

Data-Driven Interpretation of Long-Term Stability of Oilwell Cements Using ML and Deep Learning Techniques

Jocin James Abraham, Mona Amani and Mahmood Amani

Petroleum Engineering Department
Texas A&M University at Qatar
Doha, Qatar

jocin.abraham@qatar.tamu.edu, mona.amani@qatar.tamu.edu, mahmood.amani@qatar.tamu.edu

Catalin Teodoriu

Mewbourne School of Petroleum and Geological Engineering
University of Oklahoma
Norman, Oklahoma, USA
cteodoriu@ou.edu

Abstract

Operational safety, efficiency and reliability have been key priorities for any oil and gas project. During drilling and production operations, while several safety barriers, controls and mechanisms are incorporated into the overall project plan, the reliability of the cement layer is vital to prevent catastrophic incidents and failures like blowouts. While properties of cement have been studied extensively in the laboratory, evaluating its long-term downhole behavior remains challenging due to practical limitations of cost, time and access. This study leverages machine learning (ML) strategies to determine the mechanical properties of cement such as compressive strength (UCS) and mechanical stability over time using available experimental laboratory data. An extensive experimental dataset of over 1,100 oilwell cement samples was tested at temperatures ranging from 25°C to 75°C, and the compressive strength (UCS) was measured to determine stability. Experimental results revealed that while higher curing temperatures (75°C) accelerated early-stage UCS development (1-3 days) compared to lower temperatures (up to 15 days), they suffered significant thermal degradation, reducing peak UCS by at least 20%. Using this laboratory data, three predictive models were developed and evaluated: Random Forest (RF), Gradient Boost (XGBoost), and Support Vector Machines (SVM). All models performed well on testing, with high R^2 values of 0.9166, 0.8866 and 0.7156, indicating that machine learning strategies can reliably predict intermediate-to-long-term cement strength under varying downhole conditions. Ultimately, this research provides operations managers and petroleum engineers with data-driven predictive tools thereby enhancing well integrity management.

Keywords

Machine Learning, Oilwell Cement, Cement Compressive Strength, Neural Network and Deep Learning.

1. Introduction

Oil and gas drilling operations require meticulous planning and execution, with a paramount focus on ensuring safe and efficient operations across the entire lifecycle of the project. Maintaining safety during drilling, completions, production, and eventual plugging and abandonment (P&A) involves a complex interplay of engineering practices, specialized equipment, and barrier matrices that collectively guarantee wellbore integrity. Modern oil and gas wells are designed with multiple primary and secondary barriers, alongside several physical and mechanical redundancies to prevent the catastrophic, unintentional flow of formation fluids into the wellbore. This well integrity framework

includes drilling fluids as a dynamic barrier, complemented by mechanical and physical barriers such as valves, blowout preventers (BOPs), packers, casing strings, and cement layers (James Abraham & Amani, 2023). Within this well integrity matrix, the cement layer plays a foundational role as a passive, permanent barrier required to remain fully operational for the entire lifespan of the well.

As the definitive first layer of protection, a comprehensive understanding of the casing-cement-formation interaction is essential. The primary responsibility of the cement layer is to provide an impermeable seal in the annular space between the casing and the surrounding rock, achieving complete zonal isolation. Achieving this seal depends entirely on the cement's ability to form a continuous, homogeneous, and flaw-free barrier (Teodoriu et al., 2016). A well-designed cementing program is critical to preventing influxes or kicks by stopping the migration of unwanted fluids into the wellbore, securing the casing and tubing strings in place, and preventing critical failures like borehole collapse or lost circulation (Abraham et al., 2021; Bourgoynne et al., 1986).

Despite significant technological leaps in drilling mechanisms, downhole completion tools, and operational methodologies, the primary materials used for oilwell cementing have remained largely unchanged for the better part of the last few decades. The formulation of oilwell cement, components and additives as well as the cement mixing guidelines utilized today are nearly identical with those formulated several decades ago by the American Petroleum Institute (API). Because of this, and due to the ever-increasing complexity of oil and gas operations, the success of a cementing job hinges heavily on precise downhole placement, slurry optimization, and a deep understanding of how the cement's mechanical and thermal properties evolve under subsurface stresses.

1.1 Subsurface Stresses and Failure Mechanisms

Failing to properly test and characterize cement behavior under downhole conditions introduces severe well integrity risks. The use of untested additives or incompatible cement formulations can lead to inadequate bonding, resulting in the creation of permeable channels leading to the loss of zonal isolation that allow formation fluids to bypass the cement barrier. Furthermore, the cement sheath is highly vulnerable to the dynamic stresses of the subsurface environment and without rigorous pre-cementing testing, cracks or fractures can occur in the cement layer again leading to loss of zonal isolation. Additionally, as the wellbore is subjected to fluctuating reservoir pressures, natural tectonic stresses, and thermal expansion or contraction, the cement can debond from the casing-formation interface. This structural failure typically manifests as microfractures within the cement matrix or the formation of a micro-annulus between the cement sheath and the steel casing.

Beyond mechanical stresses acting on the cement layer, the chemical stability of the cement barrier is also constantly affected by the downhole environment. Sustained interaction with corrosive formation fluids and chemically reactive injected fluids such as carbon dioxide, hydrogen sulfide, hydrocarbons, and corrosive brines can accelerate chemical degradation via corrosion or erosion. Each of these chemical and physical interactions must be thoroughly evaluated for long-term compatibility to prevent the failure of the cement layer and the subsequent loss of well control (Al-Ramis & Teodoriu, 2020).

2. Literature Review

Although oilwell cements have been studied for decades (API, 1979; Carter & Evans, 1964; Ghabezloo et al., 2008; James Abraham et al., 2023; Saleh et al., 2018; Teodoriu et al., 2013; Walter et al., 2003), a critical disconnect remains, with no single study being able to characterize the different cement formulations, and the evolution of the cement properties over time. The vast majority of historical literature relies on conventional laboratory testing mostly at standard conditions. Even with real-world testing conditions, the studies have been severely constrained by the scope and feasibility of recreating and monitoring evolution of properties under downhole conditions. Most studies are also often highly specific, isolating a single physical property under controlled conditions that fail to mirror the multi-axial, simultaneous stresses encountered downhole.

To bridge this gap, researchers have progressively focused on the long-term thermal and mechanical evolution of cement barriers. Early foundational work by Saunders and Walker (1954) established that elevated downhole temperatures induce thermal dehydration. This process accelerates setting times and triggers severe regression of cement mechanical strength over extended periods. Work done by Teodoriu et al. (2013) demonstrated that incompatible slurry formulations subjected to high-pressure, high-temperature (HPHT) conditions rapidly compromise the cement structural integrity, leading to barrier failure. Literature has also noted that over extended

timelines, the mechanical strength of oilwell cement evolves dynamically, which is a critical consideration for long-term plugging and abandonment strategies (Walter et al., 2003). Operational loads, such as the high pressures encountered during hydraulic fracturing can trigger mechanical failure of the cement (Cerasi et al., 2020; Kim et al., 2016), ultimately leading to the formation of flow pathways through the damaged cement layer during active drilling and production phases (Al-Ramis & Teodoriu, 2020).

Additionally, the physical properties of the cement can also be severely compromised even before the cement is cured downhole. During placement and circulation, the head and tail of the cement slurry column are highly susceptible to contamination from drilling muds, which alters the slurry's chemistry and degrades the final mechanical properties of the set cement. High-temperature downhole zones further complicate this process, disrupting standard placement techniques and cement curing mechanics, which underscores the need for alternative formulations to mitigate temperature-induced placement risks (Arbad et al., 2022). All these demonstrate the challenges in accurately predicting the long-term properties of the cement under downhole conditions using conventional field and laboratory testing methods.

The rapid evolution of machine learning (ML) and data-driven modeling can offer several solutions for optimizing oilwell cementing programs. While the oil and gas sector has only adopted data-driven strategies recently, civil and mechanical engineering disciplines have long established a vast foundation of literature using data-driven workflows to model concrete performance, especially for construction. Rather than relying solely on repetitive, destructive laboratory testing, supervised ML strategies present a significant opportunity to predict crucial mechanical parameters of oilwell cement such as Poisson's ratio, compressive strength, tensile strength, and yield strength. With the availability of several decades of data on cementing programs, data-driven solutions can be developed with reasonable accuracy to predict the mechanical properties of oilwell cement and the long-term behavior, especially under downhole condition.

Building on these cross-disciplinary insights, recent studies have begun adapting these algorithms to model and predict properties of oilwell cement. Nyakilla et al. (2023; 2022) aggregated datasets from baseline oilwell cement slurries and those enhanced with novel additives. By applying algorithms such as Gaussian Process Regression, Neural Networks, Regression Trees, Back Propagation Neural Networks (BPNN), and Group Method of Data Handling (GMDH), they successfully forecasted both Unconfined Compressive Strength (UCS) and Shear Bond Strength (SBS). Advanced neural networks and clustering techniques have also been successfully deployed to automate and support wellbore integrity analysis (Abid et al., 2023) and optimize diagnostic workflows for plugging and abandonment operations (Bose et al., 2021; Camerini et al., 2023). These breakthroughs demonstrate that supervised machine learning models are highly capable of mapping hidden patterns within massive, multi-variable material datasets. By leveraging these predictive tools, engineers can radically accelerate oilwell cement formulation testing and proactively design slurries tailored to withstand downhole environments.

This study aims to develop a framework for utilizing experimental laboratory data along with data driven ML models to accurately predict the properties of cement downhole, particularly the uniaxial compressive strength. This study addresses the limitations in conventional studies by comprehensively evaluating the long-term evolution of mechanical properties and chemical stability across various oilwell cement classes and providing an assessment of their performance and durability, especially during long-term operations. Experimental studies are conducted using three different classes of oilwell cement - Class C, Class G and Class H, and this information is used to model several supervised ML models to predict cement properties in the long term.

3. Methodology

To evaluate oilwell cement performance across varying subsurface conditions, this study several classes of cement standardized by the American Petroleum Institute (API, 1979). While the API categorizes oilwell cements into classes A through H based on target depth, chemical composition, and environmental resistance, this research utilizes Classes C, G, and H due to their widespread adoption in standard drilling and completion operations (API, 1979). While Classes G and H are utilized widely for their chemical stability and broad adaptability across high-pressure, high-temperature (HPHT) regimes, Class C cement is utilized for specific zonal requirements (James Abraham et al., 2023). Specialized cement slurry formulations are also required to adapt to weak geological formations, adjust slurry density, and control thickening intervals and setting times. For this, oilwell cement formulations integrate an array of additives including fluid loss control agents, retarders, thickening agents, as well as loss control additives which are mixed in

specified concentrations for specific operations. However, to establish clear mechanical and chemical baselines before modeling these complex formulations, this study primarily prioritizes the behavior of neat cement formulations. The formulation, mixing and curing conditions, sample size, and classification are shown below in Table 1.

Table 1. Information on the samples tested, experimental conditions including ageing periods

Classification	Description	Samples Tested	Ageing Periods	Temperatures	Curing Conditions
Class 1	Class C Cement	366	1-366 days	25°C, 75 °C	Dry, Wet
Class 2	Class G Cement	543	1-2866 days	25 °C, 75 °C	Dry
Class 3	Class H Cement	210	1-1012 days	25 °C, 75 °C	Dry, Wet

Preparation of cement samples and formulation is performed as per API Standard 10B (1979) guidelines to ensure experimental reproducibility. Mixing was conducted using standardized high-shear cement blenders to ensure a uniform slurry matrix as shown in Figure 1. The fresh slurries were subsequently cast into cubical molds. To map the progressive development cement properties, the specimens were transitioned into specialized curing environments mimicking limited downhole thermal conditions. Three different curing conditions were evaluated – at ambient room temperatures (25°C) vs. higher downhole temperature condition (75°C), wet vs. dry curing as well as ageing periods ranging from a single day to over a year for specific long-term stability tracking (Abraham et al., 2024).



Figure 1. Experimental testing of the samples - curing of samples in temperature-controlled water baths, stress testing and acoustic testing

A total repository exceeding over 1,000 samples was generated and tested to measure their properties including density, viscosity, specific gravity, other rheological properties, sample weight and dimensions, gel strength, as well as mechanical properties including uniaxial compressive strength (UCS), acoustic properties (UPV), failure envelope, as well as stress and strain maps by subjecting the cement samples to destructive and non-destructive analysis. This extensive testing regimen aimed to construct a highly representative dataset which could be utilized for training advanced analytical models, including predictive tools for long-term cement properties. While the broader database developed from these experimental investigations, as well as those performed earlier by Abraham et al. (2024; 2023) tracks long-term degradation, a vital subset of this was utilized in this study to focus on the initial 30-day window - a critical phase where compressive strengths reach their peak before stabilizing.

The experimental data was then utilized to develop and train machine learning (ML) models which could help evaluate and scale their lab-derived properties to a field scale, especially over the long term. Of the different mechanical parameters evaluated for oilwell cement, uniaxial compressive strength (UCS) was of critical focus within this study. To predict UCS without relying solely on destructive testing, several supervised machine learning techniques were developed to evaluate functional parameter relationships. Specifically, three distinct supervised algorithmic frameworks - XGBoost (XGB), Random Forests (RF), and Support Vector Machines (SVM) were developed and trained using the information from the dataset. The predictive accuracy and operational efficiency of each resulting model were then rigorously evaluated using a testing dataset against one another to identify the most robust predictive approach.

4. Data Preparation and Processing

By utilizing supervised machine learning (ML) algorithms, the objective was to build ML models that could evaluate the various parameters and determine complex, time-dependent patterns in the mechanical properties to effectively predict them during the operational lifecycle of a well. Given oilwell cement integrity was the primary objective, uniaxial compressive strength (UCS) was isolated as the primary target variable. Three distinct supervised learning models were developed and tested to predict UCS using non-destructive inputs like ultrasonic pulse velocity (UPV), density, rheological properties and curing conditions.

The various ML models were trained and validated using a dataset containing data from 1,119 individual samples of API Class C, G, and H cement specimens. The dataset contained information on the cement sample curing duration (age), ambient exposure temperature, sample dimensions, cross-sectional surface area, mass, density, specific gravity, acoustic behavior through ultrasonic pulse velocity (UPV) as well as the unconfined compressive strength (UCS), which was the target variable. The statistical spread of the various variables in the dataset is shown below in Table 2, demonstrating an average UCS of approximately 34 MPa along with a mean acoustic velocity of 16 m/s.

Table 2. Dataset used for machine learning models, various parameters used and their statistical data

	Age (days)	Temperature (C)	Average Length (mm)	Average Width (mm)	Area (sq mm)	Weight (gm)	Sample SG	Density (Kg/m ³)	UPV (m/s)	UCS (Mpa)
count	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000	1119.000000
mean	20.695764	37.205541	50.850232	50.333798	2555.59538	239.533771	2.134656	17.782583	16.350849	34.616175
std	111.546279	21.586882	3.043095	4.390890	241.16278	48.028689	0.227064	1.890641	2.808605	19.862339
min	0.170000	23.000000	25.030000	25.060000	1154.47000	40.800000	0.010000	0.060000	12.400000	2.340000
25%	3.000000	25.000000	50.830000	50.670000	2574.21000	238.790000	1.990000	16.610000	14.400000	16.520000
50%	7.000000	25.000000	50.900000	50.940000	2592.51000	250.180000	2.050000	17.080000	15.900000	32.710000
75%	21.000000	25.000000	50.970000	52.235000	2659.04500	266.605000	2.220000	18.480000	17.900000	51.300000
max	2246.000000	75.000000	60.870000	55.870000	3216.22000	321.260000	3.240000	26.960000	28.400000	90.080000

To evaluate the variance and detect any anomalies or outliers across key features like density, dimensions, UPV, and UCS, the feature space was visually mapped using descriptive box plots (Figure 2). Though outliers were detected for several features, those values were not excluded during model training and testing, since their impact on model performance was limited. Additionally, ageing of cement samples over several months typically results in several outlier points across the properties of the cement samples, but these cannot be excluded since their inputs are necessary for long-term prediction of properties.

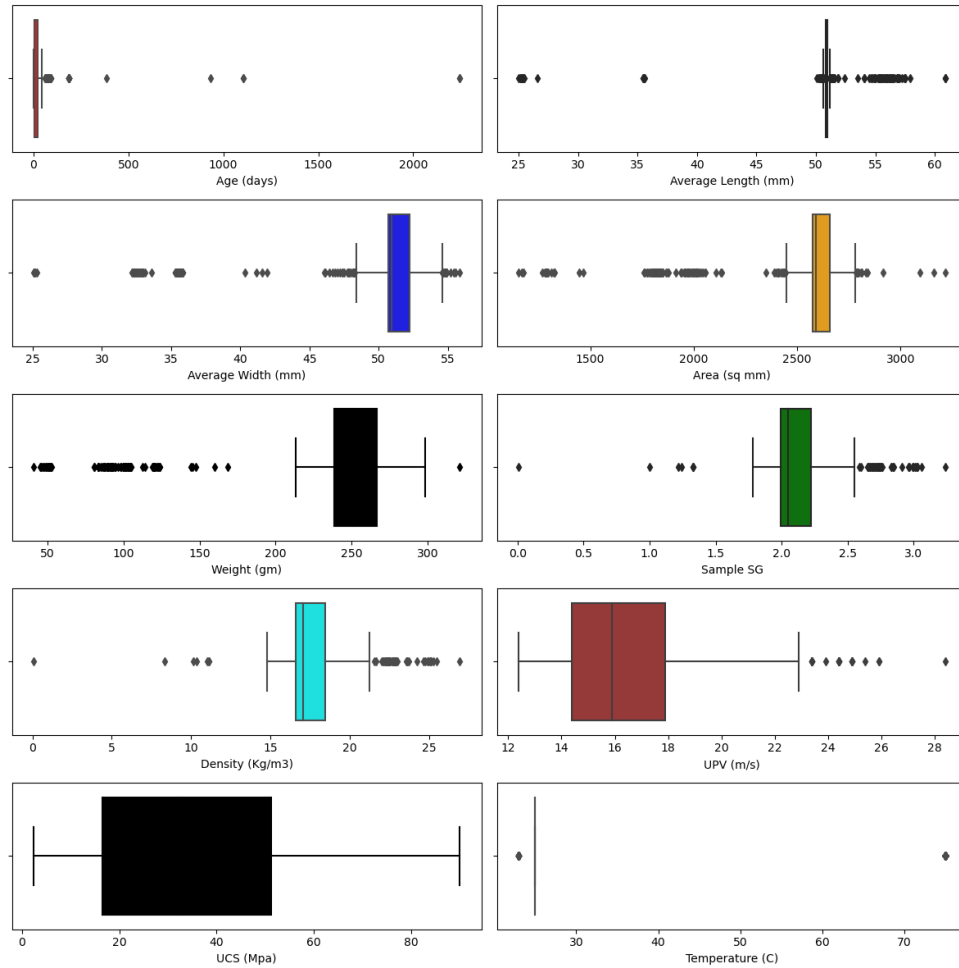


Figure 2. Outlier evaluation for the dataset – no outliers were excluded in training the models

Additionally, to uncover hidden collinearities between different feature parameters, a Spearman correlation matrix was constructed as shown in Figure 3. Identifying these co-linearities is vital as an optimized dataset can help to accelerate model training, prevent overfitting, and optimize computing resource allocation without sacrificing accuracy. The correlation values, bound between -1 and 1, revealed a strong mathematical dependence between specimen density and UPV, alongside a highly significant direct correlation linking UCS to UPV. Crucially, near-perfect direct collinearity was identified between raw bulk density and specimen specific gravity (Sample SG). To eliminate redundant variance, specific gravity was excluded from the input layer. All other mechanical and physical attributes were preserved to guarantee the availability of sufficient datapoints for the predictive models.

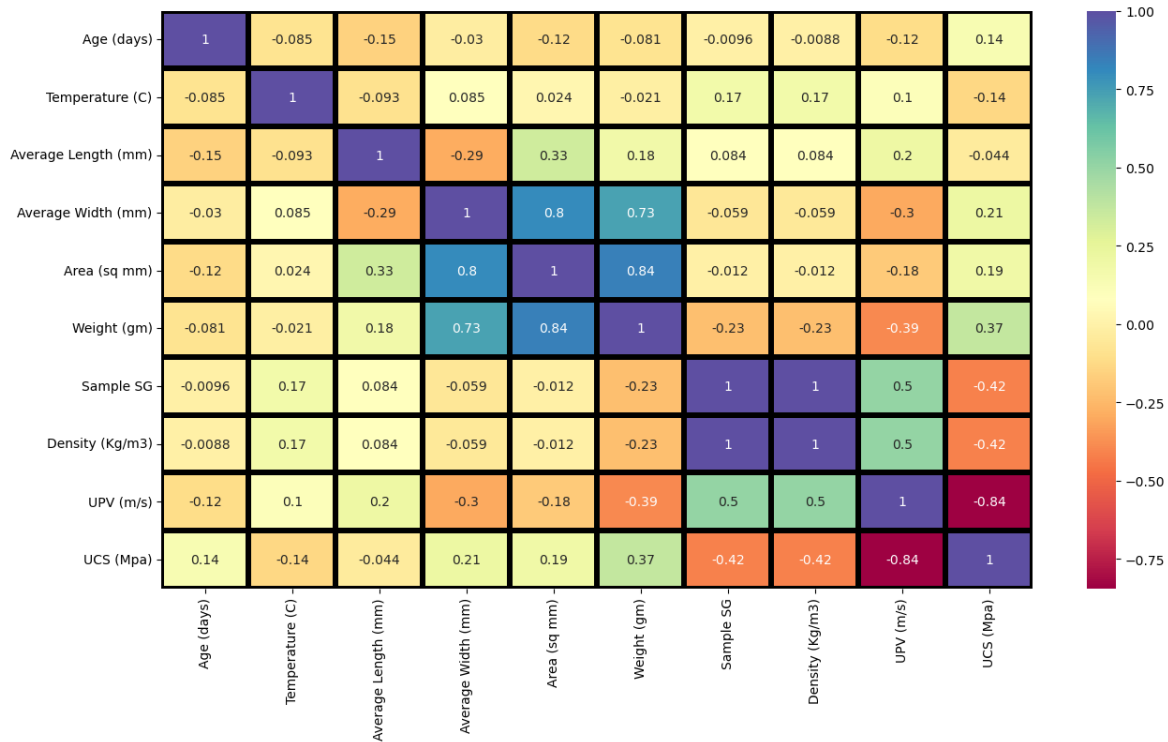


Figure 3. Spearman correlation matrix to identify relationships and collinearities between various parameters – Sample SG was excluded from the feature dataset

For further optimization of the dataset, input variables were scaled to prevent features with larger absolute magnitudes from dominating the training algorithm, especially since outliers were not excluded. This was achieved using the MinMaxScaler, a utility within Python's Scikit-Learn framework. The MinMaxScaler scales the data within a standard [0, 1] range by assigning the largest datapoint as 1 and the smallest datapoint as 0, while all the other datapoints are scaled between these, as shown in Equation 1.

$$\text{MinMaxScaler: } x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} * (max - min) + min \quad (1)$$

To evaluate model reproducibility and ensure accuracy, the complete dataset was also split into a training and a testing subset. Utilizing Scikit-Learn's train_test_split function, 70% of the datapoints were designated for active model training, while the remaining 30% was held back as an independent testing set. Model performance and predictive reliability were subsequently evaluated and quantified using cross-validation scores, coefficient of determination (R^2) metrics, as well as the calculated errors (MAE, MSE, RMSE).

5. Results and Discussion

Experimental investigation of the cement samples was performed to evaluate the UCS of cement samples over time and at different downhole conditions. The data collected from these investigations were utilized to develop a comprehensive dataset of cement properties, which was then utilized for developing ML models.

5.1 Experimental Results

Experimental analysis revealed clear differences in how oilwell cement formulations develop structural strength when subjected to different thermal regimes and curing conditions. As can be seen in Figure 4, the baseline experimental testing captured a vital thermodynamic trade-off between the rate of hydration and long-term compressive strength of oilwell cement. Exposure to higher thermal conditions (75°C) initially accelerated hydration kinetics. Across all three standard oilwell cement classes, samples reached high Unconfined Compressive Strength (UCS) values rapidly, with the cement slurry stabilizing shortly after placement. Conversely, samples cured at room temperature (25°C) followed a much slower development path. These specimens required an extended curing window, often exceeding 30 days on

average to steadily build compressive strength (UCS). However, despite the slow buildup of UCS, samples cured at room temperature (25°C) ultimately surpassed their high-temperature counterparts, achieving superior absolute mechanical strength (UCS) between day 25 and day 30.

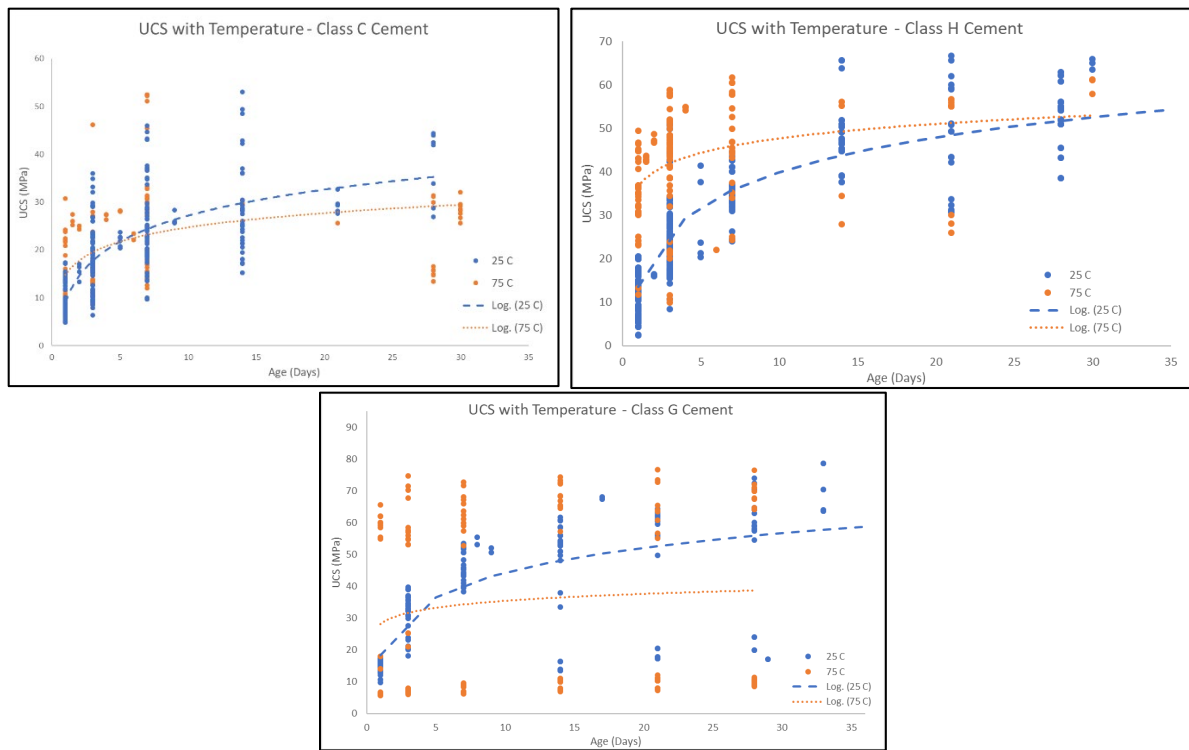


Figure 4. Experimental results of the oilwell cement samples showing their behavior with temperature over time

Among the different cement classes, API Class G achieved the highest overall performance, making it a strong candidate for demanding downhole environments. However, sustained thermal exposure acted as a primary degradation mechanism across every single Portland-based oilwell cement formulation. Thermal aging cut peak structural capacity significantly, forcing a drop in absolute UCS that exceeded 20% in some cases. This consistent downward trend across Classes C, G, and H confirms that testing at standard laboratory conditions can fundamentally overestimate their long-term structural reliability under true downhole conditions. Furthermore, cement strength regression was also noticed over longer ageing periods, indicating that higher temperatures can impair cement strength and performance over time.

All Class C, H, and G cement specimens were also tested for their acoustic velocities (UPV) before destructive compressive testing. This approach aimed at determining how acoustic properties evolve during matrix curing and to uncover any underlying relationships between a sample's acoustic profile and its ultimate uniaxial compressive strength (UCS). The experimental results revealed a distinct inverse relationship between acoustic properties and UCS development. As the cement slurries transitioned from a liquid state to a consolidated matrix, increasing the overall UCS, the corresponding UPV metrics systematically decreased across all tested classes as shown in Figure 5. This acoustic-mechanical behavior is directly related to the physical chemistry of the cement matrix, and this correlation can be incredibly valuable for advanced predictive data analytics to determine residual downhole compressive strengths purely by using non-destructive ultrasonic logging tools.

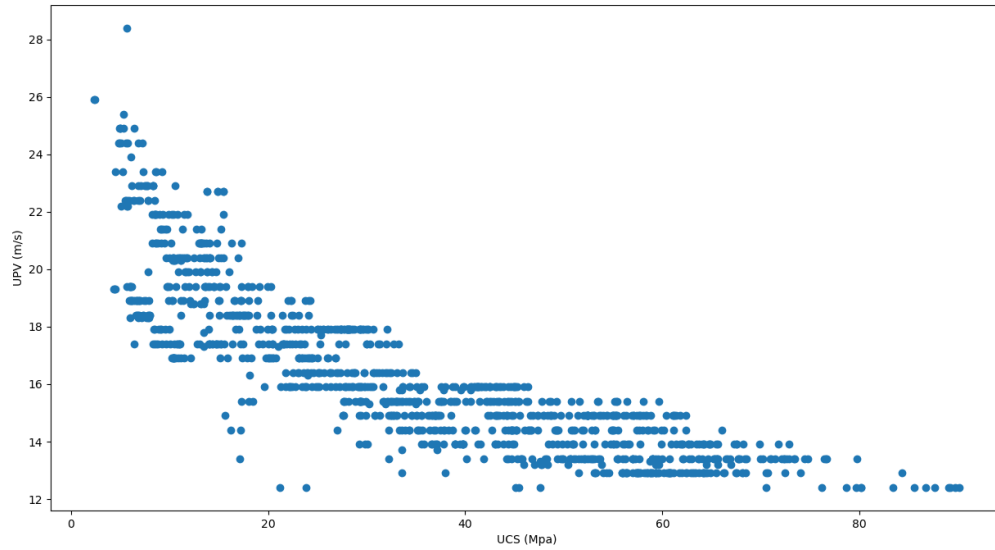


Figure 5. Relationship between uniaxial compressive strength (UCS) and acoustic pulse velocities (UPV) of samples

5.2 Evaluation of Machine Learning Models

To evaluate the accuracy of the developed machine learning models in predicting the uniaxial compressive strength (UCS) of oilwell cement, three distinct algorithms - Random Forest (RF), Gradient Boosting (serving as the proxy for XGBoost), and Support Vector Machines (SVM) - were trained and tested. The models utilized 70% of the data for training and the remaining 30% for testing, with all input features normalized using a MinMax scaler. The predictive performance on the testing dataset was rigorously quantified using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The trained model was then tested using the testing subset, and the predicted values were compared to the actual values of UCS as shown in Figure 6.

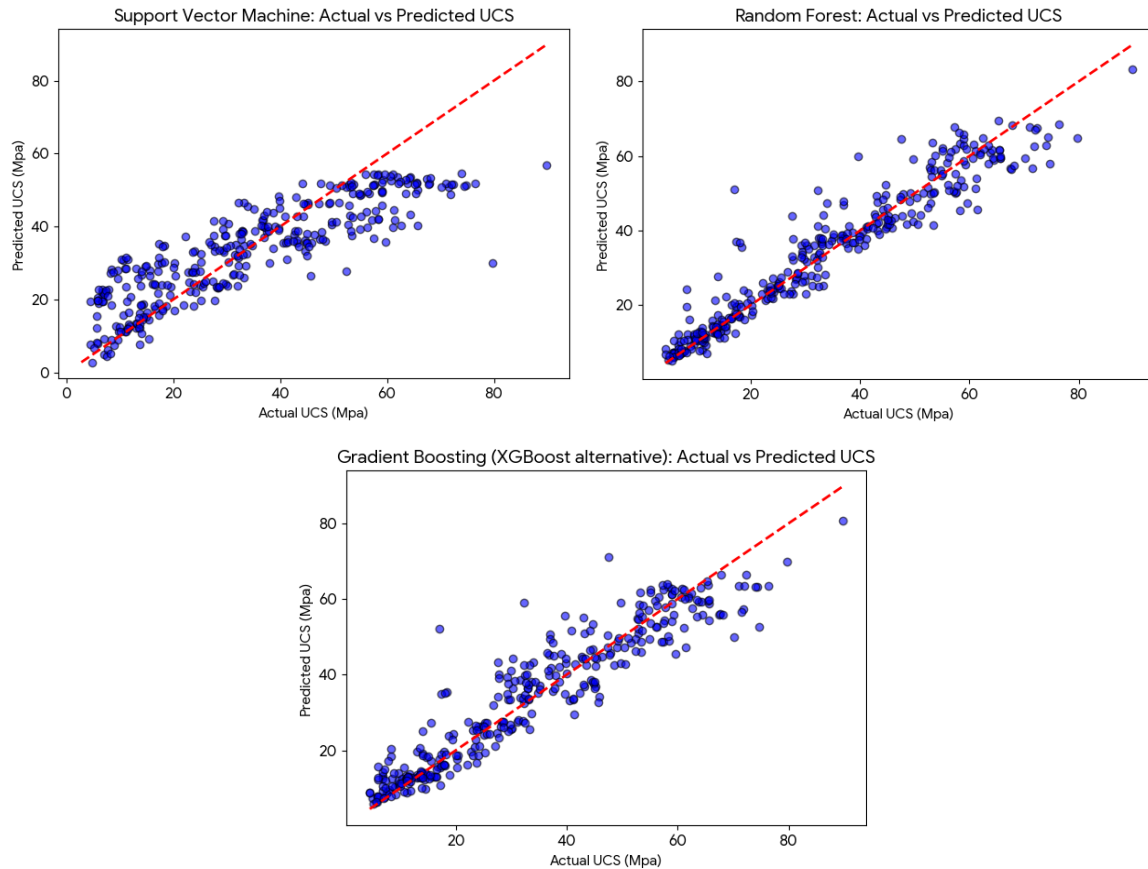


Figure 6. Relationship between the actual and predicted values of UCS for different models – Support Vector Machines (SVM), Random Forest (RF) and Gradient Boosting (XGBoost)

When evaluating the actual and predicted values, the R^2 value measures the proportion of variance in the dependent variable (UCS) that is predictable from the independent input features. The random forest model demonstrated the highest predictive accuracy, achieving an R^2 score of 0.9166. This implies that over 91.6% of the variability in the cement's compressive strength can be successfully explained by the model using physical and mechanical parameters such as age, temperature, ultrasonic pulse velocity (UPV), and dimensional metrics. The gradient boosting model also exhibited strong predictive power with an R^2 value of 0.8866. While slightly trailing behind random forest, it underscores the inherent capacity of tree-based methods to capture complex trends in cement property datasets. Conversely, the support vector machine yielded the weakest correlation, with an R^2 of 0.7156. This indicates that roughly 28.4% of the variance remains unexplained by the SVM model, making it the least reliable option among the three for this application.

Significant evaluation into error metrics provided in Table 3 is also essential to validate the robustness and operational accuracy of the models, as large prediction errors in cement strength could lead to catastrophic structural failures in wellbore integrity. Mean Absolute Error (MAE) provides a direct assessment of the average expected error magnitude. The Random Forest model achieved an MAE of 3.8140 MPa, meaning its predictions deviate from the true laboratory-measured UCS by less than 4 MPa on average. Gradient Boosting followed with an MAE of 4.8061 MPa, whereas the SVM model exhibited a much higher average deviation of 8.0992 MPa. MSE and RMSE are highly sensitive to larger errors because they square the residual values before averaging. For the Random Forest model, the RMSE (5.6929 MPa) is only moderately higher than its MAE (3.8140 MPa). This tight variance between the two errors indicates a highly consistent prediction profile with a minimal presence of severe outliers or wildly inaccurate predictions across the testing spectrum. For the SVM model, the RMSE and MAE values were substantially higher, indicating that the SVM model suffers from occasional but significant prediction failures resulting in poorly fitting certain zones of the dataset (e.g., extremely high or low UCS boundaries).

Table 3. Model validation metrics including the coefficient of determination (R^2), MAE, MSE and RMSE

Model	R^2	MAE (MPa)	MSE (MPa ²)	RMSE (MPa)
Random Forest	0.9166	3.814	32.4092	5.6929
Gradient Boosting (XGBoost alternative)	0.8866	4.8061	44.0533	6.6373
Support Vector Machine (SVR)	0.7156	8.0992	110.5387	10.5137

The validation process confirms that ensemble tree-based machine learning approaches are highly viable for predicting oilwell cement properties. The actual versus predicted plots as well as the coefficient of determination (R^2) values reinforce this conclusion, showing a tight clustering along the ideal 45-degree line for Random Forest and Gradient Boosting, whereas the SVR scatter plot displays significant dispersion. With agreeable outcomes on the error parameters (MAE, MSE, RMSE) as well, the Random Forest model can be recommended as the primary predictive tool for assessing UCS for the available dataset.

6. Conclusion

The structural integrity of a wellbore depends heavily on the long-term reliability of its cement sheath, which serves as a critical barrier throughout the life of the well. This comprehensive study analyzed a dataset of over 1,000 samples to evaluate the mechanical properties of API Class C, G, and H oilwell cements. The experimental findings revealed that while all three cement classes reach a stable structural threshold after roughly 30 days of hydration, their final performance is heavily dictated by composition and environmental conditions. API Class G cements consistently outperforming Class H and Class C variants in absolute uniaxial compressive strength (UCS), indicating they are preferred for high temperature, high pressure operations downhole. Across all evaluated classes, thermally aged cement samples (75°C) exhibited a noticeably lower ultimate UCS compared to those allowed to cure slowly at room temperature (25°C). Furthermore, a distinct, inverse exponential relationship was established between the development of UCS and its corresponding Ultrasonic Pulse Velocity (UPV). This acoustic signature can be utilized as a highly effective, non-destructive proxy for mapping internal cement matrix consolidation, offering a reliable methodology to infer downhole cement strength without resorting to destructive physical testing.

To scale this experimental data, supervised machine learning models - Random Forest (RF), Gradient Boosting (XGB), and Support Vector Machines (SVM) - were trained on a normalized, 70/30 split dataset to forecast UCS. The computational validation confirmed that ensemble tree-based architectures possess an innate capacity to map the highly complex, non-linear variables governing cement strength degradation. The Random Forest model demonstrated the highest predictive accuracy and operational reliability, with an R^2 of 0.9166. Gradient Boosting followed closely behind with an R^2 of 0.8866, whereas the SVM model delivered a poorly correlated fit. A rigorous analysis of residual error metrics further validated the superiority of the Random Forest framework. It achieved a low Mean Absolute Error (MAE) of 3.8140 MPa and a Root Mean Squared Error (RMSE) of 5.6929 MPa. This minimal variance between average and squared error indicators demonstrates that the Random Forest algorithm maintains an exceptionally consistent prediction profile with an absence of severe outliers. Visually confirmed by the tight clustering of data points along the ideal 45-degree line on the actual-versus-predicted UCS plots, the Random Forest model is highly recommended as the primary predictive tool for data-driven well integrity assessments.

References

- Abid, K., Srivastava, S., Tellez, M. L. R., Amani, M., & Teodoriu, C. Experimental and machine learning study of thermal conductivity of cement composites for geothermal wells. *Geothermics*, 110, 102659, 2023. <https://doi.org/10.1016/j.geothermics.2023.102659>
- Abraham, J. J., Carvero, A., Devers, C., Teodoriu, C., & Amani, M. *Stress Failure Assessment and Potential Well Integrity Issues with Different Oilwell Cement Classes and Formulations in Complex Wells: GOTECH*, 2024. <https://doi.org/10.2118/219117-MS>
- Abraham, J. J., Devers, C., Teodoriu, C., & Amani, M. *The Need for a Comprehensive Cement Database—A Novel Approach to Best Practices by Cataloging Cement Properties: Abu Dhabi International Petroleum Exhibition&Conference*, 2021. <https://doi.org/10.2118/208112-MS>
- Al-Ramis, H., & Teodoriu, C. Critical Leak and Mechanism (A Review) with application to Oil, Gas and Geothermal Well Integrity. *Oil and Gas Magazine*, 26–32, 2020.
- API. *API RP 10B: Recommended Practice for Testing Oil-Well Cements and Cement Additives* (21st ed.). American Petroleum Institute, 1979. <https://www.apiwebstore.org/standards/10B>

- Arbad, N., Emadi, H., & Watson, M. A comprehensive review of geothermal cementing from well integrity perspective. *Journal of Petroleum Science and Engineering*, 217, 110869, 2022. <https://doi.org/https://doi.org/10.1016/j.petrol.2022.110869>
- Bose, S., Zhu, L., Zeroug, S., Merciu, I. A., Constable, K., Hemmingsen, P. V., Berg, E., Wielemaker, E., Govil, A., & Kalyanaraman, R. S. *Acoustic Evaluation of Annulus B Barriers Through Tubing for Plug and Abandonment Job Planning: Offshore Technology Conference*, 2021. <https://doi.org/10.4043/31302-MS>
- Bourgoyne, Jr., Adam T., Millheim, K. K., Chenevert, M. E., & Young, Jr., F. S. *Applied Drilling Engineering*. Society of Petroleum Engineers, 1986. <https://doi.org/10.2118/9781555630010>
- Camerini, I. G., Ferreira, G. R. B., de Souza, L. P. B., Hidalgo, J. A. S., Correia, T. M., Rodrigues, A. S., & Batista, J. H. G. *Machine Learning Assisted Cement Integrity Evaluation During Plugging and Abandonment Operations: ADIPEC*, 2023. <https://doi.org/10.2118/216950-MS>
- Carter, L. G., & Evans, G. W. A Study of Cement-Pipe Bonding. *Journal of Petroleum Technology*, 16(02), 157–160, 1964. <https://doi.org/10.2118/764-PA>
- Cerasi, P., Ghaderi, A., Taghipour, A., Todorovic, J., Gawel, K., Coquard, R., Edvardsen, L., Agofack, N., & Bhuiyan, M. H. Testing methods to assess improved well cementing and remediation. *Geothermal Rising Annual Meeting*, 2020. <https://api.nva.unit.no/publication/0198cc4e3b93-019b3066-9e10-4338-87ae-f98d6f21212e>
- Ghabezloo, S., Sulem, J., Guédon, S., Martineau, F., & Saint-Marc, J. Poromechanical behaviour of hardened cement paste under isotropic loading. *Cement and Concrete Research*, 38(12), 1424–1437, 2008. <https://doi.org/https://doi.org/10.1016/j.cemconres.2008.06.007>
- James Abraham, J., & Amani, M. *Field Evaluation and Assessment of Wellbore, Casing and Cement Integrity During Critical High-Pressure Operations in Existing and Abandoned Wells: Gas&Oil Technology Showcase and Conference*, 2023. <https://doi.org/10.2118/214035-MS>
- James Abraham, J., Devers, C., Teodoriu, C., & Amani, M. *A Comprehensive Evaluation of Properties of Common and Regional Cement Formulations from a Well Integrity Standpoint: Offshore Technology Conference*, 2023. <https://doi.org/10.4043/32419-MS>
- Kim, J., Moridis, G. J., & Martinez, E. R. Investigation of possible wellbore cement failures during hydraulic fracturing operations. *Journal of Petroleum Science and Engineering*, 139, 254–263, 2016. <https://doi.org/https://doi.org/10.1016/j.petrol.2016.01.035>
- Nyakilla, E. E., Jun, G., Kasimu, N. A., Robert, E. F., Innocent, N., Mohamedy, T., Shaame, M., Ngata, M. R., & Mabeyo, P. E. Application of machine learning in the prediction of compressive, and shear bond strengths from the experimental data in oil well cement at 80 °C. Ensemble trees boosting approach. *Construction and Building Materials*, 317, 125778, 2022. <https://doi.org/https://doi.org/10.1016/j.conbuildmat.2021.125778>
- Nyakilla, E., Gu, J., Charles, G., Ricky, R., Hussain, W., Iqbal, S., Charles Kalibwami, D., Alareqi, Ahmed. G., Shaame, M., & Ngata, M. Application of Novel Machine Learning in the Prediction of Compressive Strength of Oil Well Cement with Reinforced Nanomaterial. *SSRN Electronic Journal*, 2023. <https://doi.org/10.2139/ssrn.4296915>
- Saleh, F. K., Rivera, R., Salehi, S., Teodoriu, C., & Ghalambor, A. *How Does Mixing Water Quality Affect Cement Properties: SPE International Conference and Exhibition on Formation Damage Control*, 2018. <https://doi.org/10.2118/189505-MS>
- Saunders, C. D., & Walker, W. A. *Strength Of Oil Well Cements And Additives Under High Temperature Well Conditions: Fall Meeting of the Petroleum Branch of AIME*, 1954. <https://doi.org/10.2118/390-G>
- Teodoriu, C., Asamba, P., & Ichim, A. *Well Integrity Estimation of Salt Cements with Application to Long Term Underground Storage Systems: SPE Europec featured at 78th EAGE Conference and Exhibition*, 2016. <https://doi.org/10.2118/180073-MS>
- Teodoriu, C., Kosinowski, C., Amani, M., Schubert, J., & Shadravan, A. Wellbore integrity and cement failure at HPHT conditions. *International Journal of Engineering and Applied Sciences*, 2, 2305–8269, 2013.
- Walter, M., Criado, M., Jorge, R., & Bianchi, G. *Design of High Toughness Cement for Effective Long Lasting Well Isolations*, 2003. <https://doi.org/10.2118/81001-MS>

Biographies

Dr. Jocin James Abraham is a post-doctoral researcher at Texas A&M University at Qatar. His research topics and areas of interest include well integrity, drilling and completions, underground containment, Carbon Dioxide and Hydrogen storage and reservoir characterization. Prior to pursuing a PhD, Dr. Jocin was working in the industry as an Operations Consultant for Al Badi Petroleum Services based in Doha, Qatar. Dr. Jocin has a master's degree from the University of Oklahoma and an Undergraduate Degree from Texas A&M, both in Petroleum Engineering. Dr. Jocin is very active in research and has several publications in journals and in conference proceedings.

Mona Amani is currently working as a Research Assistant at Texas A&M University at Qatar. She has a bachelor's degree in electrical and computer engineering from Texas A&M University at Qatar as well as a master's degree in industrial engineering from Texas A&M University in College Station, Texas, USA.

Dr. Mahmood Amani is a Professor of Petroleum Engineering in Texas A&M University at Qatar. Dr. Amani has a B.S. Degree in Mechanical Engineering from Wichita State University, an M.S. in Natural Gas Engineering from Texas A&M University-Kingsville, and a Ph.D. in Petroleum Engineering from Texas A&M University. After completing his Ph.D., Dr. Amani worked for few years in the Exploration and Production Technology Department of Texaco in Houston, Texas, before moving into Academia. Dr. Amani joined the faculty of the Petroleum Engineering Department of Texas A&M University in 2003 and relocated to Qatar in 2005 to start the new Petroleum Engineering Program of Texas A&M University in Qatar. Dr. Amani is very active in teaching and research and has published more than 220 articles in various journals and conference proceedings and holds two US Patents.

Dr. Catalin Teodoriu is a professor of Petroleum Engineering at the University of Oklahoma, where he holds the Mewbourne Chair #6. With over 25 years of experience across industry and academia, he previously served as an assistant professor at Texas A&M and headed the Drilling, Completion, and Workover Department at Clausthal University of Technology, Germany. An expert in drilling technology, well integrity, and failure analysis, Dr. Teodoriu specializes in developing new threaded connections for OCTG and complex mechatronic systems for AI and machine learning applications. His recent research includes leading the world's largest drilling vibration setup and establishing a long-term cement repository. He has authored over 320 publications, including 85 peer-reviewed articles, and received the prestigious SPE Drilling Engineer Award. Dr. Teodoriu is also widely recognized for delivering customized industry courses on drilling, completion, and geothermal engineering.