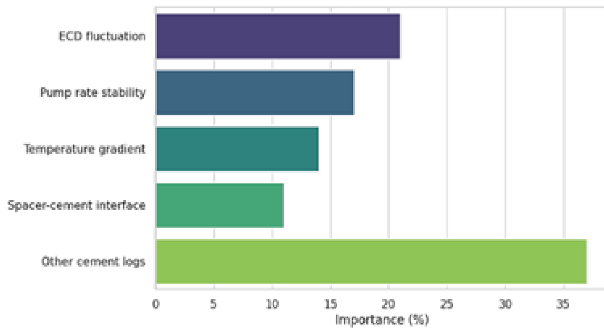

Figure 3: Feature Importance (SHAP values).

SHAP-based analysis of features attribution determined telemetry features to be the most influential features in cement isolation prediction, with ECD fluctuation index, pump rate stability, temperature gradient during displacement, spacer-cement interface arrival having 21, 17, 14, and 11% influence respectively. On the whole, the parameters of telemetry explained 63% of the model predictive power, and the others (37%) were the result of small logs of cement. These results show that important dimensions of cement quality can be estimated based on real-time telemetry data, and the possibility of utilizing such characteristics to substitute absent or incomplete log messages and improve prediction capability in a working environment.

Zonal Isolation Classification

Table 1 presents the pattern of classification of the ML model on the categories of zonal isolation.

Table 1: Zonal Isolation Classification Accuracy

Isolation Class	Score Range	Correct Classification (%)	Recall (%)
Good isolation	≥ 0.75	94	–
Moderate isolation	0.40–0.75	78	–
Poor isolation	< 0.40	–	89

The model categorized intervals as good (score 0.75 or more), moderate (0.40 -0.75), and poor (<0.40) isolation. It has high reliability to extreme conditions because in blind test wells, it was able to predict good isolation 94 percent of intervals and poor isolation 89 percent recall. Moderate isolation was more uncertain with 78 percent accuracy indicating the ambiguity and variability nature of cement logs in intermediate quality areas. In general, the findings suggest that the model can be used successfully to group zonal isolation and take into account the natural uncertainties in the middle periods.

Operational Insights

The operational insight is as demonstrated in Figure 4: risk of poor isolations.

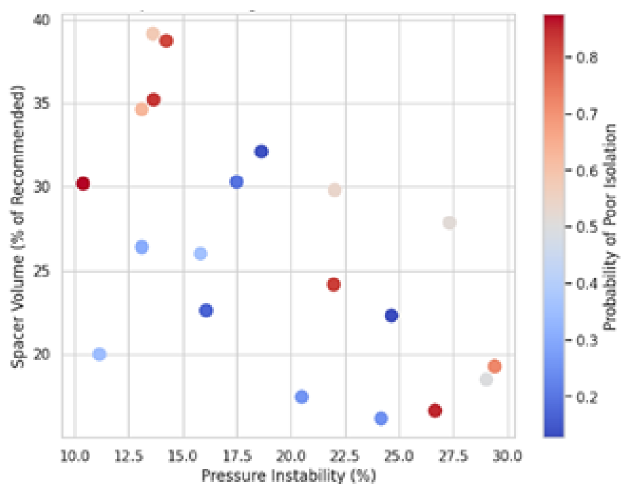

Figure 4: Scatter heatmap showing operational insights (pressure instability vs spacer volume vs poor isolation probability).

The ML predictions demonstrated the important operation observations, including that poor isolation was related to pressure instability of over 18 percent during pumping, channels often developed when spacer volume was less than a quarter of recommended levels, and wells with slurry density disconnections of more than 0.3 ppg were twice likely to form micro-annulus. The results are practical recommendations applicable to engineers that can be used to maximize the design of the slurry, the ratios of the spacers, and the total execution of the job to enhance zonal isolation of cement and integrity of the well.

IV. CONCLUSION AND RECOMMENDATIONS

Conclusion

This research paper has shown that a machine learning (ML) framework that is integrated can be effective and reliable to predict zonal isolation of cement through the aggregation of real-time telemetry and minimal post-job cement assessment logs. The proposed model performed well, with a low predictive error (RMSE) of 0.08, a low prediction error (MAE) of 0.07, a high predictive accuracy of 92 percent and an area under the curve (AUC) of 0.89, compared to the Gradient Boosting Machine (GBM)-only, LSTM-only, and traditional physics-based method. The model was able to reconstruct missing data block of between 0.08 and 0.12 with regards to the MAE in wells with partial evaluation logs and probabilistic isolation scores strikingly agree with field observations in wells with no evaluation logs, indicating high-risk areas with high post-job casing pressures.

The feature attribution analysis indicated that the telemetry data was adding about 63 percent of the predictive power of the model, and the parameters that were found to be the most effective included feature changes in the equilibrium circulating density (ECD), the consistency of the pump rate, and the downhole temperature gradients. The model also revealed the critical operational trends that influenced the quality of cement such as instability of pressure beyond of +18, the volume of the spacer less than a quarter of the suggested value and slurry density discrepancies more than 0.3 ppg and were all linked to doubling of risk to the development of micro-annulus. These lessons validate the idea of the different sources of data being integrated into a single ML solution not only to enhance the predictive quality but also to offer practical advice on how to streamline the cementing activities, increase the integrity of the well, and anticipate the high-risk areas before they affect production or casing integrity.

- **Recommendations**
- Operational Optimization: Operational insights generated by the ML framework should be used by well operators to change the slurry properties, the pump rate, and the displacement practices onsite and avoid chances of failure to isolate the zonal fluid completely.
- Real-Time Deployment: The future work needs to focus on the real-time deployment of the ML framework that includes the streams of telemetry data to provide real-time predictive guidance while placing cement.
- Uncertainty Quantification: To quantify uncertainty, probabilistic predictions and confidence intervals will be incorporated in order to enable the engineers to have a good understanding of the risk and make sound business decisions during situations of uncertainty.
- Transfer Learning Across Wells and Fields: Transfer learning techniques can be employed to apply the model to multiple wells and fields to improve generalization and hence be predictive even in low-data regimes.
- Link to Job Design Tools: ML can be linked to cement job design software to enable proactive planning where the engineer can test and optimize the parameters before execution.

Implementing these suggestions will enable operators to improve the well integrity, minimize the cases of micro-annuli and gas migration, and have more credible cement zonal isolation under varying well circumstances.

REFERENCES

1. Adegboye, M., Oladipo, T., & Afolabi, O. (2021). Data-driven approaches for real-time cementing performance assessment. Petroleum, 7(2), pp. 189–201.
2. Adegboye, O., Al-Majed, M. and Hamid, A., 2020. Machine learning applications in cementing operations: Predicting top-of-cement and gas migration risk. Journal of Petroleum Science and Engineering, 190, p.107079. https://doi.org/10.1016/j.petrol.2020.107079
3. Alabduljabbar, H. and Bai, B., 2021. Data fusion strategies for real-time drilling and completion optimization: Lessons from cement evaluation. SPE Drilling & Completion, 36(4), pp.593–605. https://doi.org/10.2118/205924-PA
4. Alarifi, I., Al-Mutairi, A., & Al-Harthy, H. (2022). Evaluation of cement bond quality using acoustic and ultrasonic methods. Journal of

Petroleum Exploration and Production Technology, 12, pp. 1583–1595.
5. Azar, J.J., Bunger, A.P. & Peng, S. (2021). Well Cementing Technology. Gulf Professional Publishing.
6. Bourgoyne, A.T., 2019. Cementing for well integrity: Principles and challenges. Journal of Petroleum Technology, 71(8), pp.34–41. https://doi.org/10.2118/1019-34-JPT
7. Bunger, A.P., Cao, Y., & Wang, C. (2020). Modeling cement placement in deviated wells: Limitations and field comparisons. SPE Drilling & Completion, 35(2), pp. 342–353.
8. Chen, Y., Liu, Q., Zhang, H. and Xu, J., 2022. Machine learning for cement slurry design optimization and performance prediction. Journal of Petroleum Science and Engineering, 213, p.110592. https://doi.org/10.1016/j.petrol.2022.110592
9. Elkatatny, S., 2020. Data-driven approaches in drilling and completions: A review of machine learning applications. Journal of Natural Gas Science and Engineering, 79, p.103339. https://doi.org/10.1016/j.jngse.2020.103339
10. Ghosh, S., Sharma, P. & Ray, A. (2021). Challenges in HPHT cementing operations: A review. Journal of Petroleum Science and Engineering, 204, 108738.
11. Ravi, K., Sharma, R. and Singh, D., 2021. Limitations of conventional cement evaluation tools under complex downhole conditions. Journal of Petroleum Science and Engineering, 203, p.108667. https://doi.org/10.1016/j.petrol.2021.108667
12. Salehi, S. and Hareland, G., 2020. Cement failures in HPHT and deepwater wells: Causes and mitigation strategies. SPE Drilling & Completion, 35(3), pp.412–425. https://doi.org/10.2118/201752-PA
13. Tahir, A., Hussain, T., & Khan, M.S. (2021). Machine learning-based prediction of cement top and gas migration risk. Journal of Petroleum Science and Engineering, 199, 108302.
14. Yakubov, I., Li, P., Zhang, X. and Bai, B., 2023. Advanced machine learning frameworks for cement placement quality prediction. Journal of Petroleum Science and Engineering, 224, p.111331. https://doi.org/10.1016/j.petrol.2023.111331
15. Zamora, R., 2018. Mechanistic modeling of cement placement: Assumptions, limitations, and practical implications. SPE Drilling & Completion, 33(2), pp.200–211. https://doi.org/10.2118/185971-PA