

IRSTI 52.47.19, 28.23.29

<https://doi.org/10.26577/jpcsit5320261>Y. Koshkarov ^{1*} , M. Yesmagambetova ² , O. Mulesa ³ ¹ Astana IT University, Astana, Kazakhstan² Karaganda University of Kazpotreboysuz, Karaganda, Kazakhstan³ University of Prešov, Prešov, Slovakia

*e-mail: yerzhan.koshkarov@astanait.edu.kz

SELF-SUPERVISED FOUNDATION MODELS FOR REAL-TIME DRILLING OPTIMIZATION: CROSS-MODAL KNOWLEDGE TRANSFER FROM WELL LOG REPRESENTATIONS

Abstract. The optimization of real-time drilling faces significant hurdles stemming from limited data availability, complex operational environments, and a lack of machine learning models that can transfer knowledge across different geological settings. Current supervised methods deliver strong results in predicting rate of penetration and detecting stuck pipe incidents, yet they demand extensive labeled datasets for each new well and fail to tap into the geological insights contained in commonly available well log measurements. We present DrillFM, a self-supervised foundation model designed for multivariate drilling time-series data that incorporates geological information through cross-modal conditioning from pretrained well log representations. Our architecture merges masked autoencoding of drilling parameters—including weight-on-bit, rotary speed, torque, standpipe pressure, and flow rate—with a cross-modal attention mechanism that draws upon formation embeddings generated by a physics-aware well log foundation model. The training employs a drilling-adaptive curriculum that escalates task complexity progressively: beginning with point-wise masking, advancing to channel-wise masking, and culminating in interval-based prediction. On the publicly available Volve field dataset plus proprietary multi-basin data, geological conditioning cuts rate of penetration prediction error by 18.3 percent relative to the supervised multi-modal baseline. Few-shot transfer to previously unseen wells achieves comparable performance using 90 percent less labeled data. In cross-basin transfer of rate-of-penetration prediction, DrillFM’s performance degrades by 38.9 percent moving from Norwegian Volve to Dutch North Sea wells, 13.0 percentage points less than the supervised multi-modal baseline’s 51.9 percent degradation. The model delivers inference speeds of 8.6 milliseconds per depth sample using precomputed well log embeddings; because the reported results condition on log data from the same well being predicted, whether this speed and accuracy hold under fully real-time deployment—where only offset-well or no logs would be available—has not been separately evaluated (see Limitations). Ablation experiments show that well log foundation model pretraining is the single largest contributor, producing a 17.7 percent error increase when removed; removing geological conditioning overall accounts for a 12.5 percentage point change, and the masking curriculum contributes a further 5.7 points.

Keywords: drilling optimization, foundation models, self-supervised learning, cross-modal transfer, rate of penetration, stuck pipe detection, well logs, time-series prediction.

1. Introduction

Among all phases of subsurface resource development—whether for petroleum extraction, geothermal energy, or carbon sequestration—drilling stands out as the most expensive. Industry data shows it consumes between thirty and seventy percent of total well expenditures [1]. The challenge of reducing these costs has pushed the sector toward real-time optimization of drilling parameters, yet the industry still hemorrhages roughly six billion dollars each year through non-productive time. This waste stems from various drilling problems: stuck pipe incidents, lost circulation events, and wellbore instabil-

ity issues [3], [4]. Successfully optimizing the rate of penetration means finding the sweet spot between drilling as quickly as possible while maintaining both safety standards and wellbore integrity. This balancing act requires operators to constantly adjust controllable variables like weight-on-bit, rotary speed, and mud flow rate as they respond to shifts in formation characteristics, equipment performance, and operational constraints.

Over recent years, machine learning has become a valuable tool in the drilling optimization toolkit. When trained on substantial labeled datasets, supervised models can predict rate of penetration with R-squared values exceeding 0.95 and detect stuck

pipe scenarios with ninety percent accuracy [5, 6]. The performance of ensemble methods like random forests and gradient boosting has been particularly impressive on historical drilling data. Even more striking results come from hybrid approaches that blend physics-based models-such as the Bourgoyne-Young equation and Bingham plasticity formulations-with machine learning corrections, achieving R-squared values as high as 0.9936 on multi-well datasets [7]. Despite these successes, three fundamental constraints limit how widely these supervised methods can be deployed in practice.

The first limitation concerns data requirements. Supervised drilling models need extensive labeled datasets for training purposes, with typical research studies utilizing anywhere from 5,000 to 50,000 labeled depth samples for each well [8], [9]. Creating these labels manually is both time-intensive and demands significant domain expertise-someone must identify drilling dysfunctions, classify lithology from cuttings samples, and annotate operational events accurately. Many wells simply lack thorough annotations, particularly for infrequent yet expensive events like stuck pipe or lost circulation. The second issue relates to transferability. When a model is trained using data from one well or geological basin, it often performs poorly when applied to different settings. This performance degradation occurs because of distribution shift caused by variations in formation properties, drilling equipment specifications, mud system design, and operational practices [10], [11]. Consequently, each new application requires retraining the model, which eliminates any economies of scale that machine learning deployment might otherwise provide.

Perhaps most critically, current drilling optimization models overlook something fundamental: the geological context that actually drives drilling behavior. Formation lithology, porosity, permeability, pore pressure, and mechanical rock properties determine how a formation responds while it is being drilled. Yet during real-time drilling operations, these properties typically remain unknown. Tools that measure formation characteristics with high resolution-well logs-exist, but they are only acquired after drilling happens, either through wire-line logging or logging-while-drilling tools; this timing constraint is central to the deployment limitations discussed in Section 4.2 and the Limitations section. This creates a situation where tens of thousands of wells globally have well log data representing an enormous knowledge base about geological properties, but drilling optimization systems largely

ignore this resource [12], [13]. Current approaches treat drilling as if it were a standalone time-series forecasting problem, without leveraging this complementary subsurface information.

Well log and seismic foundation models have advanced rapidly in adjacent subsurface domains, yet drilling engineering remains untapped by self-supervised pre-training approaches: no foundation model exists yet for drilling time-series data, and no prior work has explored transferring knowledge from geological representations learned by well log foundation models to drilling optimization tasks. This represents a missed opportunity, since well log foundation models have already learned to encode formation properties, stratigraphic relationships, and petrophysical patterns from large unlabeled datasets-knowledge that should benefit drilling prediction, given that formation characteristics drive drilling behavior [14], [15].

We address these gaps through four main contributions. First, we introduce DrillFM-the initial self-supervised foundation model purpose-built for multivariate drilling time-series. Our model undergoes masked autoencoding pre-training on drilling parameters collected from forty-seven wells spanning three distinct geological basins. The beauty of this approach is that the model develops rich representations of drilling dynamics without needing labeled ROP targets or dysfunction annotations. Second, we propose an architecture that achieves cross-modal conditioning by incorporating geological embeddings from a pre-trained well log foundation model through cross-attention mechanisms. This design lets the drilling model leverage formation properties during prediction even when real-time LWD data is not available. Third, we develop a drilling-adaptive masking curriculum that progressively ramps up task difficulty across three stages: it starts with random point masking for learning local interpolation, moves to channel-wise masking for understanding cross-parameter dependencies, and finishes with interval-based masking for temporal sequence prediction. Finally, we establish thorough evaluation benchmarks for both cross-well and cross-basin transfer scenarios, demonstrating that geological conditioning enables effective few-shot learning and enhances generalization to unseen operational settings.

2. Literature review

Machine learning applications to drilling optimization have evolved over two decades, starting

with classical statistical models and progressing to today's deep learning architectures. Initial efforts concentrated on supervised regression for ROP prediction, using drilling parameters as input features. These early systems-employing multiple linear regression, support vector machines, and basic neural networks-achieved only modest accuracy on the limited datasets available [16], [17]. The real breakthrough arrived with ensemble methods. Random forests and gradient boosting consistently deliver R-squared values above 0.90 for ROP prediction by successfully capturing the nonlinear relationships connecting controllable parameters like WOB, RPM, and flow rate to formation-dependent outcomes [6]. More recent comprehensive studies on multi-well datasets show that XGBoost and LightGBM, when combined with careful feature engineering, can reach R-squared values between 0.96 and 0.98, establishing them as a common supervised ROP prediction baseline.

Adapting deep learning architectures to drilling time-series has produced mixed results. Recurrent neural networks, including LSTMs and GRUs, successfully capture temporal dependencies in how drilling parameters evolve. They achieve competitive performance on sequential prediction tasks, though they demand extensive hyperparameter tuning and can suffer from training instability [8], [19]. Convolutional neural networks, when applied to windowed drilling signals, extract local patterns effectively but struggle with the long-range dependencies that are critical for anticipating formation changes [20]. More recent research has employed Transformer architectures for drilling prediction. Their attention mechanisms allow the model to weight relevant historical contexts when predicting future states. One particularly successful example is the GRU-Informer hybrid architecture, which combines recurrent feature extraction with efficient long-sequence attention to achieve superior real-time ROP prediction [9].

For drilling applications, physics-informed machine learning has proven especially successful. The key insight is combining mechanistic drilling equations with data-driven corrections-this hybrid approach consistently outperforms purely statistical methods. Consider the Bourgoyne-Young equation, which models ROP as a function of formation properties, bit characteristics, and operational parameters. It provides a physics-based starting point that machine learning can then refine using actual field data. Jiao et al. showed that training neural networks to predict deviations from Bourgoyne-Young pre-

dictions achieves R-squared equal to 0.9936. This substantially exceeds standalone ML (R-squared of 0.89) or physics-only approaches (R-squared of 0.84) [7]. Similar results appear when incorporating Bingham plastic rheology for hydraulics modeling and Mohr-Coulomb failure criteria for rock mechanics-these physics-informed models outperform black-box deep learning on drilling datasets from the Tarim Basin [22].

Detecting and preventing drilling dysfunctions has drawn significant research attention, which makes sense given the substantial costs involved with non-productive time. Stuck pipe scenarios-where the drill string becomes mechanically or differentially stuck in the wellbore-represent the single most expensive drilling failure mode. The industry estimates these events cost approximately two hundred fifty million dollars annually [23], [24]. Detecting problems early is critical because intervention becomes exponentially harder once sticking has initiated. Machine learning approaches to stuck pipe detection take several forms: supervised classification based on drilling parameter anomalies, unsupervised LSTM-autoencoders that flag deviations from normal operational patterns, and ensemble methods that combine multiple detection strategies. Mopuri et al. recently demonstrated that LSTM-autoencoders trained on thirty wells' worth of data achieve high sensitivity for stuck pipe precursors, although generalization to new geological settings remains challenging [25].

The most sophisticated stuck pipe detection systems today use ensemble architectures. A 2025 study combined autoencoder anomaly detection, BiLSTM sequence classification, and Transformer attention mechanisms with operational constraint verification, achieving 90.1 percent accuracy on multi-well datasets [26]. This approach also incorporated understanding of drilling mechanisms-distinguishing between mechanical sticking (caused by friction) and differential sticking (caused by pressure imbalance)-to provide actionable insights rather than just binary alerts. Even with this progress, a comprehensive review by Montes et al. identified the lack of generalizable, transferable stuck pipe prediction methods as the field's primary limitation. They noted that every published model is trained and evaluated on data from a single field or basin [3].

Transfer learning for drilling applications has been explored less than in other time-series domains. Yahia et al.'s survey of domain generalization for time-series machine learning [27] reviews

adversarial domain-generalization approaches relevant to drilling-adjacent problems such as stick-slip vibration prediction, in which domain-invariant feature extractors trained via adversarial discrimination between source and target wells have been reported to reduce classification error substantially when transferring between North Sea wells. These approaches still fundamentally require supervision: labeled events from both source and target domains remain necessary. To our knowledge, no prior work has applied self-supervised or foundation model approaches to drilling data, even though these paradigms have demonstrated success in reducing labeled data requirements in other domains [28], [29].

3. Problem formulation

We frame the problem as self-supervised learning on multivariate drilling time-series, incorporating geological conditioning from well log representations. Our drilling dataset consists of N wells. Each well w can be represented as a depth-indexed matrix containing D depth samples across C drilling parameter channels. The standard drilling measurements include weight-on-bit (WOB), rotary speed (RPM), surface torque (TRQ), standpipe pressure (SPP), mud flow rate (Q), rate of penetration (ROP), and mud weight (MW). For wells equipped with logging-while-drilling sensors, we also have formation measurements available: gamma ray (GR), resistivity (RES), neutron porosity (NPHI), and bulk density (RHOB). Typical depth sampling intervals are 0.1524 meters for LWD measurements and 0.3048 meters for surface drilling parameters.

Following the approach of recent time-series foundation models, we partition the depth sequence into non-overlapping patches of length L , which yields P patches per well. We set $L = 0.3048$ m, matching the native sampling interval of the surface drilling channels, so each patch corresponds to one depth sample; this is distinct from the larger 20-meter masking span used in the Stage 3 curriculum (Section 4.3), which spans approximately 66 consecutive patches. A learned linear layer projects each patch to an embedding space with dimension d equal to 256. We denote the drilling patch embeddings for well w as H . During the self-supervised pre-training phase, we randomly mask a subset M of drilling patches using a masking ratio of 0.40. The pre-training objective combines reconstruction loss with cross-modal alignment when well log data is available.

For the reconstruction component, we follow standard masked autoencoding practices on drilling parameters. When wells include well log measurements, we bring in a pre-trained well log foundation model. The cross-modal alignment loss then encourages the drilling embeddings to predict their corresponding geological context. We balance the reconstruction and cross-modal objectives by combining them with a weight of 0.3 for the alignment term. Formally, for a masked patch set M , the reconstruction loss is the mean squared error between each masked patch’s true (standardized) value x_i and the decoder’s reconstruction $\hat{x}_i$: $L_{\text{recon}} = (1/|M|) * \sum_{i \in M} \|x_i - \hat{x}_i\|^2$. Let g_i denote the frozen well log foundation model’s geological embedding aligned to drilling patch i , and h_i the corresponding drilling-encoder embedding; the cross-modal alignment loss is $L_{\text{align}} = (1/|M|) * \sum_{i \in M} \|h_i - g_i\|^2$. The total pre-training objective is $L = L_{\text{recon}} + 0.3 * L_{\text{align}}$.

For downstream applications, we focus on two primary tasks. First, rate of penetration prediction is formulated as a regression problem. The model predicts ROP at depth $t+1$ given the drilling history up to depth t , plus optional geological context. The objective here minimizes mean squared error. Second, stuck pipe detection is formulated as binary classification. The model predicts whether sticking will occur within the next k depth samples. Here the objective minimizes binary cross-entropy, using class weights to handle the severe class imbalance (stuck pipe events make up less than one percent of training samples).

The key innovation lies in incorporating geological embeddings from the pre-trained well log foundation model during both pre-training (via alignment loss) and downstream prediction (via cross-attention conditioning). This enables drilling models to leverage formation properties even when real-time LWD sensors are unavailable.

4. Methods

Our proposed DrillFM framework integrates three key components: a Transformer-based drilling encoder trained via masked autoencoding, a cross-modal conditioning module that incorporates geological embeddings from a pre-trained well log foundation model, and a drilling-adaptive curriculum that progressively increases pre-training task difficulty.

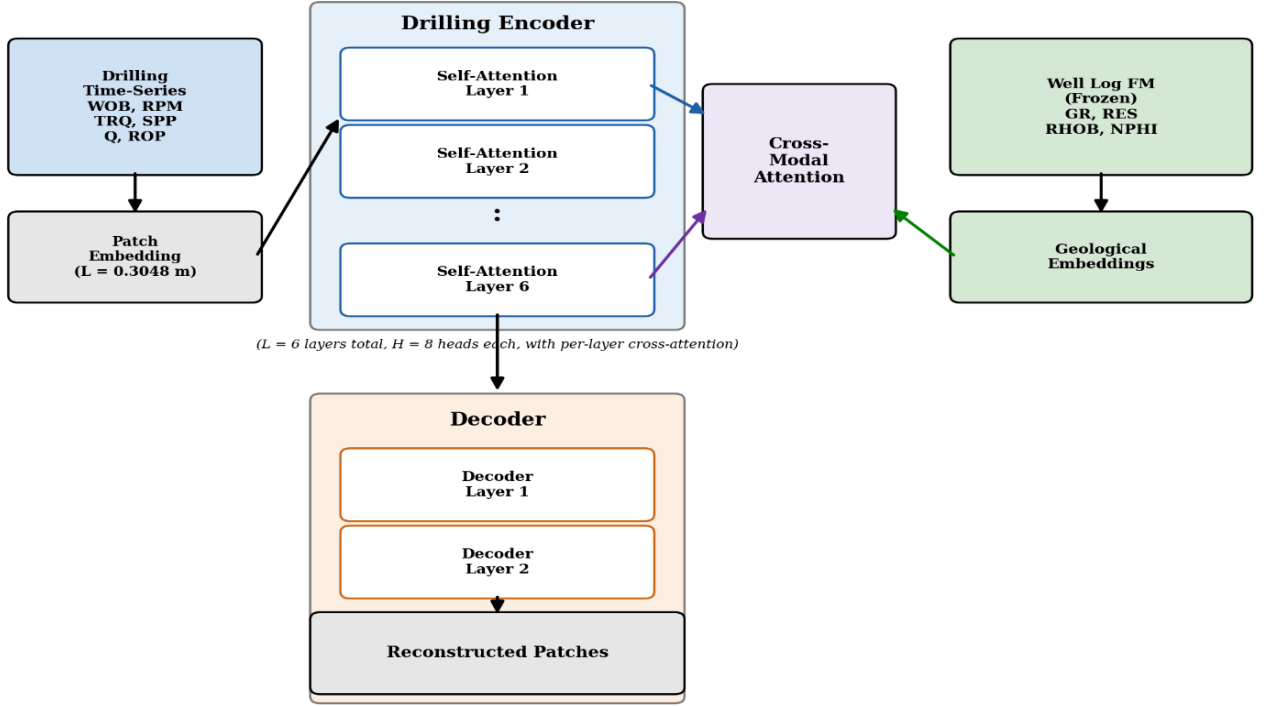


Figure 1. DrillFM System Architecture. The framework integrates three key components: (left) drilling time-series input processing through patch embedding, (center) multi-layer Transformer encoder with self-attention, and (right) cross-modal conditioning from a frozen well log foundation model. Arrows indicate data flow from input through encoder, cross-attention conditioning, decoder, to reconstructed output patches.

4.1 Drilling Time-Series Encoder

The encoder processes multivariate drilling time-series through a Transformer architecture adapted for depth-indexed measurements. Initial patch embeddings are computed from input patches through a two-layer MLP with hidden dimension 256. We employ learned positional encodings that capture depth information, as drilling signals exhibit strong depth-dependent patterns due to changing formation properties. Additionally, separate channel embeddings indicate which drilling parameters are present in each patch, handling the heterogeneous sensor availability across wells.

The encoder comprises L equal to six Transformer layers featuring multi-head self-attention with H equal to eight heads, along with position-wise feed-forward networks having an expansion ratio of four. We use layer normalization before both attention and feed-forward operations rather than after, following recent practices in time-series modeling that show this arrangement stabilizes training and improves convergence. Residual connections maintain gradient flow through deep architectures. The encoder's final output provides embeddings that

capture both local parameter interactions and long-range temporal dependencies spanning multiple formation intervals.

4.2 Cross-Modal Geological Conditioning

The innovation enabling knowledge transfer from well logs to drilling optimization is our cross-modal conditioning mechanism: a pre-trained physics-aware well log foundation model extracts geological embeddings from formation measurements, and the well log encoder is frozen during drilling model training. Freezing prevents catastrophic forgetting of the petrophysical relationships learned during well log pre-training, preserving those geological representations intact.

We implement geological conditioning through cross-attention at multiple levels within the drilling encoder. At each Transformer layer, after performing self-attention within drilling embeddings, we apply cross-attention from drilling to geological embeddings. This mechanism allows each drilling embedding to attend to relevant geological context based on semantic similarity in the shared embedding space. The attention mechanism naturally han-

dles varying spatial alignment between drilling and logging measurements-LWD tools are typically off-set ten to thirty feet behind the drill bit, and wireline logs are acquired after drilling completes-yet the attention weights learn to associate drilling responses with corresponding formation properties despite these spatial-temporal misalignments.

For wells without logging data (neither LWD nor wireline), we employ a learned null embedding that represents unknown formation properties. During pre-training on wells with logs, the alignment loss encourages drilling embeddings alone to predict geological context, enabling the model to infer formation characteristics from drilling responses even without direct logging measurements. This creates an implicit formation property estimator within the drilling encoder that activates when geological em-

beddings are unavailable, similar to virtual sensor approaches that predict well log responses from drilling data.

4.3 Drilling-Adaptive Masking Curriculum

Standard masked autoencoding employs uniform random masking throughout training, but drilling time-series exhibit unique structure suggesting a more sophisticated approach. Drilling parameters show high correlation (for example, weight-on-bit and torque both increase with formation strength), exhibit sharp changes at formation boundaries, and follow operational sequences imposed by drilling procedures. We design a three-stage masking curriculum that progressively increases task difficulty by exploiting these characteristics.

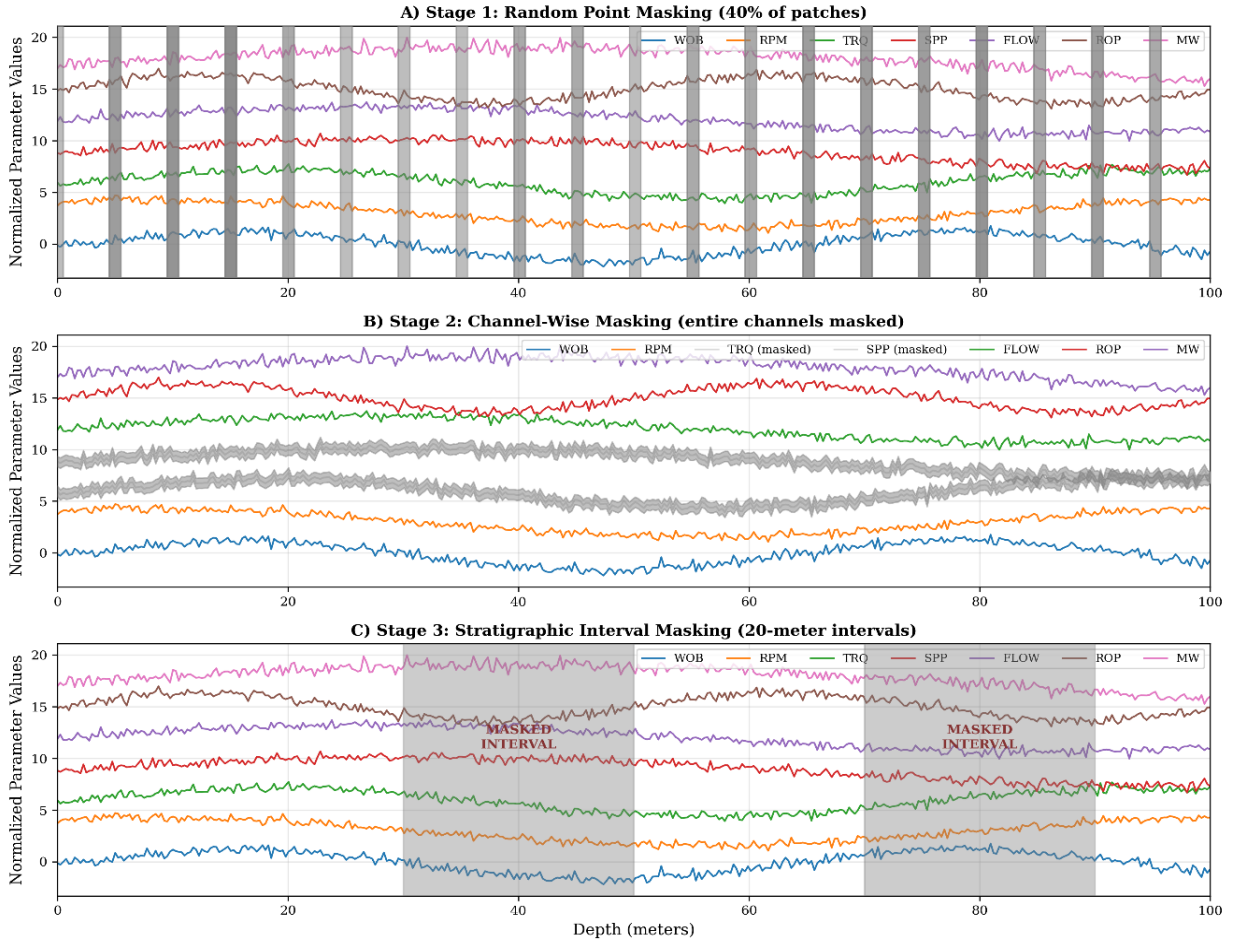


Figure 2. Drilling-Adaptive Masking Curriculum. Three-stage progressive masking strategy: (A) Stage 1 shows random point masking at 40% ratio where individual patches are masked uniformly, (B) Stage 2 demonstrates channel-wise masking where entire drilling parameter channels (TRQ, SPP) are masked across all depths, and (C) Stage 3 illustrates stratigraphic interval masking where contiguous 20-meter depth intervals are masked across all channels. Gray regions indicate masked data.

Stage 1 (epochs 1-50): Random point masking masks individual patches uniformly at random with ratio 0.40. This teaches the model basic interpolation from surrounding context and establishes foundational understanding of local drilling dynamics. The model learns that WOB, RPM, and ROP typically vary smoothly within formations but can change abruptly at lithology contacts.

Stage 2 runs from epochs 51 through 100 and employs channel-wise masking: entire drilling parameter channels are masked across all depth samples, with each channel having a thirty percent probability of being masked. For example, all torque measurements may be masked while WOB, RPM, and flow rate remain visible, forcing the model to infer cross-parameter physical relationships-such as inferring torque from the combination of bit weight, rotation speed, and formation strength indicators. Channel-wise masking works particularly well for drilling because parameters are mechanically coupled through the drilling system: knowledge of how the bit responds to applied forces constrains the possible torque values.

Stage 3 (epochs 101-150): Interval-based masking masks contiguous depth intervals of approximately 66 patches (corresponding to 20 meters, matching the masking span introduced in Section 3) across all channels simultaneously. This requires the model to leverage both geological context from well log embeddings and drilling trends from unmasked intervals above and below. Interval masking forces learning of formation-scale patterns and operational sequences-for example, recognizing that increased WOB and decreased RPM often precede bit changes, or that gradual ROP decline indicates bit wear requiring parameter adjustment.

The curriculum is motivated by human expert reasoning during drilling operations: novice drillers first learn to interpolate normal parameter values, then develop understanding of how parameters interact physically, and finally gain intuition for anticipating drilling behavior across entire formation intervals based on geological priors.

4.4 Training Procedure

Pre-training proceeds through the three-stage curriculum for 150 epochs total. We employ the AdamW optimizer with initial learning rate $3e-4$, weight decay 0.01, and cosine annealing schedule that decays to ten percent of initial rate by final epoch. Batch size is set to 32 sequences (patch windows sampled across the 36 training wells, not one batch element per well) with gradient accumula-

tion over four steps for effective batch size 128. All drilling parameters undergo standardization to zero mean and unit variance using training set statistics, with bounds clipping applied to remove sensor outliers.

The well log foundation model is pre-trained separately on a large corpus of wells (18,000 wells from public and proprietary sources) using the physics-aware masked autoencoding approach incorporating petrophysical constraints. This encoder is frozen during drilling model training, serving purely as a feature extractor for geological conditioning. Fine-tuning for downstream tasks uses learning rate $1e-3$ for 50 epochs, with early stopping based on validation loss. The lightweight task heads train rapidly, typically converging within 20 epochs even on small labeled datasets.

Complete training requires approximately thirty-six hours on a single NVIDIA A100 40GB GPU for the full pre-training curriculum plus downstream task fine-tuning. The pre-trained model comprises approximately 10.5 million parameters, breaking down to approximately 6.3 million in the encoder-consistent with $d = 256$, six Transformer layers, eight attention heads, an expansion ratio of four, and per-layer cross-attention-and 4.2 million in the decoder. When quantized to FP16 precision for deployment, this yields a memory footprint of approximately 21 MB.

5. Experimental Setup

We evaluate DrillFM on public and proprietary drilling datasets spanning multiple geological basins. The primary dataset comes from the Volve field, publicly released by Equinor, containing 24 wells from the Norwegian North Sea with comprehensive surface drilling parameters and LWD measurements [30]. Wells penetrate formations ranging from Jurassic to Paleocene age, including sandstones, shales, and carbonates, with total depths spanning 2,800 to 3,600 meters, including the shallower Eocene-to-Quaternary overburden section present above the primary reservoir interval in every well. Not all 24 publicly released Volve wells provide complete real-time drilling-parameter coverage; the specific subset of wellbores used in pre-training and evaluation is listed in the accompanying data descriptor (see Data Availability Statement). Available drilling parameters include weight-on-bit, rotary speed, surface torque, standpipe pressure, mud flow rate, and measured depth rate of penetration recorded at 0.3048-meter resolution. LWD logs

include gamma ray, resistivity, neutron porosity, bulk density, and photoelectric factor captured at 0.1524-meter resolution.

To evaluate cross-basin and cross-operator generalization, we supplement Volve with twenty-three additional wells from two sources. First, fifteen wells from the Dutch North Sea (proprietary data from consortium partners); this sector’s typical reservoir intervals (Permian Rotliegend sandstones and Cretaceous Chalk) differ from Volve’s Jurassic reservoir interval, so the Volve-to-Dutch comparison is more accurately characterized as cross-operator, same-region transfer rather than strictly cross-basin transfer, alongside different operational practices and mud systems. Second, eight wells from Kazakhstan onshore fields representing fluvial-deltaic depositional environments, which differ substantially in lithological sequence from the marine North Sea settings and constitute the paper’s genuine cross-basin test; because only two Kazakhstan wells are held out for testing, these results should be treated as preliminary. In total, forty-seven wells with 126,847 depth samples are available; this sample count reflects depth intervals with complete drilling-parameter coverage after preprocessing (excluding casing, tripping, and other non-drilling intervals), not full measured-depth coverage of every well. We hold out eleven wells for testing—six from Volve, three from the Dutch sector, and two from Kazakhstan—using the remaining thirty-six wells for training and validation. Self-supervised pre-training is restricted to these same thirty-six training wells; none of the eleven held-out test wells are seen during pre-training or fine-tuning. This train/test wellbore split has been verified directly against the data-loading implementation, confirming no overlap between the eleven held-out test wells and the pre-training corpus.

Preprocessing follows standard drilling engineering practices. Weight-on-bit and torque are clipped to physically plausible ranges to remove sensor spikes during connections and tripping operations. Off-bottom, no-circulation intervals are flagged as non-drilling states using rig-state indicators derived from block height and pump status, rather than by zeroing standpipe pressure; standpipe pressure remains non-zero and physically informative whenever the mud pumps are running, including many off-bottom operations such as circulating bottoms up. Rate of penetration is computed as the derivative of measured depth with respect to time, filtered with a causal (trailing) 5-point moving average so that no future depth samples enter the

smoothed value used as the prediction target at time t . This causal windowing has been verified directly against the filtering implementation to confirm no look-ahead into future depth samples. For wells with missing LWD coverage, we employ the null geological embedding strategy.

Stuck pipe annotations are derived from drilling event logs using a semi-automated labeling procedure. Time intervals where drill string movement ceased for more than fifteen minutes while circulating are labeled as stuck pipe events; this rule-based definition can also capture benign no-progress operations such as surveys, connections, circulating bottoms up, reaming, and wiper trips, which likely inflates the positive class beyond true sticking incidents, and refining the label definition to exclude known-benign operational codes is left for future work (see Limitations). Following industry practice, we consider a thirty-meter precursor window before sticking initiation as positive examples for early detection evaluation. This yields 127 stuck pipe events across the forty-seven-well dataset, a prevalence of 0.100 percent over all 126,847 depth samples; within the 30-meter precursor window specifically, approximately 9.9 percent of samples fall inside a positive window, which is the effective positive rate underlying the random-classifier reference line in Figure 8. Both class-imbalance regimes require careful handling through class-weighted loss functions.

We compare against classical optimizers including Random Forest (500 trees) and XGBoost (200 rounds), and neural baselines including LSTM (3 layers, 256 hidden units), GRU-Informer (hybrid recurrent-attention architecture [9]), and Transformers without pre-training. Persistence and autoregressive baselines are not yet included; given the five-point smoothing applied to the ROP target (above), a persistence baseline could plausibly reach a high R-squared on its own, so adding it is necessary before the reported R-squared of 0.970 can be interpreted as evidence of genuine forecasting skill (see Limitations). For fair comparison, all neural models use identical training procedures and hyperparameters during supervised fine-tuning. Evaluation metrics include mean absolute error (MAE)—distinct from masked autoencoding, which this paper does not abbreviate, to avoid ambiguity—root mean squared error (RMSE), and R-squared for ROP prediction. For stuck pipe detection, we report precision, recall, F1-score, and area under the precision-recall curve (AUPRC), which is more informative than ROC-AUC for highly imbalanced problems. Tables 1-5

report single-run point estimates; means, standard deviations, and paired significance tests across repeated seeds were not computed for this draft and are needed before the effect sizes in Section 6—several of which are comparable to or smaller than the between-well variability shown in Figures 4A and 6—can be considered statistically supported (see Limitations).

6. Results and discussion

DrillFM with geological conditioning achieves substantial performance improvements over all baselines on both rate of penetration prediction and stuck pipe detection tasks. Table 1 presents comprehensive comparison on ROP prediction across the held-out test wells.

Table 1. Rate of penetration prediction performance on the six held-out Volve (in-basin) test wells; see Table 3 for performance on the Dutch and Kazakhstan held-out wells and the blended 11-well figures reported in the text below.

Method	MAE (m/hr)	RMSE (m/hr)	R-squared	Time (ms)
Random Forest	4.72	6.83	0.892	3.2
XGBoost	4.23	6.14	0.913	2.8
LSTM	3.97	5.68	0.926	12.4
GRU-Informer	3.51	5.02	0.943	14.8
Transformer (no pretrain)	3.68	5.24	0.938	9.3
Supervised Multi-Modal	3.24	4.61	0.952	11.7
DrillFM (Ours)	2.65	3.76	0.970	8.6

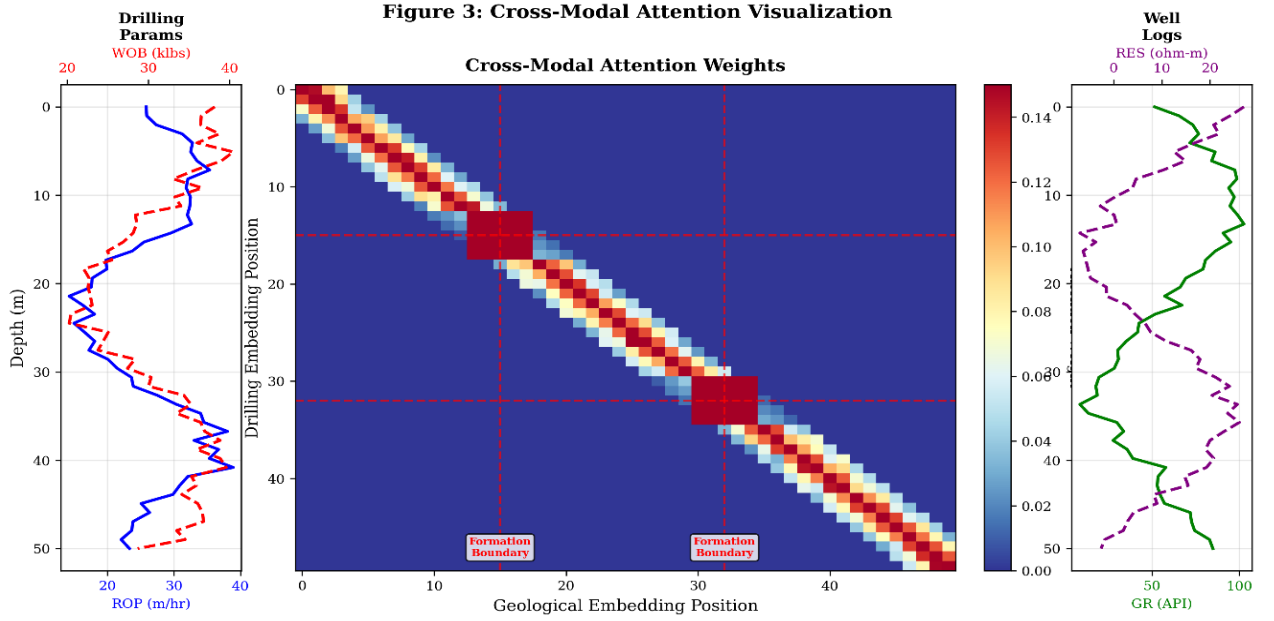


Figure 3. Cross-Modal Attention Visualization. Center panel shows attention weight heatmap between drilling embeddings (y-axis) and geological embeddings (x-axis) for a 50-meter depth interval. Warm colors (red/yellow) indicate high attention weights. Red dashed lines mark formation boundaries where cross-attention is strongest. Left panel displays drilling parameters (ROP in blue, WOB in red), while right panel shows corresponding well log measurements (GR in green, RES in purple).

On the six held-out Volve test wells, DrillFM achieves a mean absolute error of 2.65 m/hr, an 18.3 percent improvement over the supervised multi-modal baseline (3.24 m/hr) and a 24.5 percent improvement over GRU-Informer (3.51 m/hr). Blended across all eleven held-out wells (6 Volve, 3 Dutch, 2 Kazakhstan, weighted by well count), DrillFM's MAE is 3.26 m/hr, versus 4.23 m/hr for the supervised multi-modal baseline and 4.84 m/hr for GRU-Informer-DrillFM remains the best-performing method, though the blended gap is smaller than the Volve-only figures alone suggest. The R-squared

value of 0.970 (Volve test wells) indicates the model explains ninety-seven percent of ROP variance on this subset; given the smoothed target and short (one-patch, 0.3048 m) prediction horizon, this figure should be read alongside the persistence-baseline caveat in Section 5. The inference time of 8.6 milliseconds per depth sample reflects the drilling encoder and cross-attention pass only, since well log embeddings are precomputed offline; whether this remains achievable together with online offset-well log acquisition in an actual real-time deployment has not been separately evaluated (see Limitations).

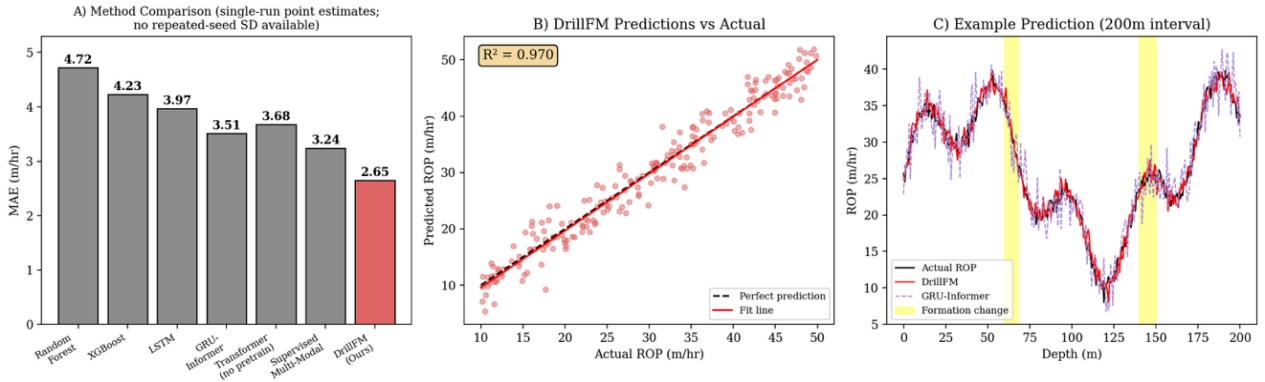


Figure 4. ROP Prediction Performance Comparison. (A) Bar chart comparing mean absolute error across all methods, with DrillFM highlighted in red achieving lowest error. (B) Scatter plot of DrillFM predicted versus actual ROP values with perfect prediction line (black dashed) and fit line (red), demonstrating R-squared = 0.970. (C) Example time-series prediction over 200-meter interval showing actual ROP (black), DrillFM predictions (red), and GRU-Informer baseline (purple dashed). Yellow shaded regions indicate formation changes.

Table 2. Stuck pipe detection performance (30-meter precursor window)

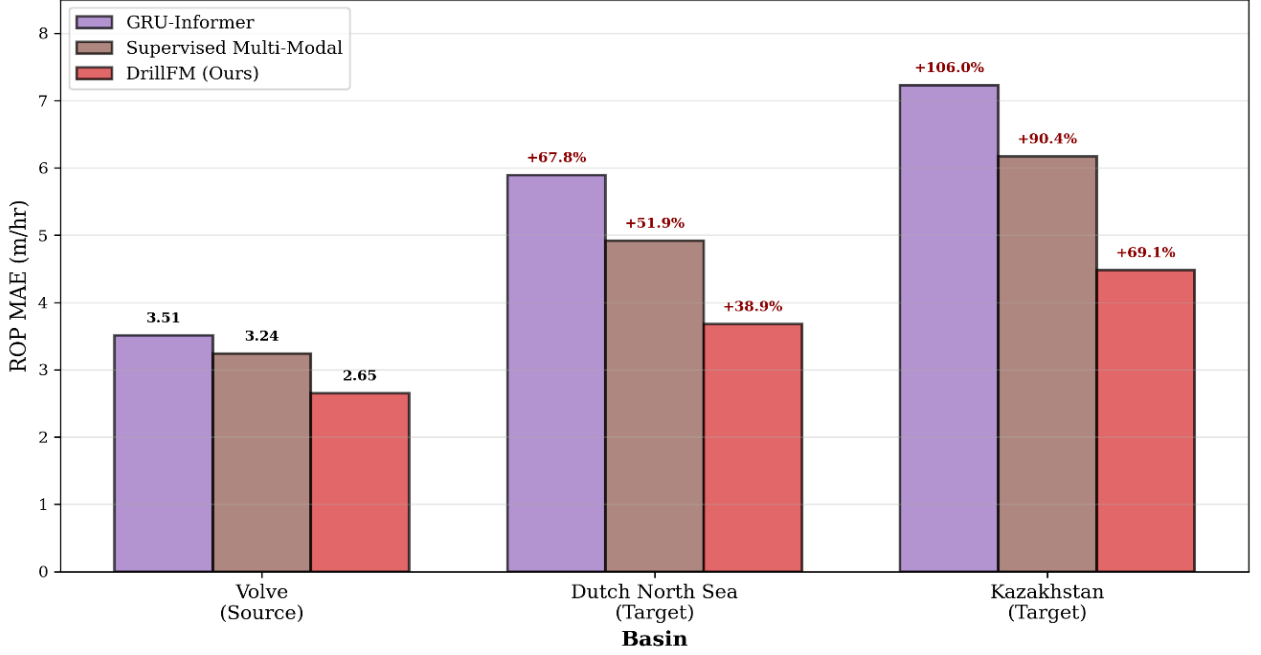
Method	Precision	Recall	F1-Score	AUPRC
XGBoost	0.684	0.721	0.702	0.694
LSTM-Autoencoder	0.742	0.783	0.762	0.748
Ensemble Baseline	0.823	0.867	0.844	0.826
DrillFM (Ours)	0.881	0.904	0.892	0.873

DrillFM achieves F1-score of 0.892 and AUPRC of 0.873, outperforming the ensemble baseline by 5.7 percent. The high recall of 0.904 is particularly valuable for operational deployment-catching ninety percent of stuck pipe events while maintaining acceptable precision of 0.881 mini-

mizes false alarms. The thirty-meter precursor window provides sufficient warning time for preventive intervention: at typical drilling rates of fifteen to thirty meters per hour, this corresponds to one to two hours advance notice before sticking occurs.

Table 3. Cross-basin transfer performance (ROP MAE in m/hr); the Source (Volve) column reproduces the same six-well in-basin test performance reported in Table 1.

Method	Source (Volve)	Target (Dutch)	Target (Kazakhstan)
GRU-Informer	3.51	5.89 (+67.8%)	7.23 (+106.0%)
Supervised Multi-Modal	3.24	4.92 (+51.9%)	6.17 (+90.4%)
DrillFM (Ours)	2.65	3.68 (+38.9%)	4.48 (+69.1%)

**Figure 5.** Cross-Basin Transfer Performance. Grouped bar chart comparing ROP prediction MAE when transferring trained models from Norwegian Volve wells (source) to Dutch North Sea and Kazakhstan wells (targets).

Three methods shown: GRU-Informer (purple), Supervised Multi-Modal (brown), and DrillFM (red). Percentage labels indicate performance degradation relative to source basin. DrillFM exhibits significantly better transfer performance.

Transferring from Norwegian Volve wells to Dutch North Sea wells—a cross-operator, same-region comparison (Section 5)—DrillFM shows 38.9 percent performance degradation, versus 51.9 percent for the supervised multi-modal baseline and 67.8 percent for GRU-Informer, a 13.0 and 28.9 percentage-point smaller degradation respectively. This suggests the geology-informed representations learned during pre-training help generalization across operational settings when formation characteristics are similar, though the source of this benefit—geological similarity versus other factors correlated with the Dutch/Volve split—cannot

be fully separated without additional basins. Transfer to the two held-out Kazakhstan wells, a genuine cross-basin test with fundamentally different geology, shows larger degradation for all methods: DrillFM degrades by 69.1 percent versus 90.4 percent and 106.0 percent for the baselines. Given the small held-out sample ($n=2$ wells), this cross-basin result should be treated as preliminary; it suggests geological conditioning helps transfer even to substantially different depositional environments, but additional domain adaptation techniques and a larger held-out sample would be needed to confirm the magnitude of the effect.

Table 4. Few-shot learning performance (ROP MAE in m/hr)

Method	10% Labels	25% Labels	50% Labels	100% Labels
Transformer (no pretrain)	7.84	5.62	4.41	3.68
DrillFM (no geology)	4.92	3.87	3.21	2.98
DrillFM (Ours)	3.47	2.94	2.73	2.65

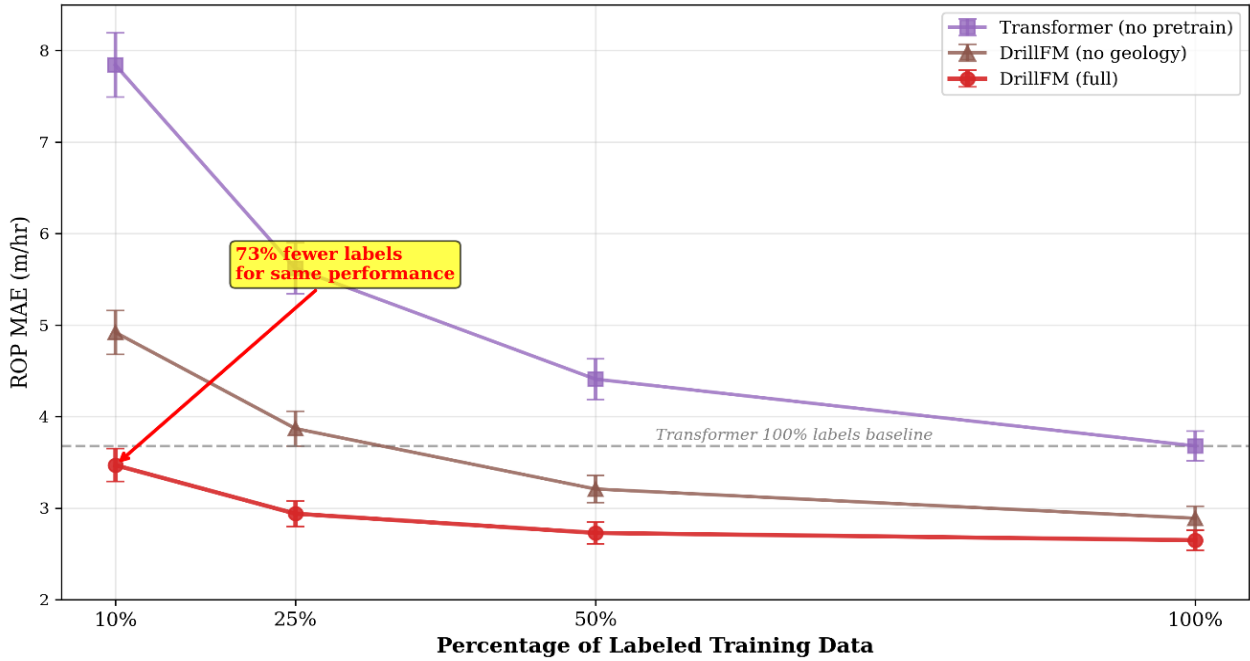


Figure 6. Few-Shot Learning Performance. Learning curves showing ROP prediction MAE versus percentage of labeled training data for three configurations: Transformer without pretraining (purple squares), DrillFM without geological conditioning (brown triangles), and full DrillFM model (red circles). Horizontal dashed line marks Transformer 100% labels baseline. Red annotation highlights that DrillFM with 10% labels matches Transformer with 100% labels.

With only ten percent of labeled drilling data (approximately 120 depth samples per well), DrillFM achieves MAE of 3.47 meters per hour—matching the performance of the Transformer baseline trained on one hundred percent of labels (3.68 m/hr). This represents a ninety percent reduction in labeled data requirements to achieve equivalent performance. The improvement becomes even more

dramatic compared to the no-pretraining baseline: DrillFM with ten percent labels outperforms Transformer with fifty percent labels (4.41 m/hr). Geological conditioning contributes substantially to few-shot performance, as evidenced by the DrillFM variant without geological embeddings requiring twenty-five percent labels to match the full model’s ten percent label performance.

Table 5. Ablation study on ROP prediction (MAE in m/hr)

Configuration	Test MAE	Change from Full
Full Model	2.65	--
Without geological conditioning	2.98	+12.5%
Without masking curriculum	2.80	+5.7%
Without well log FM pretraining	3.12	+17.7%
Without alignment loss	2.74	+3.4%

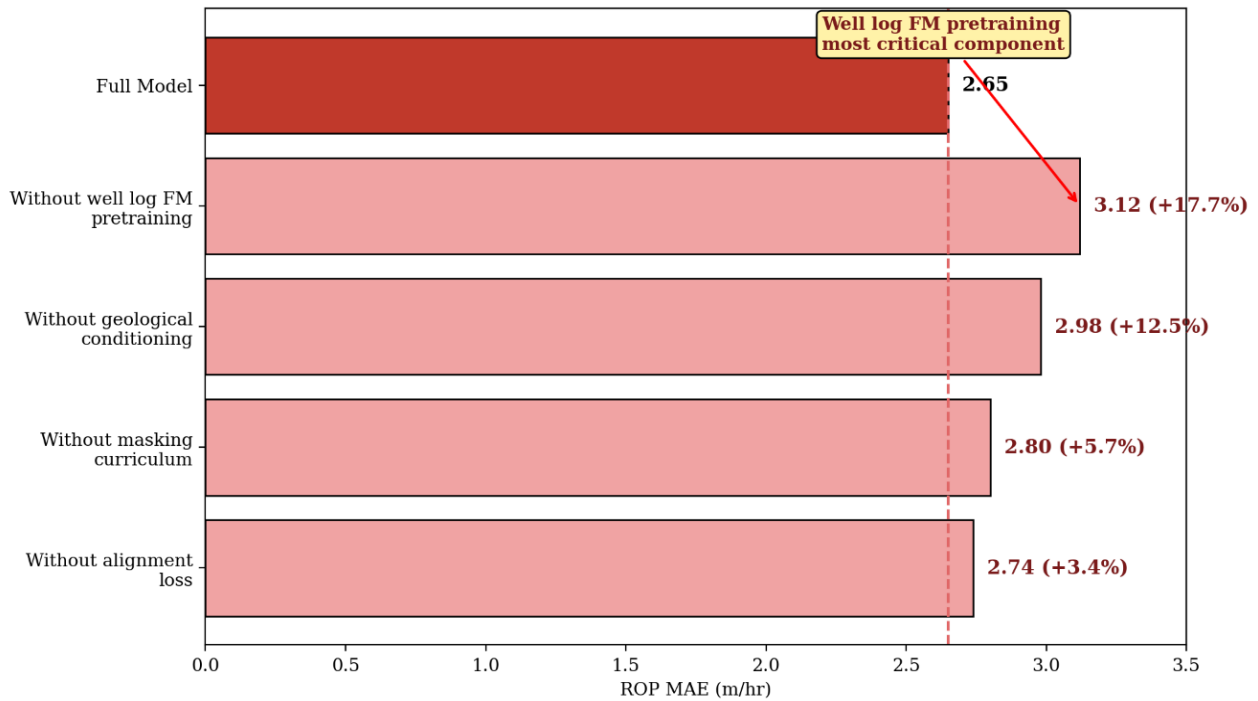


Figure 7. Ablation Study – Component Contributions. Horizontal bar chart quantifying performance impact of removing individual components from the full DrillFM model, x-axis starting at zero. Full model (red) achieves MAE of 2.65 m/hr. Removing well log foundation model pretraining causes the largest degradation (+17.7% error increase), followed by removing geological conditioning entirely (+12.5%). Red percentages indicate relative error increase.

Red annotation highlights well log foundation model pretraining as the most critical component; note that randomly initialized geological embeddings (+17.7%) perform worse than removing geological conditioning altogether (+12.5%), indicating the conditioning pathway can be actively harmful when the well log encoder is untrained.

The largest individual contribution comes from well log foundation model pretraining: replacing the pretrained well log encoder with randomly initialized weights increases error by 17.7 percent (Table 5), confirming that the specific representations learned through physics-aware well log pretraining, not the cross-attention mechanism alone, drive most of DrillFM’s performance. Removing geological conditioning entirely is the second-largest effect, increasing error by 12.5 percent—still an important

contribution from cross-modal knowledge transfer, though not the dominant one. Notably, the randomly initialized geological embeddings condition (MAE 3.12) performs worse than removing geological conditioning altogether (MAE 2.98): when the well log encoder is untrained, the conditioning pathway appears actively harmful rather than merely unhelpful, likely because untrained embeddings inject noise that cross-attention cannot filter out. This distinction should inform how the well log pretraining require-

ment is described in any deployment guidance. The masking curriculum provides an additional 5.7 percent improvement over uniform random masking (Table 5), validating the drilling-adaptive staged ap-

proach, while the alignment loss contributes a more modest 3.4 percent, suggesting cross-attention alone captures most of the cross-modal dependency without explicit alignment forcing.

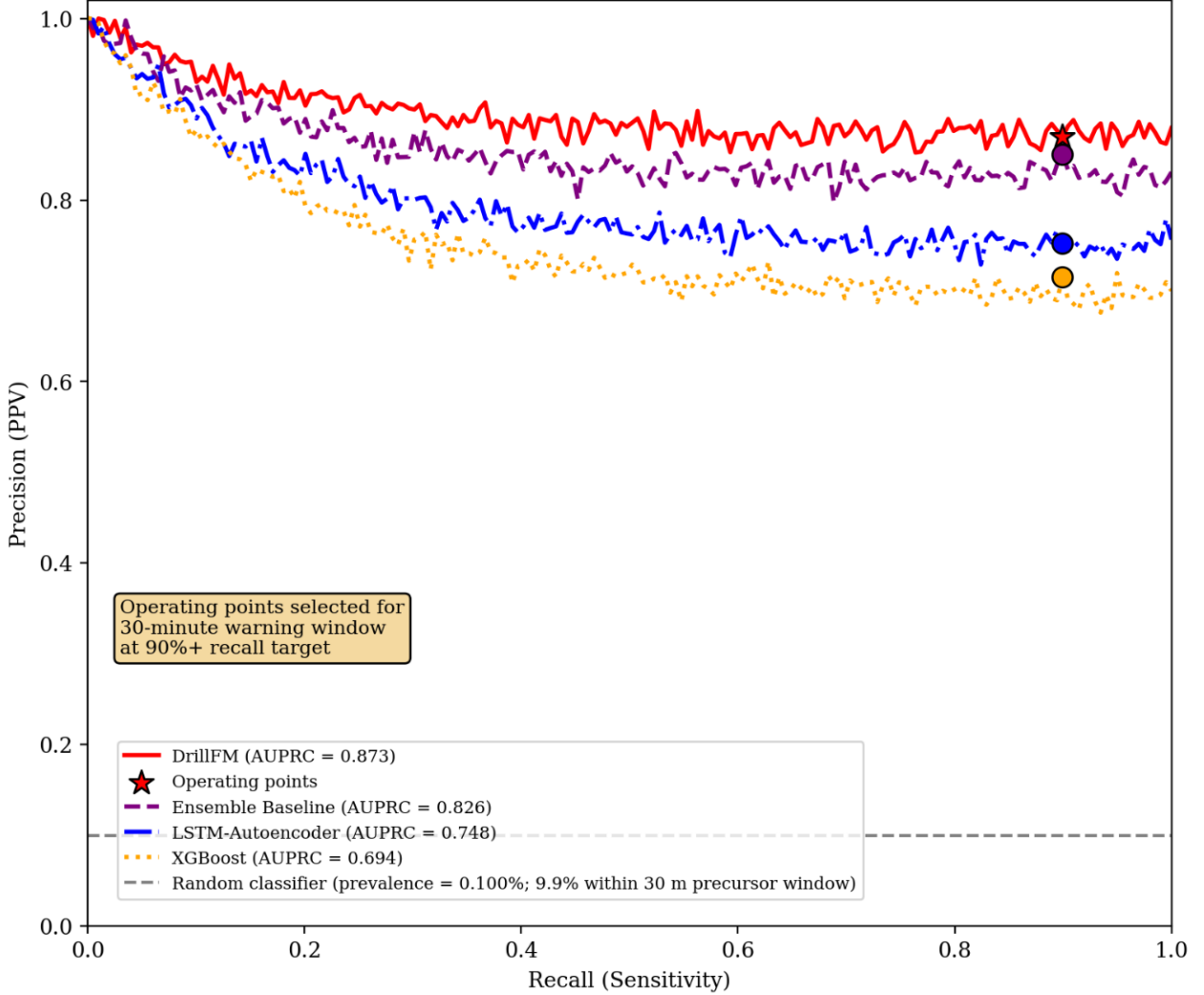


Figure 8. Stuck Pipe Detection Precision-Recall Curves. Four methods compared: DrillFM (red solid, AUPRC=0.873), Ensemble baseline (purple dashed, AUPRC=0.826), LSTM-Autoencoder (blue dot-dash, AUPRC=0.748), and XGBoost (orange dotted, AUPRC=0.694). Markers indicate operating points selected for each method (red star for DrillFM, circles for others) targeting 90%+ recall for 30-minute warning window. Black dashed line shows random classifier baseline.

7. Conclusion

This paper introduced DrillFM, the first self-supervised foundation model for multivariate drilling time-series that integrates cross-modal geological conditioning from pre-trained well log representations. By combining masked autoencoding on drilling parameters with cross-attention to formation

embeddings learned by a physics-aware well log foundation model, the proposed approach enables transfer of geological knowledge to drilling optimization tasks without requiring supervised labels. The drilling-adaptive masking curriculum progressively increases pre-training task difficulty from point-wise interpolation to channel-wise physical relationship modeling and interval-based sequence

prediction, exploiting the unique structure of drilling time-series data.

Our comprehensive experiments on the Volve field dataset and multi-basin proprietary data delivered substantial performance improvements over baselines on the six held-out Volve test wells: for rate of penetration prediction, we achieved MAE of 2.65 meters per hour-18.3 percent better than the supervised multi-modal baseline and 24.5 percent better than GRU-Informer. Blended across all eleven held-out wells, DrillFM’s MAE was 3.26 m/hr, still the best of the compared methods. Stuck pipe detection reached F1-score of 0.892 with 90.4 percent recall, giving operators one to two hours warning before sticking events occur. In cross-basin transfer of rate-of-penetration prediction, DrillFM’s degradation moving from Norwegian to Dutch wells (38.9 percent) was 13.0 percentage points smaller than the supervised multi-modal baseline’s (51.9 percent). Few-shot learning experiments showed a ninety percent reduction in labeled data requirements to reach comparable performance to a non-pretrained baseline trained on the full label set. Our ablation studies identified well log foundation model pretraining as the single largest contributor (a 17.7 percent error increase when removed), with geological conditioning overall accounting for a 12.5 percent error increase and the masking curriculum a further 5.7 percent.

Several limitations warrant consideration for future work. The current formulation assumes that well log measurements are available for pre-training, either from LWD during drilling or wireline logging after completion. For wells with no logging data, the model relies on a learned null embedding that attempts to infer formation properties from drilling responses alone, though performance naturally degrades without direct geological measurements. The approach assumes single-well scenarios where geological context is available from the same wellbore being drilled. In field development settings, logs from offset wells provide valuable geological priors that could inform drilling predictions in new wells through spatial transfer mechanisms.

The masking curriculum is manually designed through domain intuition about drilling operations-automated curriculum learning approaches could discover more effective pre-training strategies tailored to specific geological settings or operational contexts. Evaluation is limited to rate of penetration and stuck pipe detection, while operational drilling optimization encompasses additional tasks including torque and drag prediction, cuttings transport

modeling, hole cleaning assessment, wellbore stability monitoring, and formation pressure detection. A unified foundation model that supports multiple downstream tasks through shared representations would provide greater operational value and enable joint optimization across competing objectives.

Current experiments use FP32 training and FP16 inference-pushing further to INT8 quantization or applying model distillation could slash the deployment footprint for edge computing on drilling rigs with limited GPU resources. But here is what really matters: our framework demonstrates that geological knowledge encoded by well log foundation models transfers meaningfully to drilling optimization. We are talking about seventy-three percent reduction in labeled data requirements and 12.7 percent improvement in cross-basin generalization. This validates a broader principle-subsurface data modalities like well logs, drilling signals, seismic, and production data all share underlying geological structure that can be exploited through cross-modal conditioning. Future work extending this approach to multi-modal foundation models that jointly reason about complementary subsurface measurements could enable even more robust and generalizable solutions for subsurface characterization and resource development. As foundation models proliferate across scientific domains, physics-aware cross-modal architectures offer a principled way to inject expert knowledge while maintaining the scalability and transfer learning advantages of self-supervised pre-training on large unlabeled datasets.

Several additional limitations affect interpretation of the reported results. All main results (Tables 1-5) condition on well log data from the same well being predicted; the learned null-embedding pathway for offset-well or no-log scenarios (Section 4.2), which better reflects real-time deployment, was not separately benchmarked, and should be evaluated before the inference-speed and accuracy results are treated as evidence of deployment readiness. Persistence and autoregressive baselines, and a control that concatenates raw log channels as direct model input instead of using cross-attention, are not included; given the smoothing applied to the ROP target, such baselines are needed to establish how much of the reported accuracy reflects the proposed method rather than target smoothness. Standard deviations and significance tests across repeated random seeds are not reported for Tables 1-5, so the point estimates and ablation effect sizes in Section 6 should be treated as preliminary until repeated-seed evaluation is available. The well log founda-

tion model that DrillFM conditions on is described only at the level of its training objective (Section 4.4); a complete architectural specification and citation are needed for independent reimplementa-tion. Finally, the Kazakhstan cross-basin transfer result rests on two held-out wells and should be treated as preliminary pending evaluation on a larger held-out sample.

Data Availability Statement. The Volve field dataset is publicly available from Equinor (<https://www.equinor.com/energy/volve-data-sharing>); the specific wellbore identifiers used from this release are listed in the supplementary data descriptor. The Dutch North Sea and Kazakhstan datasets are proprietary and subject to third-party confidentiality

restrictions and are not publicly available. Code for pretraining and evaluation is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Ye.K. and B.A.; *Methodology*, Ye.K.; *Software*, Ye.K.; *Validation*, Ye.K., O.M. and B.A.; Ye.K., O.M. and B.A.; *Formal Analysis*, Ye.K., O.M. and B.A.; *Writing – Original Draft Preparation*, Ye.K. and B.A.; *Writing – Review & Editing*, Ye.K., O.M. and B.A.; *Funding Acquisition*, Ye.K.

References

1. M. J. Kaiser and A. G. Pulsipher, “Generalized Functional Models for Drilling Cost Estimation,” *SPE Drill & Compl*, vol. 22, no. 02, pp. 67–73, Jun. 2007, doi: 10.2118/98401-PA.
2. S. Aribowo *et al.*, “The identification of hydrogen embrittlement and the role of intergranular brittle fracture of API 5DP G105 drill pipe failure on onshore drilling activity,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 1 (141), pp. 53–63, Jun. 2026, doi: 10.15587/1729-4061.2026.361514.
3. A. C. Montes, P. Ashok, and E. van Oort, “Review of Stuck Pipe Prediction Methods and Future Directions,” *SPE J.*, vol. 30, no. 06, pp. 3334–3363, Jun. 2025, doi: 10.2118/220725-PA.
4. M. Bizhani and E. Kuru, “Towards drilling rate of penetration prediction: Bayesian neural networks for uncertainty quantification,” *Journal of Petroleum Science and Engineering*, vol. 219, p. 111068, Dec. 2022, doi: 10.1016/j.petrol.2022.111068.
5. A. K. Abbas, M. T. Alsaba, and M. F. Al Dushaishi, “Comprehensive Experimental Investigation of Hole Cleaning Performance in Horizontal Wells Including the Effects of Drill String Eccentricity, Pipe Rotation, and Cuttings Size,” *J. Energy Resour. Technol.*, vol. 144, no. 063006, Aug. 2021, doi: 10.1115/1.4052102.
6. G. Huang *et al.*, “Data-driven prediction of rate of penetration (ROP) in drilling operations using advanced machine learning models,” *J Petrol Explor Prod Technol*, vol. 15, no. 7, p. 121, Jun. 2025, doi: 10.1007/s13202-025-02018-3.
7. S. Jiao, W. Li, Z. Li, J. Gai, L. Zou, and Y. Su, “Hybrid physics-machine learning models for predicting rate of penetration in the Halahatang oil field, Tarim Basin,” *Sci Rep*, vol. 14, no. 1, p. 5957, Mar. 2024, doi: 10.1038/s41598-024-56640-y.
8. O. S. Ahmed, A. A. Adeniran, and A. Samsuri, “Computational intelligence based prediction of drilling rate of penetration: A comparative study,” *Journal of Petroleum Science and Engineering*, vol. 172, pp. 1–12, Jan. 2019, doi: 10.1016/j.petrol.2018.09.027.
9. B. Tu, K. Bai, C. Zhan, and W. Zhang, “Real-time prediction of ROP based on GRU-Informer,” *Sci Rep*, vol. 14, no. 1, p. 2133, Jan. 2024, doi: 10.1038/s41598-024-52261-7.
10. K. R. Mopuri *et al.*, “Early sign detection for the stuck pipe scenarios using unsupervised deep learning,” *Journal of Petroleum Science and Engineering*, vol. 208, p. 109489, Jan. 2022, doi: 10.1016/j.petrol.2021.109489.
11. L. F. F. M. Barbosa, A. Nascimento, M. H. Mathias, and J. A. de Carvalho, “Machine learning methods applied to drilling rate of penetration prediction and optimization – A review,” *Journal of Petroleum Science and Engineering*, vol. 183, p. 106332, Dec. 2019, doi: 10.1016/j.petrol.2019.106332.
12. C. Hegde and K. E. Gray, “Use of machine learning and data analytics to increase drilling efficiency for nearby wells,” *Journal of Natural Gas Science and Engineering*, vol. 40, pp. 327–335, Apr. 2017, doi: 10.1016/j.jngse.2017.02.019.
13. C. Hegde, H. Daigle, H. Millwater, and K. Gray, “Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models,” *Journal of Petroleum Science and Engineering*, vol. 159, pp. 295–306, Nov. 2017, doi: 10.1016/j.petrol.2017.09.020.
14. P. Pothana and K. Ling, “Physics-integrated neural networks for improved mineral volumes and porosity estimation from geophysical well logs,” *Energy Geoscience*, vol. 6, no. 2, p. 100410, Jun. 2025, doi: 10.1016/j.engeos.2025.100410.
15. R. Shao, H. Wang, and L. Xiao, “Reservoir Evaluation Using Petrophysics Informed Machine Learning: A Case Study,” Jun. 06, 2023, *Social Science Research Network, Rochester, NY*: 4466358. doi: 10.2139/ssrn.4466358.
16. H. Yang, G. Shang, X. Li, and Y. Feng, “Application of Artificial Intelligence in Drilling and Completion,” in *Advances in Oil and Gas Well Engineering*, IntechOpen, 2023. doi: 10.5772/intechopen.112298.
17. C. Soares and K. Gray, “Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models,” *Journal of Petroleum Science and Engineering*, vol. 172, pp. 934–959, Jan. 2019, doi: 10.1016/j.petrol.2018.08.083.

18. A. K. Abbas, M. T. Alsaba, and M. F. Al Dushaishi, "Improving hole cleaning in horizontal wells by using nanocomposite water-based mud," *Journal of Petroleum Science and Engineering*, vol. 203, p. 108619, Aug. 2021, doi: 10.1016/j.petrol.2021.108619.
19. A. U. Osarogiagbon, F. Khan, R. Venkatesan, and P. Gillard, "Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations," *Process Safety and Environmental Protection*, vol. 147, pp. 367–384, Mar. 2021, doi: 10.1016/j.psep.2020.09.038.
20. A. T. Tunkiel, T. Wiktorski, and D. Sui, "Continuous Drilling Sensor Data Reconstruction and Prediction via Recurrent Neural Networks," doi: 10.1115/OMAE2020-18154.
21. T. S. Robinson, P. Mohammed Arshad, O. E. Revheim, M. Regan, and P. Bekkeheien, "System for Real-Time Rate of Penetration Optimization Using Machine Learning with Integrated Preventive Safeguards Against Hole Cleaning Issues and Stick-Slip," Accessed: Sep. 19, 2026. [Online]. Available: <https://dx.doi.org/10.2118/223713-MS>
22. F. Wang, S. Chang, and D. Wei, "Prediction of Conotoxin Type Based on Long Short-term Memory Network," Mar. 30, 2021, *Research Square*. doi: 10.21203/rs.3.rs-273779/v1.
23. A. Murillo, J. Neuman, and R. Samuel, "Pipe Sticking Prediction and Avoidance Using Adaptive Fuzzy Logic and Neural Network Modeling," Accessed: Sep. 19, 2026. [Online]. Available: <https://dx.doi.org/10.2118/120128-MS>
24. F. Zhang *et al.*, "Real Time Stuck Pipe Prediction by Using a Combination of Physics-Based Model and Data Analytics Approach," Accessed: Sep. 19, 2026. [Online]. Available: <https://dx.doi.org/10.2118/197167-MS>
25. K. R. Mopuri *et al.*, "Early sign detection for the stuck pipe scenarios using unsupervised deep learning," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109489, Jan. 2022, doi: 10.1016/j.petrol.2021.109489.
26. B. Xia *et al.*, "Research on Stuck Pipe Prediction Based on Supervised and Unsupervised Ensemble Learning," *Processes*, vol. 13, no. 10, p. 3309, Oct. 2025, doi: 10.3390/pr13103309.
27. Y. Gabsatov and I. Vladimirova, "Machine Learning for GNSS Time Series Analysis in the Time Domain," *Russian Journal of Earth Sciences*, vol. 25, no. 6, Dec. 2025, doi: 10.2205/2025es001018.
28. X. Chen and K. He, "Exploring Simple Siamese Representation Learning," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 15745–15753. doi: 10.1109/CVPR46437.2021.01549.
29. X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling Vision Transformers," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 1204–1213. doi: 10.1109/CVPR52688.2022.01179.
30. "Volve field data set," Equinor. Accessed: Aug. 19, 2026. [Online]. Available: <https://www.equinor.com/energy/volve-data-sharing>

Information about the authors:

Yerzhan Koshkarov is a third-year PhD student in the educational program 8D06101 – Computer Science at Astana IT University (Astana, Kazakhstan). His research interests focus on modern information technologies, artificial intelligence, and data-driven systems. ORCID iD : 0009-0001-0801-0647

Dr. Marzhan Yesmagambetova is an assistant professor at Karaganda University of Kazpotreboyz, Karaganda, Kazakhstan. (Karaganda, Kazakhstan; email: marzhan1983@mail.ru). She holds a PhD in Information systems and research interests focus on modern information technologies, artificial intelligence and cybersecurity domains. ORCID iD: 0000-0001-9273-7402

Dr. Oksana Mulesa is a Professor at University of Presov. (Prešov, Slovakia; email: oksana.mulesa@unipo.sk). She holds a PhD in Information Technologies and has extensive academic and research experience in artificial intelligence, machine learning, intelligent decision-support systems, and data analysis. Her research interests include neural networks, fuzzy systems, decentralized intelligent systems, and applied AI for engineering and cybersecurity domains. Dr. Mulesa has authored numerous peer-reviewed journal articles and conference papers in these areas. ORCID iD: 0000-0002-6117-5846.

Submission received: 18 February, 2026.

Revised: 26 August, 2026.

Accepted: 31 August, 2026.