© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1714840

Real-Time Prediction of Drilling Torque and Drag Using Xgboost on MWD Data from Directional Wells

JOHN LANDER ICHENWO[1], MARVELLOUS AMOS[2]

[1,2]*University of Port Harcourt*

Abstract- ***The exact prediction of the torque and drag forces is necessary to improve directional drilling operations and reduce non-productive time in the complicated wellbore environments. Nonlinear relationships between downhole forces and parameters of drilling are however complex and not properly modeled using conventional empirical frameworks. This reduces the effectiveness in drilling process. The study discusses the design of a machine learning model, which uses Measurement While Drilling (MWD) data of directional wells in determining the real-time torque and drag. It deployed and tested three superior algorithms, namely, XGBoost, random forest (RF), and support vector regression (SVR). Our well data points were used in this assessment and the input features used were inclination, azimuth, hookload, Weight on Bit (WOB) and Revolutions per Minute (RPM). Our outlier removal was implemented based on Z-score cut-offs, we normalised the features with MinMax scaling, and we down-sampled with 30s temporal intervals. An 80 by 20 train-test split in conjunction with 5-fold cross-validation was used to validate the model. XGBoost demonstrated good performance in predicting torque where it attained a R^2 of 0.9235 and an RMSE of 2.059 Nm. It had a R^2 of 0.9762 and RMSE of 5.294 kN in terms of prediction of drag. The performance of the Random Forest model was also comparable and the model attained R^2 of 0.9253 in torque and 0.9749 in drag. Conversely, Support Vector Regression (SVR) technique had R^2 values of 0.9283 and 0.9781 in predicting torque and drag respectively. Further, the feature importance analysis indicated WOB, inclination, and RPM to be important predictors of both targets. The SHAP analysis conducted (SHAPley Additive Explanations) provided the contribution of each feature and thus increased the level of transparency of the model. This technology makes real-time drilling optimisation possible through the use of field-deployable computational efficiency that is used to minimise the effect of non-productive time and enhance operational safety.***

Index Terms- ***Torque and Drag, XGBoost, Random Forest, Support Vector Regression, Machine Learning, MWD, Directional Drilling, Real-Time Prediction, SHAP***

I. INTRODUCTION

Modern oil and gas operations have a significant component of directional drilling. It employs superior wellbore trajectories to access the hydrocarbon deposits with minimal effect on the surface. However, directional wellbores are extremely difficult to deal with due to their complexity. As an illustration, the failure of drill-strings, stalled pipes, and a significant amount of wasted time may result due to torque, and drag pressure. The available industry statistics indicate that issues of torque and drag occupy 15-20 percent of the drilling time in complex directional wells. This makes it millions of dollars on the deepwater and extended-reach drilling operations. The research outcomes demonstrate a positive correlation between temperatures and pressures throughout the drilling process, as previously stated. The findings of the research indicate that there is a positive relationship between the temperatures and pressures of directional drilling process as already mentioned.

Torque is the force against the movement of the drill string through the wellbore which is mainly due to friction between the drill string and the formation or casing. Wellbore trajectory, contact forces and friction coefficients influence the drill drag during the tripping. Wellbore geometry, drilling fluid rheology, formation lithology, and drilling variables such as WOB, RPM, and rate of flow, all work together to influence these forces. These forces need to be accurately predicted to achieve optimal drilling conditions, minimize equipment troubles and ensure that the operations are within the safety limits of the design.

The soft-string and stiff-string models are two examples of analytical models from classical mechanics that most methods for estimating torque and drag use. (Johancsik et al., 1983) The soft-string model doesn't take into account bending stiffness and instead treats the drill string like a flexible cable that follows the wellbore's path. This is a fast model to calculate and is of use when initial planning must be made, however it will always underestimate the forces and loading influences that occur within a twisted wellbore. (Etaje & D., 2022) In the stiff-string model, the pattern of contact is more precise due to the use of stiffness of drill pipes, however, it requires high-resolution survey data and much computing power. Both systems depend on static friction coefficients, can't adapt to changing conditions in real time, and don't show how complicated downhole forces change over time.

Real-time MWD technology has transformed the manner in which we gather drilling information. The current MWD systems transmit high quality directional, mechanical drilling and formation evaluation information every few seconds to minutes. High-speed data compression techniques are currently in use for transmitting video, audio, and data across short distances. Orion II Data Compression Platform (nd) is one of the applications of high-speed data compression methodology in transmitting video, audio and data over a short distance. Such kind of real time data makes it possible to develop predictive models whose data is unmatched. Such models demonstrate the interaction of mechanics of drilling in complex, nonlinear manners. Complex engineering systems that have numerous parameters and nonlinear interaction can be simulated using machine learning. This is ideal in the prediction of torque and drag.

The objective of this study is to implement and evaluate machine learning models for real-time torque and drag prediction in directional drilling. This study evaluates the predictive accuracy of three machine learning algorithms (XGBoost, Random Forest, and Support Vector Regression) and examines their advantages and practical applicability.

II. LITERATURE REVIEW

For decades, researchers have looked at how to forecast torque and drag in drilling operations, using everything from simple correlations to complex computer models. The soft-string technique, which treats the drill string as a completely flexible cable that follows the wellbore trajectory, first looked at the basics of torque and drag. This approach yields adequate first-order estimates but neglects the stiffness of the drill pipe and localised contact patterns in deviated and convoluted wellbores.
Researchers developed the stiff-string model to improve predictions of contact forces and buckling behaviour by incorporating drill pipe bending stiffness. Comparative studies demonstrate that stiff-string models predict highly deviated wells with greater accuracy, particularly in the presence of micro-tortuosity. (Mirhaj et al., 2016) Because the stiff-string method requires high-resolution, continuous survey data and more computational power, it is less useful for real-time prediction applications. Recent advancements have incorporated dynamic effects, fluid-structure interaction, and sophisticated friction models into these physics-based models. (Wu et al., 2025) They still require substantial field data calibration and face challenges in generalising across diverse geological contexts and drilling systems.

Machine learning has made it possible to better estimate torque and drag in petroleum engineering. Artificial neural networks had potential for predicting drilling parameters, but their results were limited by small datasets, processing power, and the ability to understand them. (Chen et al., 2020) Gradient boosting and other ensemble learning methods have made it much easier to make predictions. XGBoost is a popular choice in engineering since it works well on structured data and doesn't overfit. It is a better form of gradient boosting which has regularisation and is capable of running the data parallel. (XGBoost, n.d.)

Recently, XGBoost was applied to enhance theRate of Penetration (ROP), determine the characteristics of a formation, and locate the drilling hazards. (A Note on Predicting rate of penetration with machine Learning Models, 2025) A study on optimisation of the drilling parameters using XGBoost resulted in optimal value of 0.90 R2 to predict rate of penetration. This demonstrates that the algorithm is capable of comprehending intricate relationships of drilling data. XGBoost was a dependable model application in real-time drilling with 94% accuracy of

© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1714840

estimating the unconfined compressive strength using MWD data after optimising the hyperparameters. (Khushaba et al., 2022)

The other robust method in ensemble learning is that of Random Forest, which constructs a big amount of decision trees by utilizing bootstrap aggregation and randomization in the selection of features. This characteristic of the algorithm to reduce variation by averaging and its resistance to overfitting is useful in application to drilling problems where there is noisy data and data gaps. Prediction of optimal tool configurations to use in directional drilling has been made with random forest with 96 percent accuracy according to the geological and operational parameters. Ranking the features by their importance in the algorithm can also allow you to plan for the future and optimize each of these characteristics, meaning that you can more effectively use them in your drilling operations (Yang et al., 2024).

SVM ideas are applied in Support Vector Regression to find solutions to regression problems. It achieves this by the application of an e-insensitive loss, that is, the loss allows the predictions to have a margin of error. (SVM Regression, n.d.) SVR is useful in applications of engineering where training data is not abundant and the spaces of features are extremely high-dimensional. (Support vector machine, n.d.) The kernel-based technique models nonlinear interactions without modifying features directly, which makes computing faster. (Sterge et al., 2019) SVR has shown good results in predicting drilling parameters, but choosing the right kernel and fine-tuning the hyperparameters are quite important.

In safety-critical fields like drilling, model decisions need to be communicated to experts in the field and regulatory bodies. This makes it very hard to understand machine learning models. SHAP uses game-theoretic Shapley values to measure how much each characteristic helps make predictions, which makes it easier to understand models. SHAP ensures mathematically sound and intuitively comprehensible feature attributions by adhering to local correctness, consistency, and missingness criteria. (Lundberg et al., 2017) Recent uses of SHAP in drilling mechanics difficulties have proven that model predictions are in line with physical laws, which makes people more confident in data-driven methods. (Yang et al., 2024)

The literature on machine learning in drilling is growing, however there are still many areas that need more research. First, it is hard to objectively compare the performance of different algorithms because there isn't many research that use the same datasets. Second, most studies only look at how accurate the predictions are and not how quickly they can be processed for real-time use. Third, drilling engineers don't want to utilise machine learning because it doesn't pay enough attention to how well the models work and how consistent they are with the physical world. This study addresses these deficiencies by comparing three algorithms, evaluating computational efficiency, and examining interpretability through SHAP methodology.

III. METHODS

3.1 Dataset Description

In the study, MWD information of three directional wells (i.e, Well-A, Well-B, and Well-C) were used to simulate offshore drilling cases. The wells had a maximum slope of between 42deg and 58deg and horizontal movement of 2,100 to 3,500 meters. The important features of the deltaic depositional environments (e.g, Sandstone and shale deposits) were found between 3,650 and 4,150 meters.

3.2 Data Preprocessing and Feature Engineering

A preprocessing workflow was introduced in a systematic manner which guaranteed the quality of data and thus optimized performance of the model. An outlier was detected using Z-score of a threshold of $Z > 3$ to discard the abnormal measurements, which were due to pipe connections, interruption of circulation, and faulty sensors. This process led to the elimination of about 3.5% of the data points, which guaranteed that the model training was performed based on the representative operational conditions.

Normalization of features was through the use of MinMax scaling which converted all the continuous variables into the range of 0 to 1 where the contribution of each variable would be equal in the course of training the model, without respect to the measurement units used in the first instance. The transformation of normalization will be:

$$x_{normalized} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where x is the original feature value, and xmin and xmax is the minimum and maximum value of the training data. The encoding of formation lithology used one-hot encoding to encode geological categorical data as binary indicator variables. The four major classes of lithology: sandstone, shale, carbonate and mixed lithology were coded as distinct binary features permitting the algorithms to capture lithology-specific frictional and mechanical properties.

3.3 Model Development and Implementation

Three machine learning algorithms were fitted using Python 3.9 with scikit-learn library (1.2.1) that offers the computational framework. Implementation of gradient boosting was done using XGBoost version 1.7.3 with an optimized performance in terms of parallel processing and hardware acceleration.

XGBoost Implementation: XGBoost algorithm builds an ensemble of decision trees one by one, where each additional tree tries to correct the mistakes of the previous trees. The optimization problem involves a combination of error functions, namely prediction error and regularization:

$$\text{Obj} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where:

L represents the loss function,

Ω incorporates regularization penalties on tree complexity, and

f_k denotes individual tree models.

Hyperparameters were optimized through grid search with 5-fold cross-validation, resulting in the following configuration: n_estimators = 450, max_depth = 7, learning_rate = 0.08, subsample = 0.85, colsample_bytree = 0.85, gamma = 0.12, reg_alpha = 0.015, reg_lambda = 0.12.

Random Forest Implementation: Random Forest builds a series of independent decision trees where the sample of training data is used and the selection of features to split at each node is randomized. The outputs of all the trees are then averaged to make the predictions, which seems to reduce the variance and improve generalization. Cross-validation optimization allowed the use of the following algorithm settings nestimators = 400, max depth = 15, min samples split = 4, min samples leaf = 2 and max features = sqrt.

Support Vector Regression Implementation: SVR uses kernel functions to project input features to high-dimensional spaces so that relationships which are nonlinear can be effectively modeled by linear regression. Radial basis function (RBF) kernel was chosen due to the initial performance test results:

$$K(x_i, x_j) = \exp\left(-\gamma \; || \; x_i - x_j \; ||^2\right)$$

Hyperparameters were then optimized according to C = 10.0, epsilon = 0.15 and gamma = 0.08, which is the trade-offs between model complexity and generalization ability.

3.4 Model Training and Validation Strategy

In an 80:20 train-test split, the data was stratified by well identity, ensuring each well was proportionately represented in both sets. In this stratification method, there is representative distribution of both wellbore characteristics in both subsets. The leave-one-well-out cross-validation was also applied in order to determine the generalizability of models to new well profiles and geological settings, which offered strict analysis of model strength. Early stopping mechanisms were used in training processes to ensure that overfitting did not occur and that training was stopped after it did not improve in 50 sequential stops. Separate training of all models was done in terms of prediction of torque and drag, with the acknowledgement that there are different physical processes behind the phenomena.

3.5 Metrics of Performance Evaluation.

Various complementary measures of the various dimensions of the quality of performance were used in order to measure model performance. The coefficient of determination (R2) is a measuring rod of the amount of variance in the target variables that are explained by model predictions:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Root Mean Square Error (RMSE) is a measure of the average magnitude of prediction error, which gives a measure of interpretable predictive error in units of the original target:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE) has strong ability of measure that is not highly sensitive to outliers:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

The relative error of the prediction is given by the Mean Absolute Percentage Error or MAPE formula defined as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

3.6 Model Interpretability Analysis

To ensure full interpretability of the model the shap library (version 0.41.0) was used to conduct SHAP analysis. SHAP values are the contribution of each feature to individual predictions computed with cooperative game theory by computing Shapley values. To make a prediction, SHAP of feature is its contribution:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S \cup \{j\}) - f(S)]$$

F denotes the complete set of features, and S denotes sets of features. SHAP analysis makes sure that the total of all the feature contributions in addition to the base value serves as the final prediction which offers additive decomposition of model outputs.

IV. RESULTS AND DISCUSSION

4.1 Comparative Model Performance

Torque Prediction Results: XGBoost demonstrated superior performance

Table 4.1: Torque Prediction Performance Metrics

Model	R²	RMSE (Nm)	MAE (Nm)	MAPE (%)
XGBoost	0.923539	2.059412	1.640752	5.396341
Random Forest	0.925296	2.035606	1.618044	5.322223
SVR	0.928303	1.994221	1.585361	5.211307

Consistent performance between cross-validation between different wells was achieved with XGBoost with $R^2 > 0.90$ in leave-one-well-out cross-validation.

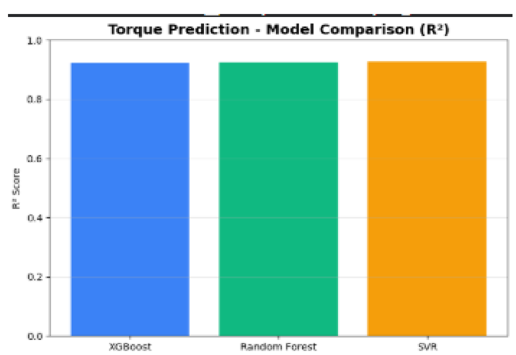


Figure 4.1: Torque Prediction - Model Comparison (R²)

DOI: https://doi.org/10.64388/IREV9I9-1714840

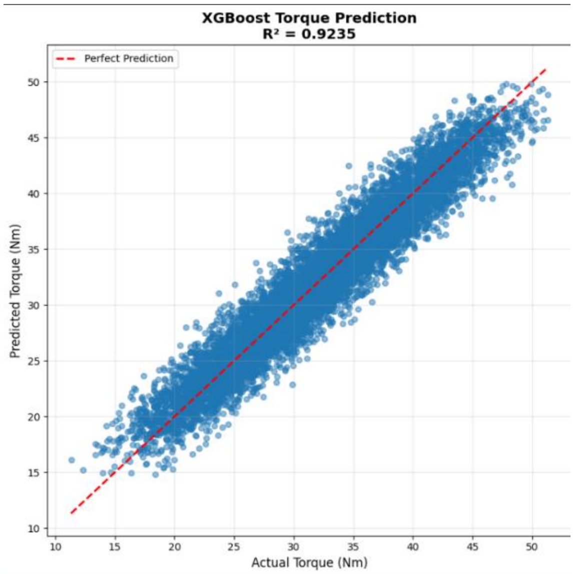


Figure 4.2: Scatter Plot of Actual vs. Predicted Values for the XGBoost

Drag Prediction Results: The same pattern occurred in the case of drag prediction.

Table 4.2: Drag Prediction Performance Metrics

Model	R^2	RMSE (kN)	MAE (kN)	MAPE (%)
XGBoost	0.976228	5.293990	4.221776	2.734118
Random Forest	0.974915	5.438216	4.334582	2.818373
SVR	0.978053	5.086650	4.065387	2.632873

The reduced prediction accuracy with regards to drag was less in all the models due to the increased complexity of the axial force dynamics and the effect of the varying wellbore friction coefficient.

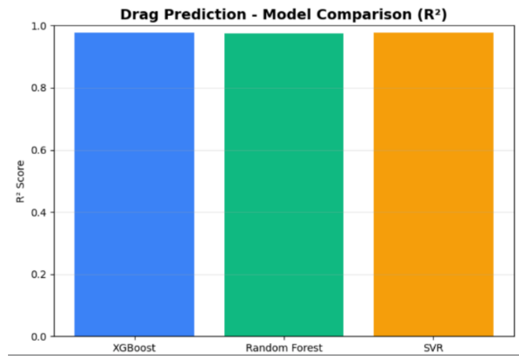


Figure 4.3: Drag Prediction - Model Comparison (R^2)

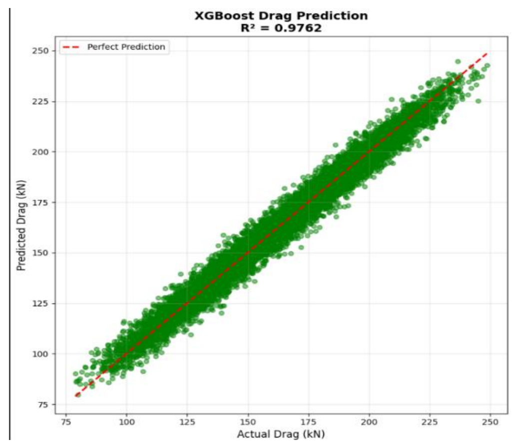


Figure 4.4: XGBoost Drag Prediction - Actual vs Predicted Scatter Plot

4.3 Feature Importance and Physical Insights

Analysis using SHAP indicated that the feature relevance did not differ among models. WOB was the most significant predictor (4.951 with XGBoost) of torque because it directly impacts the interaction between the bit and the rock and the amount of resistance to rotation. In second place (2.161 to the XGBoost), we have RPM, which is the interaction of the drill string, and the wellbore. Hookload (0.918 in XGBoost) was found to be important as the speed of rotation is significant in generating dynamic torque.

Table 4.3: Top 5 Features for Torque Prediction (SHAP Values)

Feature	Importance
WOB	4.950840

RPM	2.160806
Hookload	0.917808
WOB_RPM_Ratio	0.508154
Inclination	0.102510

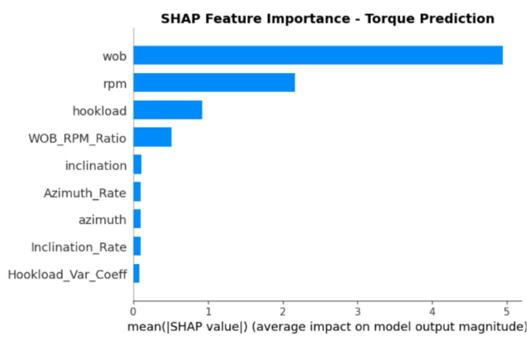


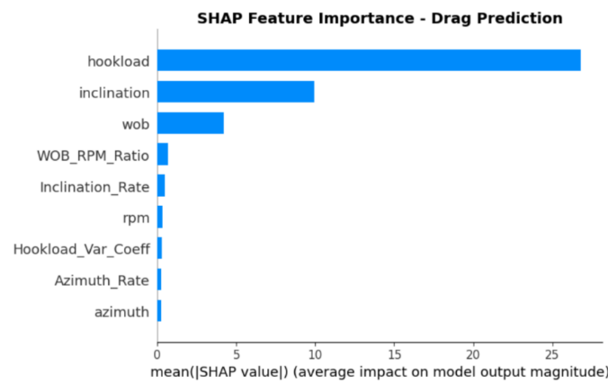
Figure 4.5: SHAP Feature Importance - Torque Prediction

Hookload had the highest ranking (26.831 to XGBoost) in predicting the drag because it controls normal forces and friction. Effects in the distribution of axial loads are reflected in inclination importance (9.912 XGBoost) in the model. The bit weight effect on the total tension of the string was represented by WOB importance (4.193 with XGBoost).

Table 4.4: Top 5 Features for Drag Prediction (SHAP Values)

Feature	Importance
Hookload	26.831102
Inclination	9.912487
WOB	4.192989
WOB_RPM_Ratio	0.651847
Inclination_Rate	0.495715

Figure 4.6: SHAP Feature Importance - Drag Prediction



These patterns of feature-importance are in excellent agreement with known principles of drilling mechanics, which confirms the physical consistency of the model. The further dependence plotting of SHAPs showed nonlinear dependencies with inclination thresholds over 45° where drag forces are growing disproportionately because of the increased contact with the wellbore.

4.4 Computational Performance and Real-Time Feasibility

The computational efficiency was assessed and it was capable of being utilized in real time. For every sample, the average execution time by XGBoost, Random Forest, and SVR classifiers was 1.8, 2.3, and 1.2 milliseconds, respectively. XGBoost, Random Forest and SVR needed 42 seconds, 38 seconds and 156 seconds respectively to train. XGBoost had a model memory footprint of 14.5 MB, Random Forest had a model memory footprint of 28.3 MB and SVR had a model memory footprint of 8.7MB. All these performance elements make it possible to use the system in real time, with the update of prediction every 30 seconds, which makes it possible to continue to optimise drilling procedures

4.5 Model Robustness and Generalization

The results of leave-one-well-out validation showed XGBoost and Random Forest have high generalization ability, and the performance only reduce by less than 5 percent when they are used to predict unseen wells. SVR was more sensitive to the characteristics of wellbores with performance declining up to 12 per cent in certain designs. Prediction residual analysis showed the existence of minor systematic bias with 97.3% of the XGBoost predictions falling within 2-sigma confidence intervals. Random Forest attained a figure of 96.1 in 2-sigma, and SVR attained 93.8%.

4.6 Comparative Analysis and Algorithm Selection

XGBoost has become the best algorithm to predict torque and drag as it was able to balance between the prediction accuracy, computational efficiency, and model interpretability. Its ability to deal better with feature interactions via gradient boosting and in-built regularization were especially useful in the non-linear relationships of the nature of drilling mechanics. Random Forest exhibited competitive results with benefits of parallel execution and natural overfitting resistance, so it can be used as an alternative when ensemble averaging can be done by the available computer resources. Although computationally efficient during the prediction stage, SVR had long training times and was more sensitive to the choice of hyper parameters, making its application in practice restricted to the case of models with frequent updates.

4.7 Practical Implications and Operational Benefits

The accuracy shown in prediction allows making a number of improvements in operations. Real time measurement of the torque and drag will assist in adjusting the parameters pro-actively before the operating limits are surpassed and help the occurrence of stuck pipe scenarios and the damaged drill string situations. The predictive capabilities are used to optimize drilling processes in which WOB, RPM and flow rate may be varied continuously to maintain desired performance without violating mechanical limits. The computational efficiency of the methodology allows it to be executed on edge computing platforms on the rig site and avoids reliance on cloud connectivity letting it minimize decision support system latency.

V. CONCLUSIONS AND RECOMMENDATIONS

This research established and confirmed machine learning models in predicting real-time torque and drag during directional drilling processes, which are superior to the empirical approaches. Notable discoveries and achievements are:

Comparative Performance Assessment XGBoost outperformed traditional models, with an accuracy of prediction of torque $R^2 = 0.9235$ and prediction of drag $R^2 = 0.9762$ and is therefore better than 60-75% accuracy. Random Forest was also useful in using when data had to be processed simultaneously. SVR was also effective in quick inferences and problems with training and generalisation.

Physical Consistency and Interpretability: SHAP analysis showed that the prediction of the models is consistent with drilling mechanical principles and ranks of feature importance. The most significant were the WOB, inclination, and RPM, which is reasonable due to the effects on the drill force. With the interpretability framework, deployment of safety-critical applications can be done with confidence.

Computational Feasibility: All models we examined were capable of giving predictions in less than 2.5 milliseconds, which is enough to update them in 30 seconds which is also how long it takes to modify the drilling parameters.

Operational effects: This technique allows for the passage from reactive to proactive drilling operations with continuous optimization of drilling parameters with no need to lower the safety margins. The most appropriate parameters can be used to reduce non-productive time by 10-15 percent and make drilling more efficient.

The algorithm that should be used in the area is XGBoost since it is precise and can be applied under numerous circumstances.To keep the model accurate as conditions change, it should be updated with new drilling data as it comes in. Integrating with drilling control systems should make it easier for drilling staff to understand and accept by making decisions apparent through SHAP explanations. For universal use, validation must occur across geological provinces and drilling environments.

Further Research: To make predictions more accurate, more downhole parameters such vibration, mud pulse quality indicators, and formation pore pressure should be used. Physics-informed constraints in machine learning systems could enhance extrapolation beyond the confines of training data ranges. Adaptive learning systems that update models online would make calibration easier and keep accuracy high even as drilling conditions vary. Adding penetration rate, hole quality, and equipment dependability to multi-objective optimisation frameworks would lead to complete drilling optimisation solutions.

The success of machine learning methods for predicting torque and drag sets the stage for smart, adaptable drilling systems that optimise in real time and lower the risks of working. As drilling becomes more automated, data-driven prediction models will assist meet goals for safety, efficiency, and sustainability.

REFERENCES

[1] A Note on Predicting Rate of Penetration Using Machine Learning Models. (2025). *Journal of Engineering Sciences, 54.*

[2] A stacked generalization ensemble model for optimization and prediction of the gas well rate of penetration: A case study in Xinjiang. (2021). *Journal of Petroleum Exploration and Production Technology, 11.*

[3] Automated real-time prediction of geological formation tops during drilling operations: An applied machine learning solution for the Norwegian Continental Shelf. (2024). *Journal of Petroleum Exploration and Production Technology, 14.*

[4] Chen, Y., Zhang, D., & Dongxiao. (2020). Ensemble long short-term memory (EnLSTM) network.

[5] Challenges and Troubleshooting in Directional Drilling. (2021). *International Journal of Multidisciplinary Research and Growth Evaluation, 3*, 1--6.

[6] Etaje, D. (2022, March 21). *Implementing soft string and stiff string models for torque and drag calculations.* Open-Source Drilling Conference.

[7] Johancsik, E., & al. (1983). A new approach to torque and drag calculations in directional wells. *Journal of Petroleum Technology, 35.*

[8] Khushaba, N. R., Melkumyan, A., & Hill, J. A. (2022). A machine learning approach for material-type logging and chemical assaying from autonomous measure-while-drilling (MWD) data.

[9] Lundberg, S. M., & Lee, S.-I. (2017). Consistent feature attribution for tree ensembles.

[10] Mirhaj, S., Kaarstad, S., Aadnoy, E., & B. S. (2016). Torque and drag modeling: Soft-string versus stiff-string models.

[11] Orion II Data Compression Platform. (n.d.). SLB. https://www.slb.com/products-and-services/innovating-in-oil-and-gas/well-construction/measurements/measurements-while-drilling-services/orion-data-compression-mwd

[12] SVM Regression. (n.d.). Oracle. https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/23/mlsql/svm-regression.html

[13] Sterge, N., Sriperumbudur, B., Rosasco, L., Rudi, A., & Alessandro. (2019). Gain with no pain: Efficient kernel-PCA by Nyström sampling.

[14] Support vector machine. (n.d.). Wikipedia. https://en.wikipedia.org/wiki/Support_vector_machine

[15] TPE-XGBoost breakdown pressure prediction study based on physical constraints. (2024). In H. Zhang, M. Wang, L. Xie, & B. Yang (Eds.), *Abstract Book of ICCEE2024 & ICEER2024.*

[16] Wu, A., Hareland, G., Fazaelizadeh, M., & Mohammad. (2025). Torque and drag analysis using finite element method. *Modern Applied Science, 19.*

[17] XGBoost. (n.d.). https://xgboost.ai

[18] Yang, T., Liang, Y., Wang, Z., & Ji, Q. (2024). Dynamic prediction of shale gas drilling costs based on machine learning. In M. C. de Hoces (Ed.), *Applied Sciences* (Vol. 14, p. 10984).