processes

MDPI

Article

Construction of an Extended-Reach Well Torque Prediction Model Based on Multi-Source Drilling Data Fusion and Analysis of Dominant Factors

Junrui Ge [1], Pengbo Li [2,*], Wei Liu [1], Bin Cai [1], Yanfei Li [1], Xuyue Chen [2], Xiujin Yuan [1] and Penglei Tang [1]

[1] CNOOC (China) Limited Shanghai Branch, Shanghai 200335, China; gejr@cnooc.com.cn (J.G.); liuwei5@cnooc.com.cn (W.L.); caibin4@cnooc.com.cn (B.C.); liyf8@cnooc.com.cn (Y.L.); yuanxj2@cnooc.com.cn (X.Y.); tangpl2@cnooc.com.cn (P.T.)

[2] School of Petroleum Engineering, China University of Petroleum (Beijing), Beijing 102249, China; chenxuyue2011@163.com

* Correspondence: 2025310614@student.cup.edu.cn

Abstract

Extended-reach wells contain long, highly inclined intervals in which drill-string/wellbore contact and cumulative friction evolve continuously with depth, making torque prediction strongly nonstationary. This study develops a multi-source torque-prediction framework using 5977 field samples from a deepwater extended-reach well in the East China Sea. To eliminate the optimistic bias caused by randomly mixing adjacent depth samples, all records were ordered by measured depth; preprocessing and feature screening were fitted only on historical data; hyperparameters were selected with forward-chaining TimeSeriesSplit; and model performance was assessed using both a fixed deep-depth stress test and expanding-window rolling-origin prediction. Training-only correlation screening and variance inflation factor analysis reduced multicollinearity before comparing ExtraTrees, XGBoost, LightGBM, BPNN, and time-aware ensemble regression. Static extrapolation to the deepest 15% of the well revealed severe distribution shift and poor long-range generalization. In contrast, 50 m rolling-forward updating produced reliable prediction: ExtraTrees achieved an MAE of 1.121, RMSE of 1.643, MAPE of 2.750%, R^2 of 0.907, and Pearson correlation of 0.964, while LightGBM achieved the lowest MAE of 1.090. ExtraTrees R^2 decreased from 0.927 at a 25 m horizon to 0.711 at 200 m, demonstrating progressive concept drift with prediction distance. SHAP analysis identified measured depth as the dominant predictive proxy, whereas WOB, flow rate, and RPM were more relevant to operational intervention. The revised framework therefore emphasizes leakage-free validation, adaptive updating, and separation of predictive from controllable factors rather than random-split interpolation accuracy.

Keywords: offshore extended-reach well; torque prediction; rolling-forward validation; concept drift; SHAP; machine learning

Check for updates

Academic Editor: Dicho Stratiev

Received: 30 July 2026
Revised: 22 August 2026
Accepted: 25 August 2026
Published: 31 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Offshore, deepwater, and structurally complex reservoirs are increasingly developed through extended-reach and cluster wells because a single platform can access a wider reservoir area while reducing the number of well slots. The corresponding long tangent and high-inclination sections, however, increase drill-string/wellbore contact length, normal force, and cumulative friction. Excessive torque and drag can aggravate poor weight

transfer, stick-slip, fatigue, bit wear, hole-cleaning problems, and stuck-pipe risk. Torque is therefore not merely a bit-cutting response; it reflects coupled bit-rock interaction, drillstring/wall friction, fluid resistance, trajectory, formation heterogeneity, and operating state. Chen et al. [1] highlighted the importance of real-time parameter optimization in offshore cluster extended-reach drilling, while Bai et al. [2] and Cao et al. [3] demonstrated the value of data-driven torque prediction for ahead-of-bit monitoring and drilling optimization. Classical formulations such as those of Johancsik et al. [4] and Ho [5] are analytical or physics-based torque-and-drag models with clear mechanical meaning, but their accuracy can be limited when friction coefficients, cutting beds, lubricity, vibration, tool condition, and contact states vary dynamically.

Machine learning provides a complementary way to learn nonlinear relationships from recorded drilling data [6]. Previous studies have addressed drilling optimization and ROP prediction [1,7–10], formation and lithology perception [11–13], porosity and lost-circulation prediction [14,15], drilling-fluid intelligence [16,17], and torque modeling [2,3,18–20]. Explainable machine learning has also been introduced to improve credibility in drilling and offshore applications [21,22]. Nevertheless, three methodological issues remain particularly important for depth-ordered drilling data. First, random train–test splitting can place adjacent depth samples in both sets, and can therefore measure local interpolation rather than genuine forward prediction. Second, correlated trajectory and bit variables can distort feature-importance rankings. Third, drilling conditions evolve with depth, so a static model calibrated in one interval may experience substantial concept drift in a subsequent interval.

Recent work further indicates a shift from static global prediction toward adaptive and physics-guided drilling intelligence. Bai et al. [2] used just-in-time learning to predict ahead-of-bit torque under changing conditions. He et al. [23] developed a physics-informed neural-network framework for drilling ROP transfer generalization, and Song et al. [24] integrated symbolic regression with time-series perception and reinforcement learning for real-time drilling optimization. Wada and Kaneko [25] reviewed the increasing integration of physics-based and data-driven models in drilling technology. These studies suggest that temporal/depth causality, transferability, and physically meaningful interpretation should be treated as core validation issues rather than secondary considerations.

Accordingly, this study revises the conventional random-split workflow and establishes a leakage-aware torque-prediction framework using multi-source field data from a deepwater East China Sea extended-reach well. The main contributions are (1) a multi-source dataset integrating drilling, drilling-fluid, logging, trajectory, and bit/BHA information; (2) a depth-ordered validation strategy combining training-only preprocessing, TimeSeriesSplit hyperparameter selection, a fixed deep-depth stress test, and rolling-origin prediction at 25–200 m horizons; and (3) VIF and SHAP analyses that explicitly distinguish variables with high predictive value from parameters that can be directly adjusted during drilling. The objective is not to maximize random-split accuracy but to quantify realistic forward generalization and identify practical conditions under which data-driven torque prediction remains reliable.

2. Data Sources and Preprocessing Methods

2.1. Data Sources and Feature Composition

The field dataset contains 5977 consecutive depth-indexed samples from 150 to 6126 m in a deepwater extended-reach well in the East China Sea. Candidate variables were grouped into drilling, drilling-fluid, logging, trajectory, and bit/BHA categories. Torque was retained only as the prediction target.

The drilling variables include hook load (WHO), weight on bit (WOB), standpipe pressure (SSP), rotary speed (RPM), flow rate (FLWpmps), bit footage (Bit Run), pure

drilling time (Bit Time), pumping time, and rate of penetration (ROP). Drilling-fluid variables include density, viscosity, chloride concentration, and fluid type. Logging variables include gamma ray (GR), spontaneous potential (SP), acoustic travel time (DT), bulk density (Rock Den), caliper (CAL), micro-spherical resistivity (RMSL), shallow resistivity (RS), and deep resistivity (RD). Measured depth (Depth), true vertical depth (TVD), and inclination characterize the trajectory, while bit size, bit type, blade number, total flow area (TFA), and directional-tool information characterize the bit/BHA system.

Variable availability was distinguished from variable controllability. Surface drilling parameters can be obtained continuously during operation, and trajectory variables are available from survey/MWD information. Logging channels can be treated as real-time inputs only when they are acquired by LWD/MWD and depth-aligned with the drilling response. Because the archived dataset does not consistently document the acquisition mode of every logging channel, the full-information model in this study is interpreted primarily as a field diagnostic/predictive model; claims of fully online deployment are limited to variables known to be available at or before the prediction depth. The multi-source input feature system used for torque prediction is illustrated in Figure 1.

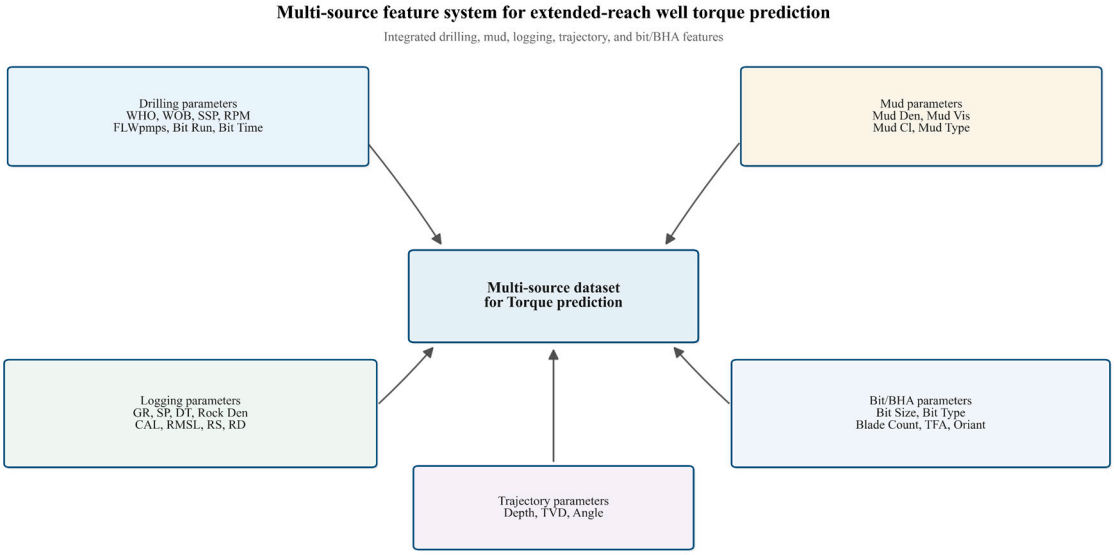

Figure 1. Multi-source input feature system used for torque prediction.

2.2. Depth-Ordered Data Cleaning and Leakage Control

All samples were first sorted by measured depth. Missing-value treatment, scaling, correlation screening, and VIF calculations were performed only with historical training data within each forward-validation stage. Forward filling was used where physically appropriate; any remaining missing numerical values were replaced using medians estimated from the corresponding training interval. The final test interval was never used to fit imputation statistics, scalers, feature-selection rules, model parameters, hyperparameters, or uncertainty calibration. This ordering eliminates future-depth information leakage that can occur when the complete well is interpolated or standardized before splitting.

Tree models do not require feature scaling, whereas BPNN inputs were scaled using parameters fitted on the relevant historical training interval. Categorical quantities that had already been numerically encoded in the archived field dataset were treated consistently across folds. Torque was never used as an input feature.

2.3. Denoising Sensitivity and Signal-Preservation Check

The originally submitted workflow applied Hampel filtering, wavelet-threshold denoising, and Savitzky–Golay smoothing [26–28] to field signals. The reviewers correctly

noted that symmetric smoothing can use future neighboring samples and that smoothing the torque target can artificially simplify the prediction task. In the revised analysis, the measured torque is therefore retained in its unsmoothed form as the ground-truth target, and filtering is treated as a sensitivity analysis rather than a mandatory preprocessing stage.

Within the historical training interval, raw inputs were compared with causal Hampel alternatives using the same TimeSeriesSplit evaluation. The raw-input configuration obtained the lowest mean ExtraTrees forward-validation RMSE (4.721). Causal Hampel variants produced RMSE values of approximately 4.786–5.015 and did not provide a consistent generalization benefit. Consequently, the main predictive results reported below use raw inputs and raw torque. A representative comparison of the raw and legacy-filtered drilling signals is shown in Figure 2. The legacy filtered traces are retained only to illustrate the magnitude of smoothing and are not used to define the revised prediction target. Spatial-frequency spectra of representative signals are provided in Figure S1. Because the archived series is indexed by measured depth rather than by uniform time, the spectral axis is reported in cycles/m rather than Hz. The diagnostic spectra show that filtering mainly attenuates higher spatial-frequency components while the dominant low-spatial-frequency energy is retained; this result is used only to assess signal preservation and does not justify smoothing the torque target.

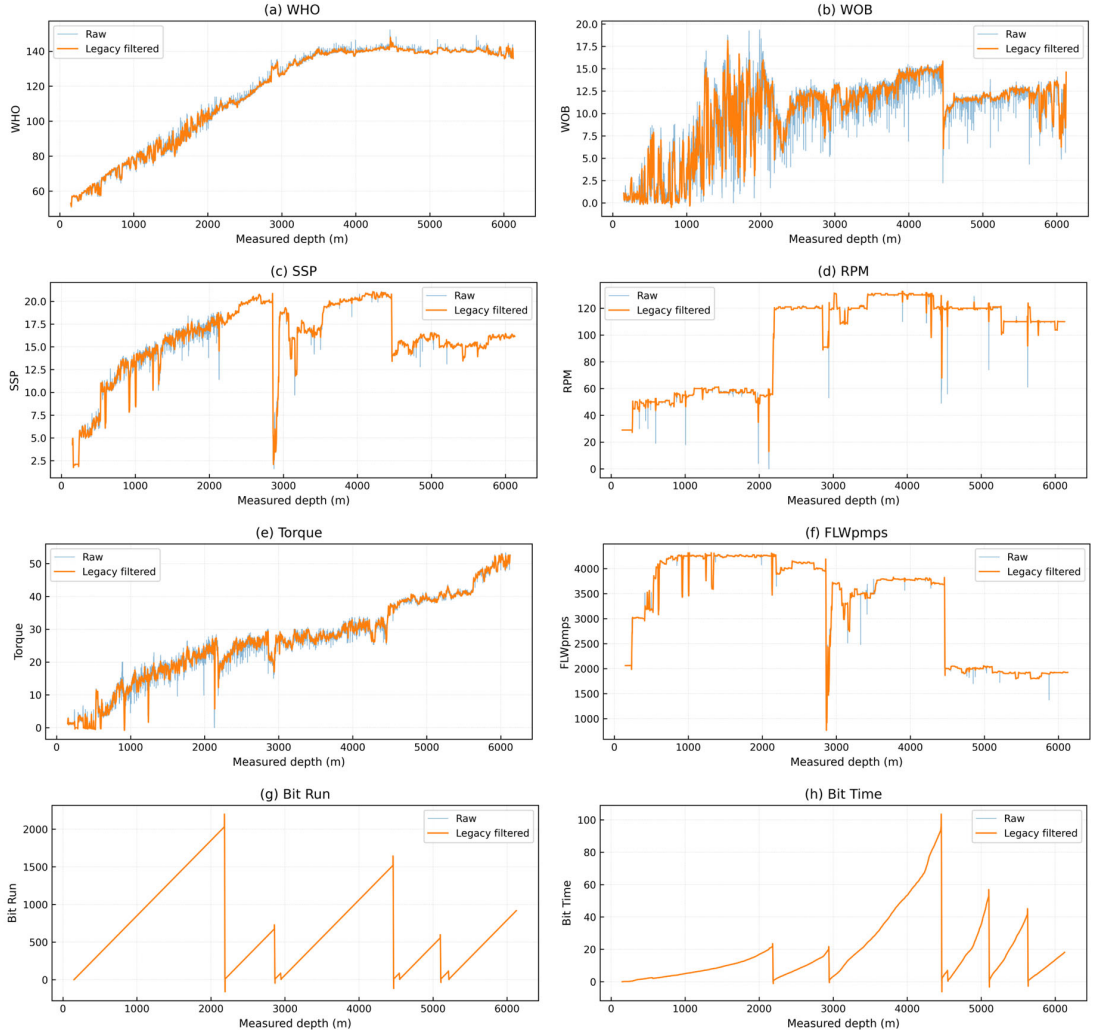

Figure 2. Representative comparison of raw and legacy-filtered drilling signals used only for denoising sensitivity diagnosis. Filtered torque is not used as the target in the revised predictive workflow.

Because the available archive is indexed at approximately 1 m depth increments rather than at uniform time intervals, frequency preservation was assessed in the spatial-frequency domain (cycles/m) rather than in Hz. A raw-versus-filtered spatial-spectrum comparison for representative signals is provided in Figure S1. This diagnostic confirms that dominant low-spatial-frequency structure is preserved, while the revised model does not rely on symmetric filtering for forward prediction.

3. Correlation and Multicollinearity Analysis

3.1. Spearman Correlation

Spearman rank correlation [29] was used to characterize monotonic relationships between torque and the multi-source variables because it is distribution-free and less sensitive to non-Gaussian outliers than Pearson correlation. The coefficient is

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}.$$

where ρ_s is the Spearman rank-correlation coefficient, d_i is the difference between the paired ranks of the ith sample, and n is the number of samples. The correlation structure is shown in Figure 3. Torque is strongly associated with measured depth and other variables that evolve with well extension, and it also responds to WOB, circulation, formation properties, and bit/BHA state.

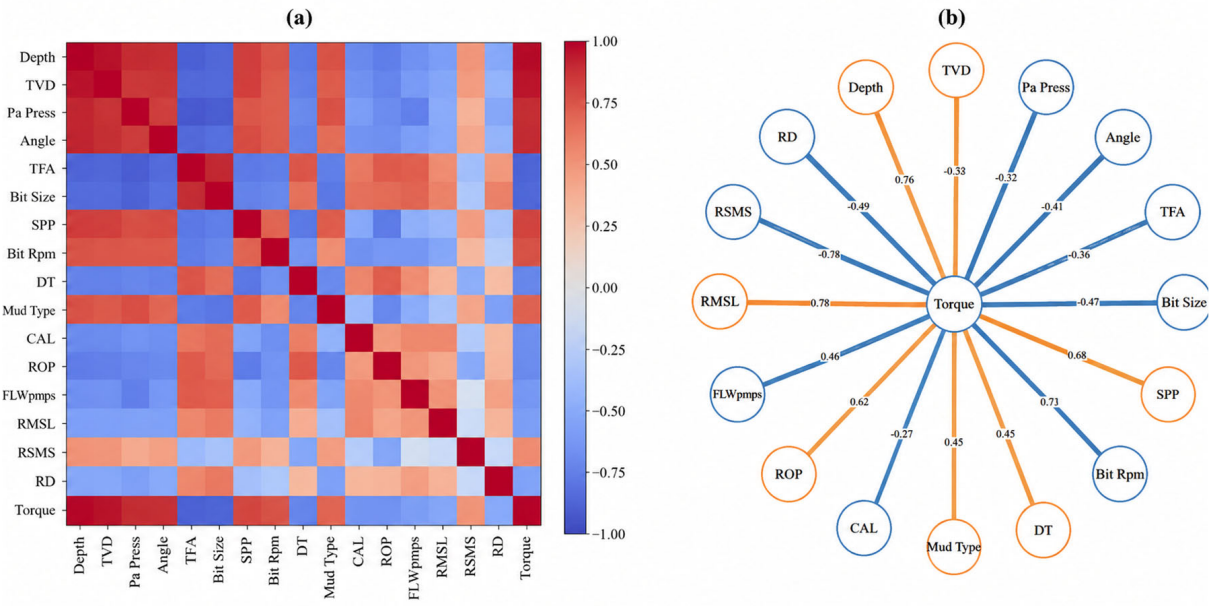

Figure 3. Correlation structure between torque and multisource drilling parameters: (**a**) Spearman correlation heatmap; (**b**) torque-centered correlation network.

3.2. Multicollinearity Screening with VIF

Correlation analysis identified severe redundancy among several variables, including Depth–TVD, Depth–WHO, Bit Time–Pump Time, and bit structure variables. To prevent redundant predictors from distorting interpretation, pairwise screening was followed by variance inflation factor (VIF) analysis using only the initial training interval. VIF was calculated as

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}.$$

where R^2_j is the coefficient of determination obtained by regressing predictor j on the remaining predictors. Variables were iteratively removed until all retained predictors

satisfied VIF < 10. The retained predictive features and their corresponding final VIF values are summarized in Table 1.

Table 1. Predictive feature set retained after training-only correlation and VIF screening.

Retained Predictor	Final VIF
Depth	5.834
WOB	3.794
CAL	2.855
ROP	2.804
Rock Den	2.598
SP	2.350
Bit Run	2.209
DT	2.157
FLWpmps	1.776
RMSL	1.513
Oriant	1.381
GR	1.343
RD	1.049
RS	1.013

The predictive feature set retained Depth, WOB, FLWpmps, Bit Run, GR, SP, DT, Rock Den, CAL, RMSL, RS, RD, Oriant, and ROP; the maximum final VIF was 5.834. Because Depth is a strong state/progression variable but is not directly controllable, a second control-oriented feature set was constructed by excluding Depth/TVD while retaining WOB, RPM, and flow rate. This auxiliary model is used only for operational interpretation and not to replace the principal predictive model.

4. Torque Prediction Model Construction and Validation

4.1. Leakage-Free Validation Design

The revised validation strategy contains two complementary levels. First, the well was divided chronologically by measured depth into an initial 70% training interval (150–4332 m; 4183 samples), a subsequent 15% validation/calibration interval (4333–5228 m; 896 samples), and a final 15% independent, deep test interval (5229–6126 m; 898 samples). Hyperparameter selection was performed only inside the 70% historical training interval using forward-chaining TimeSeriesSplit. After model selection, the first 85% of the well was refitted and the deepest 15% was retained as a strict long-range extrapolation stress test.

Second, because field drilling proceeds sequentially and new observations become available continuously, an expanding-window rolling-origin evaluation was adopted as the main engineering validation. Starting from the initial 70% historical interval, the model predicted the next unseen depth block. After measurements from that block became available, they were appended to the historical training set before predicting the subsequent block. Prediction horizons of 25, 50, 100, and 200 m were evaluated. No observation at a deeper future depth was used to generate an earlier prediction.

4.2. ExtraTrees Regression Model

ExtraTrees [30] averages a set of highly randomized regression trees and can capture nonlinear interactions in tabular drilling data while limiting variance through ensemble averaging. Compared with related randomized tree ensembles, such as Random Forests [31], ExtraTrees additionally randomizes candidate split thresholds. In the revised workflow, its hyperparameters were not fixed a priori; the number of trees, maximum depth, feature-sampling fraction, and minimum split/leaf sizes were selected using randomized search with TimeSeriesSplit and validation RMSE as the objective. The schematic structure of the XGBoost regression model is shown in Figure 4. The blue dots represent the input features, while the different colors are used to distinguish nodes at different levels of the regression trees.

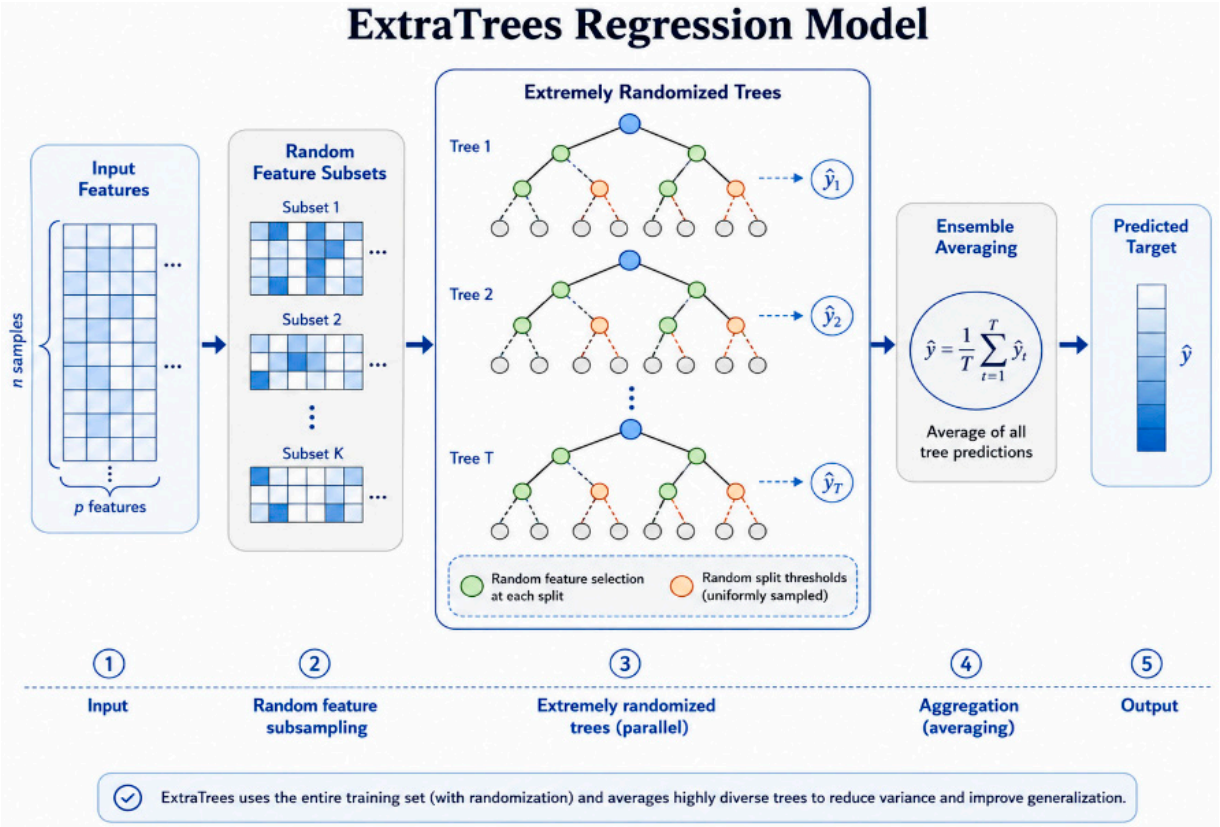

Figure 4. Schematic structure of the ExtraTrees regression model.

4.3. Extreme Gradient Boosting (XGBoost)

XGBoost [32] constructs boosted regression trees sequentially, with each tree correcting residual errors from the previous ensemble. Tree number, learning rate, depth, sample fractions, gamma, and regularization coefficients were optimized using the same forward-validation strategy so that model selection remained consistent with the ordered nature of the drilling sequence. The schematic structure of the XGBoost regression model is shown in Figure 5. The dots represent model nodes or input elements, while the different colors are used to visually distinguish different components of the model structure.

4.4. LightGBM

LightGBM [33] uses histogram-based learning and leaf-wise growth to provide efficient nonlinear regression for high-dimensional tabular data. The number of trees, learning rate, leaf number, tree depth, child-sample number, feature/bagging fractions, and regu-

larization coefficients were selected through the same TimeSeriesSplit-based search. The schematic structure of the LightGBM regression model is shown in Figure 6.

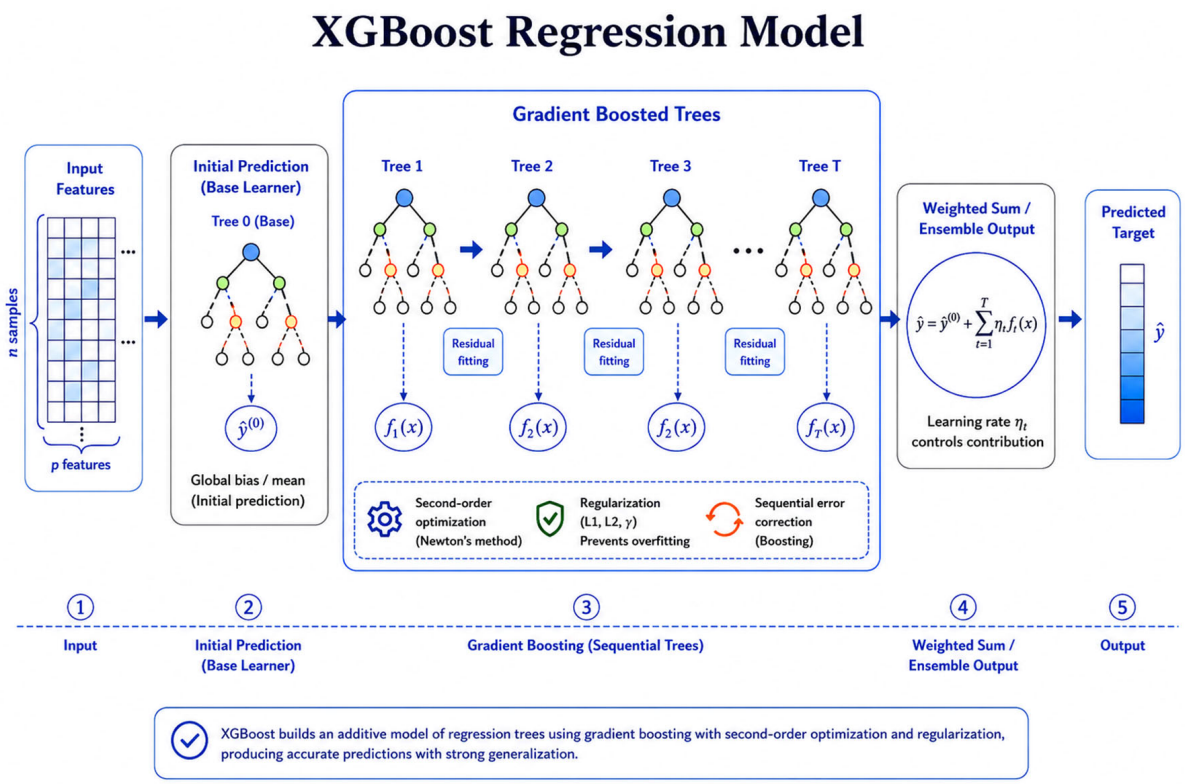

Figure 5. Schematic structure of the XGBoost regression model.

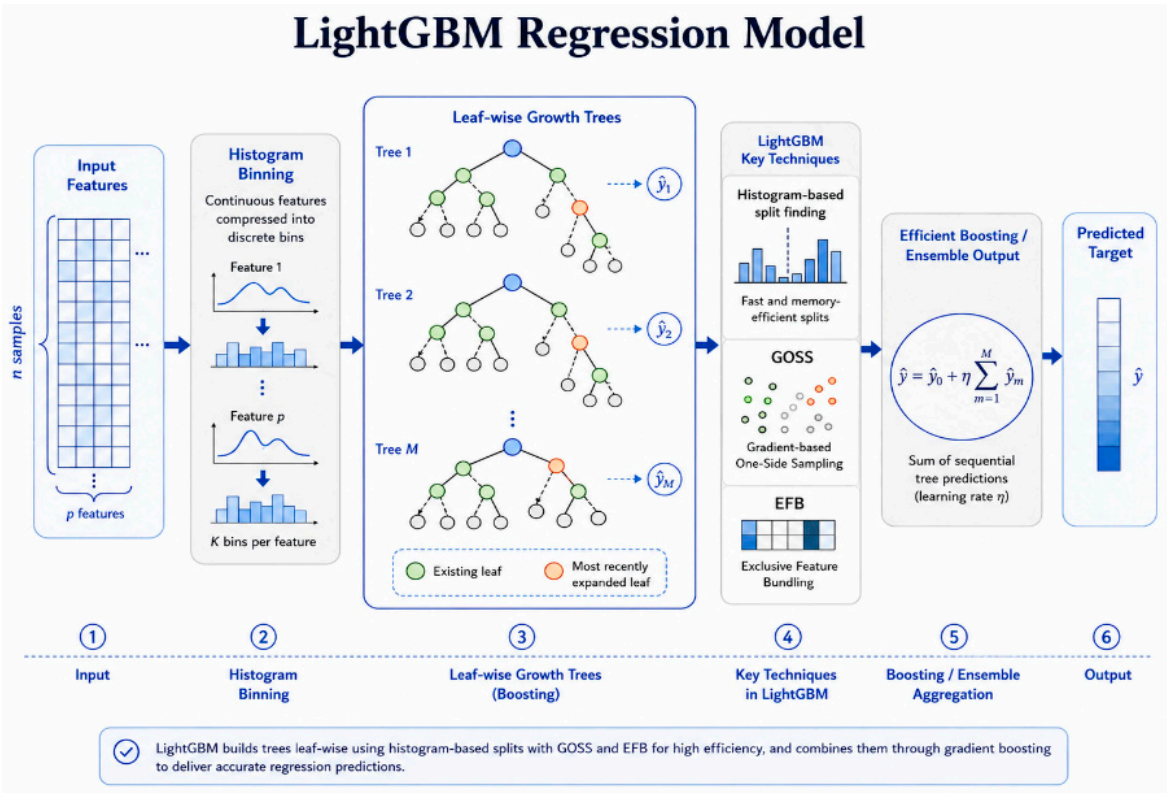

Figure 6. Schematic structure of the LightGBM regression model.

4.5. Backpropagation Neural Network (BPNN)

BPNN [34] was used as a nonlinear neural baseline. Candidate networks were trained with explicit forward validation rather than with a randomly mixed validation subset. The selected diagnostic configuration contained hidden layers of 64 and 32 neurons, ReLU activation, dropout of 0.2, Adam optimization [35], learning rate 0.001, and batch size 32. Early stopping was based on the subsequent deeper validation interval. The schematic structure of the BPNN regression model is shown in Figure 7. The blue dots represent neurons, while the different colors are used to distinguish the input, hidden, and output layers.

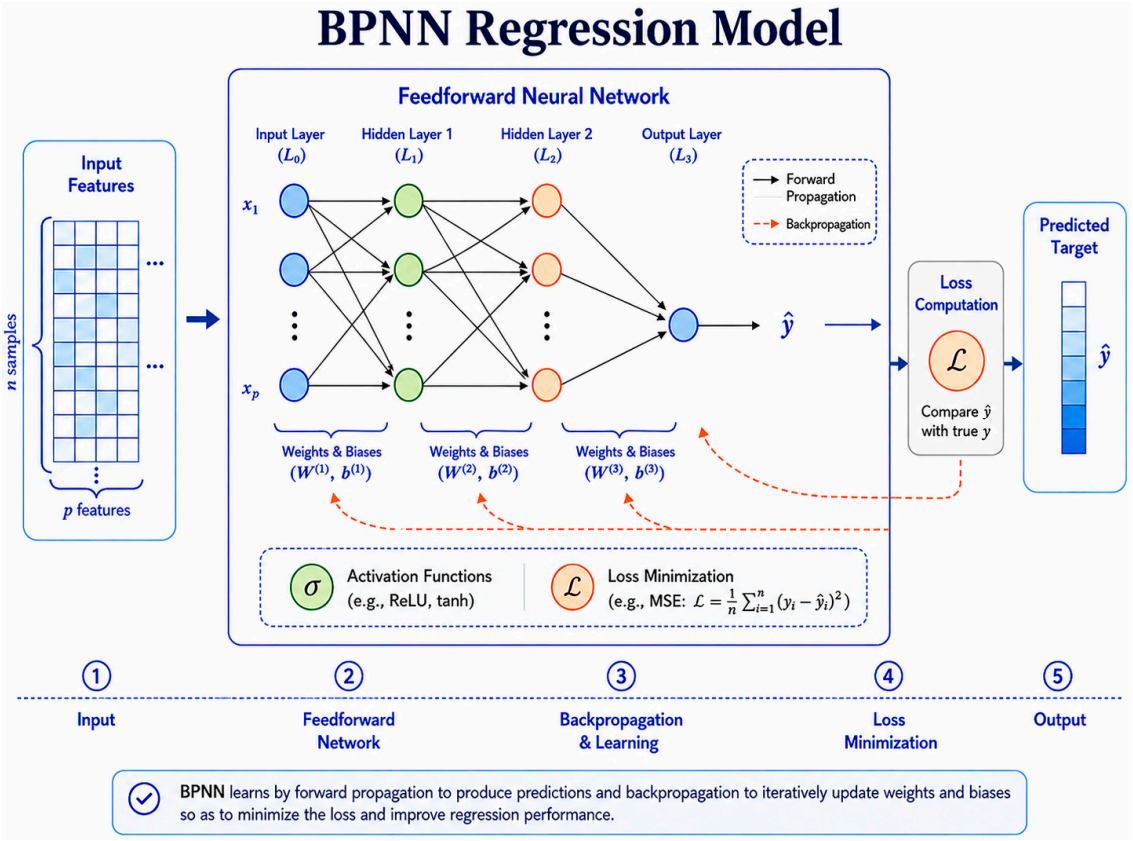

Figure 7. Schematic structure of the BPNN regression model.

4.6. Time-Aware ER Ensemble Regression

The ensemble-regression (ER) model [36] combines ExtraTrees, XGBoost, LightGBM, and BPNN predictions in a two-layer stacking architecture. The randomly shuffled five-fold OOF procedure used in the original submission has been replaced by time/depth-aware forward-chaining stacking. Each base learner is trained only on shallower historical observations and produces meta-features for a subsequent deeper validation interval. Linear regression, Ridge, LASSO, and ElasticNet were compared as meta-learners; LASSO ($\alpha = 0.001$) produced the lowest forward-validation RMSE in the revised diagnostic run and was therefore used for the strict holdout ER result. The structure of the time-aware ER ensemble-regression model is illustrated in Figure 8. The dots represent model nodes, while the different colors are used to visually distinguish different components and stages of the ensemble architecture.

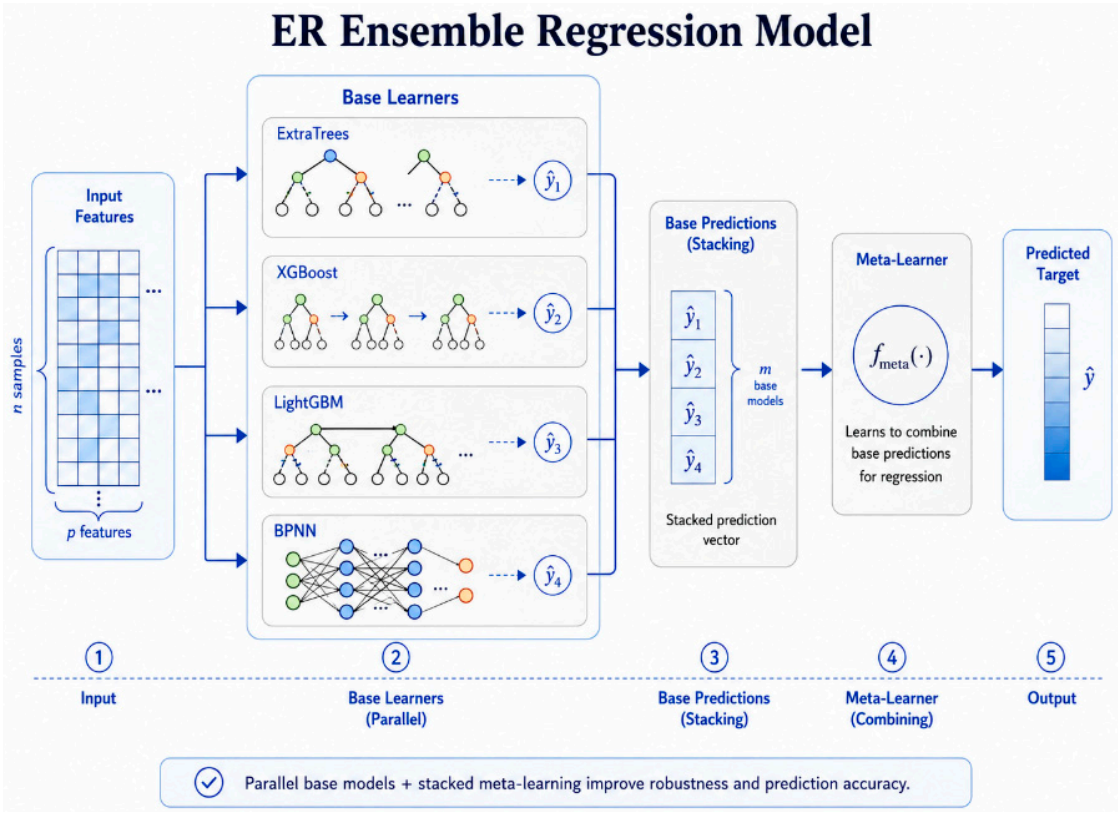

Figure 8. Structure of the ER ensemble-regression model. In the revised workflow, stacking folds follow the measured-depth order rather than randomly shuffled K-fold partitions.

4.7. Hyperparameter Optimization and Robustness

The search ranges and selected configurations are summarized in Table 2. All tree-model searches used TimeSeriesSplit and selected the configuration with the minimum mean forward-validation RMSE. The BPNN diagnostic search compared dropout values of 0.10 and 0.20, while keeping the 64–32 architecture, learning rate (0.001), and batch size (32) fixed.

Table 2. Forward-validation model configurations used in the revised analysis.

Model	Search Space/Candidates	Selected Configuration
ExtraTrees	n_estimators: 150–800; max_features: 0.50–1.00; min_samples_split: 2–15; min_samples_leaf: 1–8; max_depth: None/10–40	150; 0.85; 15; 8; 20
XGBoost	n_estimators: 150–800; learning_rate: 0.01–0.12; max_depth: 3–10; subsample/colsample_bytree: 0.60–1.00; gamma: 0–0.20; reg_alpha: 0–0.50; reg_lambda: 0.50–10	150; 0.03; 4; 1.0; 1.0; 0.5; 1.0
LightGBM	n_estimators: 150–800; learning_rate: 0.01–0.12; num_leaves: 15–95; max_depth: −1/5–15; min_child_samples: 10–50; feature_fraction/bagging_fraction: 0.60–1.00; reg_alpha: 0–0.50; reg_lambda: 0.50–5	800; 0.03; 15; 7; 20; 1.0; 0.6; 0.5; 0.5
BPNN	Hidden layers: 64–32; dropout: 0.10/0.20; learning rate: 0.001; batch size: 32	64–32; 0.20; 0.001; 32
ER	Meta-learner candidates: Linear Regression, Ridge, LASSO, ElasticNet	ET + XGB + LGBM + BPNN; LASSO (α = 0.001)

For the ER meta-learner, Linear Regression, Ridge, LASSO, and ElasticNet were compared under the same forward-validation logic. Their diagnostic forward RMSE values were 12.0320, 12.0318, 12.0311, and 12.0314, respectively. The differences were small; LASSO ($\alpha = 0.001$) was retained because it gave the lowest forward RMSE, rather than because of a material accuracy advantage.

The tree-model search minimized mean RMSE under TimeSeriesSplit. The corresponding forward-validation RMSE values were 4.721 for ExtraTrees, 4.282 for XGBoost, and 4.145 for LightGBM. The selected BPNN diagnostic configuration gave a forward-validation RMSE of 2.616. Because stochastic learners can depend on initialization, the ExtraTrees rolling-forward prediction was additionally repeated with five seeds (0, 7, 21, 42, and 100).

5. Model Evaluation Metrics

Prediction performance was evaluated using MAE, MSE, RMSE, R^2, MAPE, and Pearson correlation. The formulas are

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

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}},$$

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

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

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

Here, y_i and $\hat{y}_i$ denote the measured and predicted torque for sample i, respectively; an overbar denotes the corresponding sample mean; and n is the number of evaluated samples. Because this study focuses on sequential forward use rather than interpolation, all primary performance claims are based on unseen deeper intervals generated by the fixed stress test or rolling-origin procedure.

6. Results and Discussion

6.1. Depth Distribution Shift and Static Extrapolation Stress Test

Figure 9 shows the depth-ordered partition and reveals a strong increase in torque with well progression. The mean torque increased from approximately 20.03 in the initial training interval to 37.47 in the validation/calibration interval and 45.24 in the final test interval. Therefore, the randomly mixed train–test procedure in the original submission placed highly similar neighboring samples in both sets and substantially underestimated the difficulty of forward prediction.

The hyperparameter search spaces and the optimal values selected for each model are summarized in Table 3. All models degraded substantially under this long-range extrapolation. ER produced the lowest RMSE (4.890) and MAE (4.190) among the five candidates, but its R^2 remained negative (-0.323), and the single models also had negative R^2. These results demonstrate that a static model fitted in shallower intervals cannot be assumed to generalize over an approximately 900 m unseen deep interval. The negative

stress-test R^2 values are therefore interpreted as evidence of distribution shift, rather than as a basis for claiming high cross-depth accuracy.

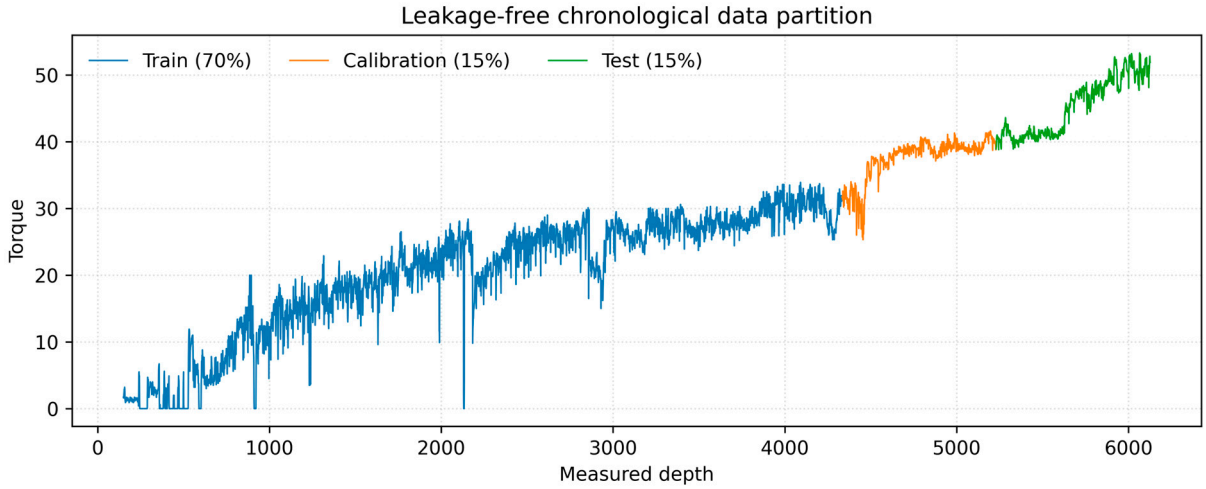

Figure 9. Leakage-free, depth-ordered partitioning of the field dataset into initial training, validation/calibration, and final test intervals.

Table 3. Static deepest-15% stress-test performance after refitting on the first 85% of the well.

Model	MAE	RMSE	MAPE (%)	R^2	Pearson r
ExtraTrees	5.570	6.992	11.550	−1.705	0.052
XGBoost	5.326	6.707	11.035	−1.489	0.522
LightGBM	4.799	6.191	9.898	−1.121	0.557
BPNN	12.770	13.523	27.553	−9.118	−0.732
ER	4.190	4.890	8.905	−0.323	−0.436

6.2. Rolling-Forward Torque Prediction

To approximate actual drilling deployment, the model was updated after each newly observed interval and then used to predict the next continuous 50 m. The 50 m rolling experiment covered 4333–6126 m and produced 1794 genuinely forward predictions. Because the BPNN and BPNN-based ER required substantially more repeated retraining and showed poor static deep-holdout behavior, the operational rolling comparison focused on the three tree-based models, which were computationally efficient for frequent updating. The predictive performance of all models on the strict depth-holdout test set is summarized in Table 4.

Table 4. Prediction performance under 50 m expanding-window, rolling-forward validation.

Model	MAE	RMSE	MAPE (%)	R^2	Pearson r
ExtraTrees	1.121	1.643	2.750	0.907	0.964
LightGBM	1.090	1.729	2.710	0.897	0.960
XGBoost	1.348	1.893	3.317	0.877	0.960

ExtraTrees achieved the lowest RMSE (1.643) and the highest R^2 (0.907) and Pearson correlation (0.964). LightGBM provided the lowest MAE (1.090) and MAPE (2.710%) with R^2 = 0.897, while XGBoost achieved R^2 = 0.877. Thus, model ranking under forward prediction differs from the originally reported random-split ranking. ExtraTrees and LightGBM emerge as the strongest practical candidates, rather than ER being treated as

universally optimal. The measured and predicted torque profiles of the five models along measured depth in the strict test interval are presented in Figure 10.

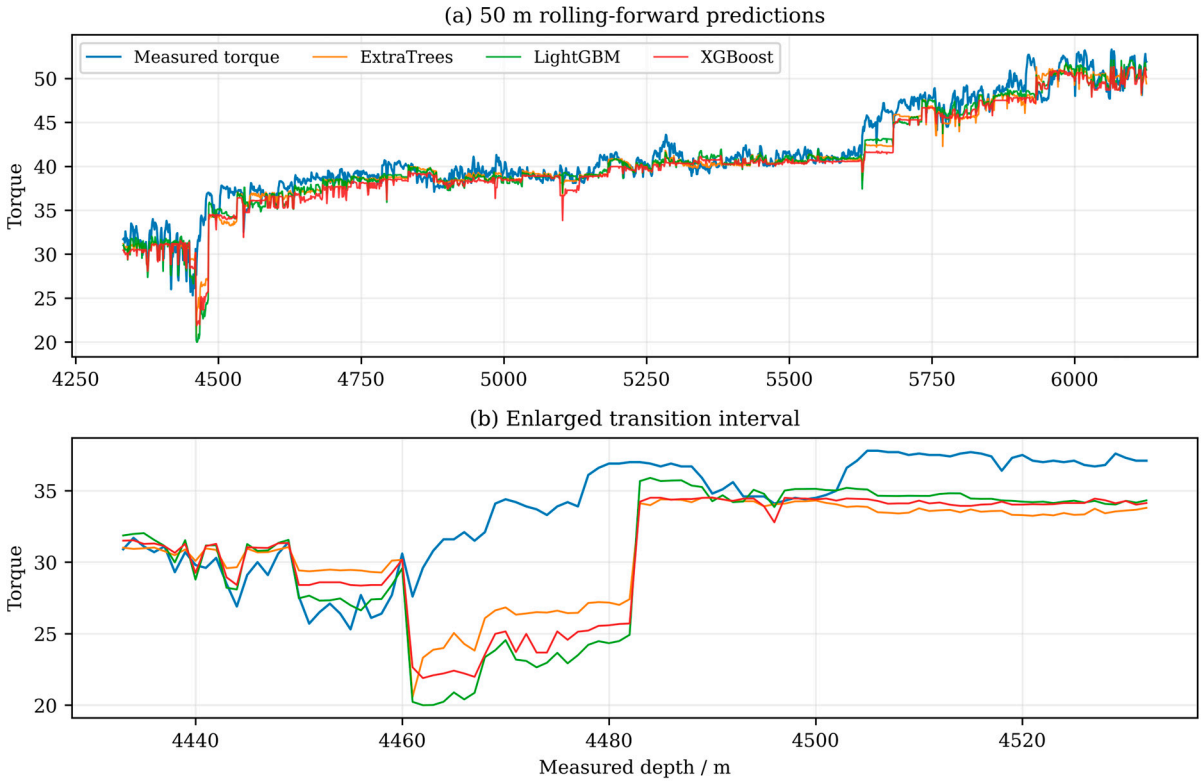

Figure 10. Measured torque and tree-model predictions under 50 m rolling-forward evaluation: (**a**) complete forward-prediction interval; (**b**) enlarged transition interval.

The depth-wise curves show that all three tree models track the progressive increase in torque, but larger discrepancies occur near abrupt regime changes. Figure 11 further diagnoses the representative ExtraTrees rolling-forward residuals. The residual distribution is concentrated around zero overall, while localized excursions occur in transition intervals, indicating that formation or operating-state changes can temporarily exceed the information represented by the current feature set.

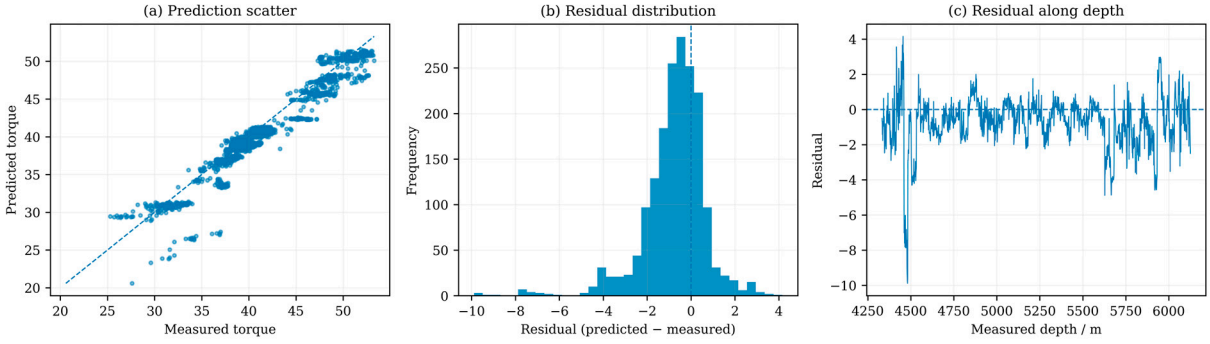

Figure 11. Error diagnosis of ExtraTrees under 50 m rolling-forward prediction: (**a**) measured vs. predicted scatter; (**b**) residual distribution; (**c**) residual variation with measured depth.

6.3. Prediction-Horizon Sensitivity and Stochastic Robustness

Prediction accuracy decreased monotonically as the look-ahead distance increased. R^2 declined from 0.927 at 25 m to 0.907 at 50 m, 0.802 at 100 m, and 0.711 at 200 m. RMSE increased from 1.454 to 2.901 over the same range. This trend quantifies the effect of concept

drift: the further the future interval lies from the most recently observed state, the less similar its formation and operating conditions become to the available training history.

The 50 m ExtraTrees experiment was repeated using five random seeds. RMSE was 1.640 ± 0.018 and R^2 was 0.9075 ± 0.0020 (mean ± standard deviation), demonstrating that the forward result is not an artifact of using a single random_state = 42 configuration. The performance of ExtraTrees and ER under the expanding-window rolling-forward validation is summarized in Table 5.

Table 5. Sensitivity of ExtraTrees rolling-forward performance to the prediction horizon.

Prediction Horizon (m)	MAE	RMSE	MAPE (%)	R^2	Pearson r
25	0.948	1.454	2.343	0.927	0.969
50	1.121	1.643	2.750	0.907	0.964
100	1.531	2.398	3.773	0.802	0.929
200	1.953	2.901	4.708	0.711	0.910

6.4. SHAP-Based Predictive and Controllable Factor Analysis

SHAP was used in place of permutation importance to improve consistency under correlated inputs. For the principal 50 m rolling ExtraTrees model, the largest mean absolute SHAP value was associated with Depth (8.63), followed by Rock Den (1.86), SP (1.82), CAL (1.53), WOB (1.40), FLWpmps (1.05), ROP (0.82), and Bit Run (0.44). Depth, therefore, acts as a strong predictive proxy for accumulated well extension and the evolving drilling state. However, SHAP importance is predictive rather than causal importance, and Depth cannot be directly adjusted by the driller.

To distinguish controllable factors from state proxies, an auxiliary control-oriented ExtraTrees model was built after excluding Depth/TVD while retaining WOB, RPM, and flow rate, and maintaining VIF < 10. Under the same 50 m rolling protocol, this model achieved MAE = 1.180, RMSE = 2.137, and R^2 = 0.843. Its SHAP ranking highlighted SSP, flow rate, formation density/caliper, RPM, WOB, and ROP. The reduction in predictive accuracy relative to the full model quantifies the information carried by depth/state proxies, while the retained performance shows that operational variables still contain useful control information. The absolute-error distributions of the five models in the training and strict test intervals are compared in Figure 12.

6.5. WOB–RPM–Torque Response and Parameter-Control Implications

A WOB–RPM response surface was constructed using the control-oriented model rather than the depth-dominated predictive model. Other variables were fixed at representative values within the observed domain. Figure 13 shows that predicted torque generally increases with WOB, while the RPM effect is nonlinear and interacts with WOB and the remaining state variables. This surface should therefore be interpreted as a model-based response within the observed data domain rather than as a causal law.

The engineering implication is that persistent high WOB combined with unfavorable rotational conditions can elevate torque, but parameter adjustment should not be based on WOB or RPM alone. Flow rate, standpipe pressure, hole cleaning, formation response, and current torque trend should be checked concurrently. This distinction directly addresses the difference between variables that explain prediction and variables that can be used for operational intervention.

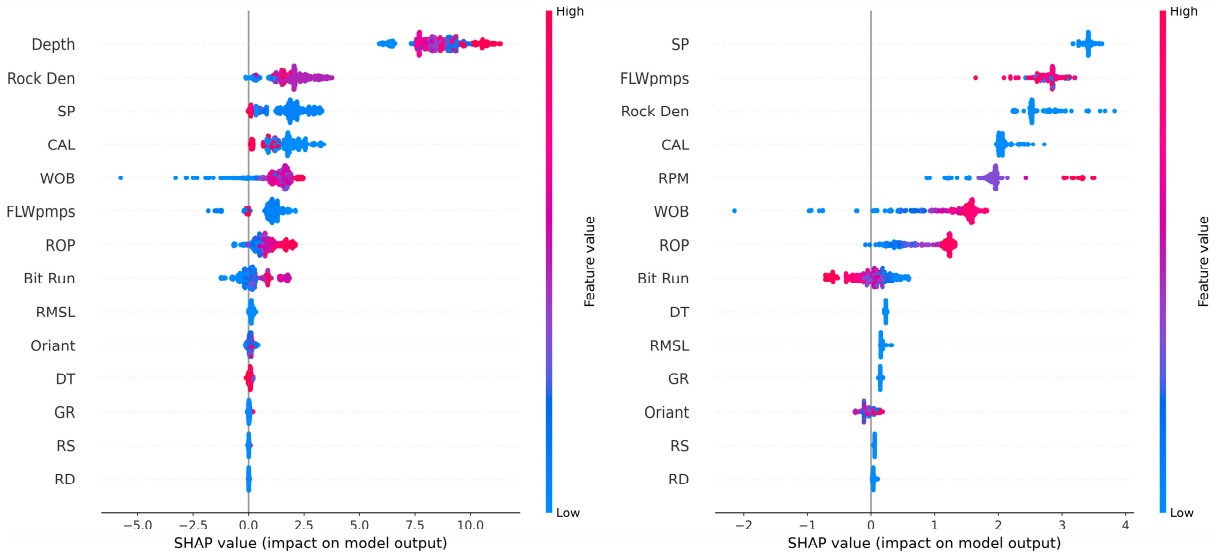

Figure 12. SHAP interpretation under rolling-forward validation: (**a**) predictive ExtraTrees model including Depth; (**b**) control-oriented ExtraTrees model excluding Depth/TVD and retaining operational variables.

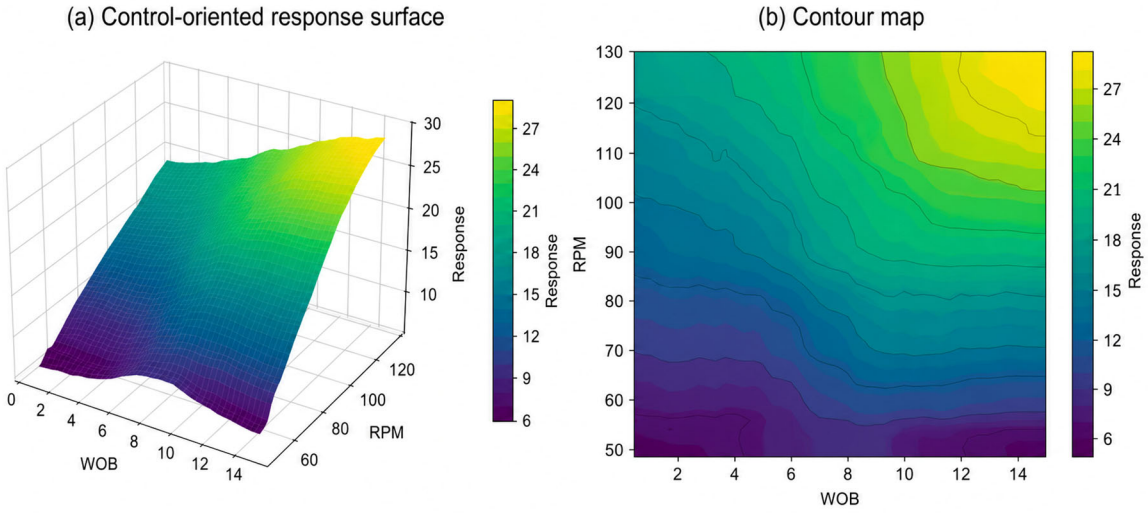

Figure 13. WOB–RPM–torque response surface derived from the control-oriented model: (**a**) three-dimensional surface; (**b**) contour map.

6.6. Static Versus Adaptive Validation, Uncertainty, and Study Limitations

Figure 14 contrasts the static deepest-15% stress test with rolling-forward validation. The large difference in RMSE demonstrates that updating the model with newly observed intervals is essential in this nonstationary well. The result is consistent with just-in-time and adaptive-learning concepts reported for drilling applications [2]. It also explains why the original random split produced R^2 values near 0.995: randomly interleaved samples largely measure interpolation within overlapping operating states rather than long-range forward generalization.

The originally submitted approximate 95% prediction interval was also re-examined. A split-conformal interval calibrated on the middle 15% of the well covered only approximately 38.5% of measurements in the final deep test interval, which is far below the nominal 95%. This failure is consistent with the strong distribution shift and violation of exchangeability. The revised manuscript therefore does not interpret the previous band

as a valid 95% prediction interval. Adaptive or rolling conformal calibration is required before uncertainty bands can be used as probabilistic risk thresholds in this setting.

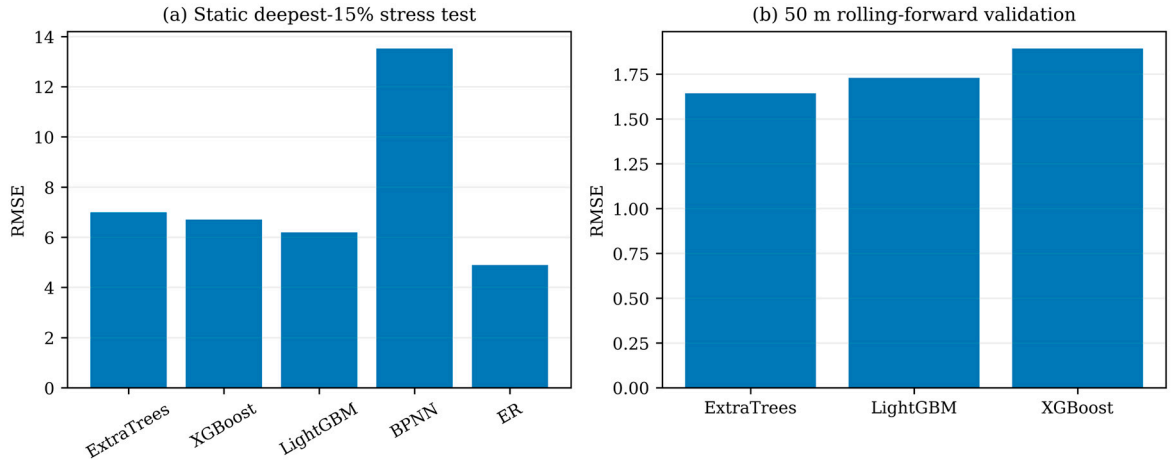

Figure 14. Comparison of RMSE under (**a**) static deepest-15% extrapolation stress testing and (**b**) 50 m rolling-forward validation. The panels represent different validation regimes and use independent vertical scales.

Several limitations remain. First, the study uses a single field well; therefore, the reported rolling-forward performance demonstrates within-well forward generalization rather than multi-well transferability. Independent-well validation is required before broader deployment. Second, the present models are data-driven and do not explicitly enforce torque-and-drag mechanics. Recent physics-informed drilling studies [23,25] indicate a promising direction for improving transferability. Third, the dataset does not contain fracture-spacing/core-image information from which a Fracture Index (FI) can be reconstructed. Wong et al. [37] showed that FI can be calculated from fracture-spacing information in core runs; incorporating such fracture descriptors or image-log equivalents is a valuable extension when corresponding data become available. Finally, logging variables should only be used for online prediction when their acquisition timing is compatible with the prediction depth.

7. Conclusions

A leakage-aware, multi-source framework was developed for torque prediction in a deepwater, extended-reach well. The principal conclusions are:

(1) Random mixing of depth-ordered samples produces overly optimistic performance because adjacent training and testing observations share highly similar operating and formation states. Strict depth-based stress testing exposed a substantial distribution shift between the shallow historical interval and the deepest section, so the original random-split $R^2 \approx 0.995$ should not be interpreted as true forward generalization.

(2) Rolling-origin updating substantially improved practical prediction. At a 50 m horizon, ExtraTrees achieved MAE = 1.121, RMSE = 1.643, MAPE = 2.750%, R^2 = 0.907, and Pearson r = 0.964; LightGBM achieved the lowest MAE of 1.090. ExtraTrees R^2 decreased from 0.927 at 25 m to 0.711 at 200 m, quantitatively demonstrating increasing concept drift with prediction distance.

(3) Training-only correlation screening and VIF analysis reduced multicollinearity to a maximum VIF of 5.834 in the principal predictive feature set. SHAP showed that Depth is the dominant predictive proxy, whereas WOB, flow rate, RPM, and related hydraulic variables are more relevant for operational intervention. A control-oriented

model excluding Depth/TVD retained an R^2 of 0.843 under 50 m rolling prediction, supporting separate treatment of predictive and controllable variables.

(4) Input smoothing did not consistently improve forward-validation accuracy, and raw measured torque was therefore retained as the prediction target. A static conformal interval also failed to preserve nominal coverage under the strong, deep-section distribution shift. These results emphasize that leakage prevention, adaptive updating, and validation under changing drilling states are more important than maximizing accuracy on randomly mixed samples.

The present results are limited to one well and do not yet incorporate explicit torque-and-drag physics, fracture descriptors, or independent-well external validation. Future work will combine additional wells, LWD-compatible real-time features, physics-informed constraints, adaptive uncertainty calibration, and just-in-time updating to improve transferability and operational reliability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pr14172801/s1.

Author Contributions: Conceptualization, J.G., P.L., W.L., B.C., Y.L. and X.C.; methodology, J.G., P.L., W.L., B.C., Y.L. and X.C.; software, J.G., P.L., W.L. and Y.L.; validation, J.G., P.L. and Y.L.; formal analysis, J.G. and P.L.; investigation, J.G., P.L. and X.Y.; resources, J.G. and P.T.; data curation, J.G. and P.T.; writing—original draft preparation, J.G., P.L. and B.C.; writing—review and editing, J.G., P.L., B.C. and X.C.; visualization, J.G. and X.Y.; supervision, J.G., X.C., X.Y. and P.T.; project administration, J.G. and P.T.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to commercial confidentiality and proprietary restrictions associated with the field drilling data. Access to the data is subject to approval by the data owner.

Conflicts of Interest: Authors Junrui Ge, Wei Liu, Bin Cai, Yanfei Li, Xiujin Yuan, and Penglei Tang were employed by CNOOC (China) Limited Shanghai Branch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Chen, X.Y.; Du, X.; Weng, C.K.; Yang, J.; Gao, D.; Su, D.; Wang, G. A real-time drilling-parameter optimization method for offshore large-scale cluster extended-reach drilling based on an intelligent optimization algorithm and machine learning. *Ocean Eng.* **2024**, *291*, 116375. [CrossRef]
2. Bai, K.K.; Sheng, M.; Zhang, H.B.; Fan, H.H.; Pan, S.W. Real-time drilling torque prediction ahead of the bit with just-in-time learning. *Pet. Sci.* **2025**, *22*, 430–441. [CrossRef]
3. Cao, W.P.; Mei, D.Y.; Guo, Y.M.; Ghorbani, H. Deep-learning prediction of drill-bit torque in directional-drilling sliding mode: Energy saving. *Measurement* **2025**, *250*, 117144. [CrossRef]
4. Johancsik, C.A.; Friesen, D.B.; Dawson, R. Torque and drag in directional wells: Prediction and measurement. *J. Pet. Technol.* **1984**, *36*, 987–992. [CrossRef]
5. Ho, H.S. An improved modeling program for computing torque and drag in directional and deep wells. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 2–5 October 1988.
6. Zhong, R.Z.; Salehi, C.; Johnson, R., Jr. Machine learning for drilling applications: A review. *J. Nat. Gas Sci. Eng.* **2022**, *108*, 104807. [CrossRef]
7. Hazbeh, O.; Aghdam, S.K.; Ghorbani, H.; Mohamadian, N.; Ahmadi Alvar, M.; Moghadasi, J. Comparison of accuracy and computational performance among machine-learning algorithms for rate of penetration in directional drilling. *Pet. Res.* **2021**, *6*, 271–282. [CrossRef]

8. Chen, H.; Duan, J.Y.; Yin, R.; Ponkratov, V.V.; Guerrero, J.W.G. Prediction of penetration rate by coupled simulated annealing–least-squares support-vector-machine learning in a hydrocarbon formation based on drilling parameters. *Energy Rep.* **2021**, *7*, 3971–3978. [CrossRef]
9. Bizhani, M.; Kuru, E. Towards drilling rate-of-penetration prediction: Bayesian neural networks for uncertainty quantification. *J. Pet. Sci. Eng.* **2022**, *219*, 111068. [CrossRef]
10. Yehia, T.; Gasser, M.; Ebaid, H.; Meehan, N.; Okoroafor, E.R. Comparative analysis of machine-learning techniques for predicting drilling rate of penetration in geothermal wells: A case study of the FORGE site. *Geothermics* **2024**, *120*, 103028. [CrossRef]
11. Sun, J.; Chen, M.; Li, Q.; Ren, L.; Dou, M.Y.; Zhang, J.X. A new method for predicting formation lithology while drilling at the bit of a horizontal well. *J. Pet. Sci. Eng.* **2021**, *196*, 107955. [CrossRef]
12. Ye, Z.H.; Guo, S.Y.; Chen, D.; Wang, H.; Li, S.D. Drilling-formation perception by supervised learning: Model evaluation and parameter analysis. *J. Nat. Gas Sci. Eng.* **2021**, *90*, 103923. [CrossRef]
13. Liang, H.; Chen, H.; Guo, J.; Bai, J.; Jiang, Y.J. Lithology identification based on the mechanical-specific-energy principle and machine-learning theory. *Expert Syst. Appl.* **2022**, *191*, 116142. [CrossRef]
14. Ameur-Zaimeche, O.; Kechiched, R.; Heddam, S.; Wood, D.A. Real-time porosity prediction using gas-while-drilling data and machine learning with reservoir-associated gas: A case study of the Hassi Messaoud field, Algeria. *Mar. Pet. Geol.* **2022**, *140*, 105631. [CrossRef]
15. Pang, H.W.; Wang, H.Q.; Xiao, Y.T.; Jin, Y.; Lu, Y.H.; Fan, Y.D.; Nie, Z. Machine learning for carbonate-formation drilling: Mud-loss prediction using seismic attributes and mud-loss records. *Pet. Sci.* **2024**, *21*, 1241–1256.
16. Agwu, O.E.; Alatefi, S.; Yusof, M.A.M.; Orun, C.B. Predicting drilling mud equivalent circulating density with precision: A critical review of modern approaches. *Petroleum* **2025**, *11*, 699–716. [CrossRef]
17. Jiang, G.C.; Dong, T.F.; Cui, K.X.; He, Y.B.; Quan, X.H.; Yang, L.L.; Fu, Y. Research status and development directions of intelligent drilling-fluid technologies. *Pet. Explor. Dev.* **2022**, *49*, 660–670. [CrossRef]
18. Oyedere, M.; Gray, K. Torque-on-bit prediction and optimization using machine-learning algorithms. *J. Nat. Gas Sci. Eng.* **2020**, *84*, 103623. [CrossRef]
19. Marquez, F.J. Drilling optimization applying machine-learning regression algorithms. In Proceedings of the Offshore Technology Conference, Virtual and Houston, TX, USA, 16–19 August 2021. OTC-30934-MS.
20. Song, X.Z.; Zhu, S.; Li, G.S.; Zeng, Y.; Guo, H.; Hu, Z. Prediction of hook load and rotary-table torque based on a BP-LSTM dual-input network. *J. China Univ. Pet. Ed. Nat. Sci.* **2022**, *46*, 76–84.
21. Gurina, E.; Klyuchnikov, N.; Antipova, K.; Koroteev, D. Making the black box brighter: Interpreting machine-learning algorithms for forecasting drilling accidents. *J. Pet. Sci. Eng.* **2022**, *218*, 111041. [CrossRef]
22. Ge, Z.G.; Zhou, X.K.; Li, Y.; Zhang, X.; Li, W. Interpretable machine-learning approaches for damage identification in drilling risers. *Ocean Eng.* **2024**, *309*, 118495. [CrossRef]
23. He, M.; Yuan, Q.-R.; Li, J.; Zhou, C.-C.; Liu, J.-W.; Chen, X.; Cao, J.-H. A novel method for drilling rate of penetration prediction and transfer generalization based on physics-informed neural networks. *Pet. Sci.* **2026**, *23*, 4009–4033. [CrossRef]
24. Song, Z.; Song, Y.; Yang, J.; Liu, B.; Gao, B.; Tang, J. A multi-objective reinforcement learning framework for real-time drilling optimization based on symbolic regression and perception. *Geoenergy Sci. Eng.* **2025**, *244*, 213392. [CrossRef]
25. Wada, R.; Kaneko, T. Integration of physics-based models and data-driven models in drilling technology. *J. Jpn. Petrol. Inst.* **2026**, *69*, 1–10. [CrossRef]
26. Hampel, F.R. The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.* **1974**, *69*, 383–393. [CrossRef]
27. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [CrossRef]
28. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
29. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [CrossRef]
30. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
31. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
32. Chen, T.; Guestrin, C. XGBoost: A scalable tree-boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
33. Ke, G.L.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient-boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
34. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
35. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.

36. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]
37. Wong, L.N.Y.; Liu, Z.; Tse, K.K.C.; Cheung, S.H.; Yu, L. A computational algorithm for calculating fracture index of core runs. *Rock Mech. Rock Eng.* **2023**, *56*, 6905–6918. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.