# Real-time Anomaly Detection for Downhole Tool Failure Using Edge AI

Volodymyr Hnatushenko[1,†], Oleksandr Pashchenko[1,*,†], Volodymyr Khomenko[1,†] and Boranbay Ratov[2,†]

[1] Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, 49005, Dnipro, Ukraine
[2] Kazakh National Research Technical University named after K.I. Satpayev, 22 Satbaev str., 050013, Almaty, The Republic of Kazakhstan

### Abstract
Downhole tool failures, including the breakage of polycrystalline diamond compact bits, the stalling of drilling motors, and the malfunction of measurement-while-drilling equipment, remain a primary cause of non-productive time, costly equipment damage, and heightened safety risks in both oil and gas drilling and water well operations. Traditional reactive maintenance strategies, which rely on post-failure analysis and simple threshold-based alarms on parameters such as weight on bit or vibration magnitude, have proven inadequate for anticipating sudden tool breakdowns, as they fail to capture complex precursor patterns and often produce an excessive number of false alarms. This paper proposes an edge artificial intelligence framework designed for real-time anomaly detection, continuously monitoring multivariate sensor streams that include triaxial vibration, torque, weight on bit, and rotational speed directly at the rig site. The methodology combines an unsupervised long short-term memory autoencoder, which learns the normal dynamics of drilling and flags deviations through reconstruction error, with a lightweight gradient boosting classifier that validates potential anomalies and reduces false positives. This hybrid model is quantized and deployed on an edge computing device (NVIDIA Jetson Orin Nano), enabling inference latencies below 30 milliseconds without relying on cloud connectivity. Experimental evaluation using field data from twelve wells demonstrates that the system detects pre-failure anomalies 15 to 30 seconds before critical tool failure, achieving 92% precision and reducing false alarms by 65% compared to conventional threshold methods. Consequently, edge-based predictive maintenance not only minimises downtime and fishing costs but also significantly enhances the operational resilience of drilling rigs as critical energy and water infrastructure.

### Keywords
real-time anomaly detection, downhole tool failure, edge artificial intelligence, LSTM autoencoder



## 1. Introduction

Downhole tools play an absolutely critical role in all forms of drilling operations, including water well drilling, oil and gas exploration, and geothermal energy extraction. These tools, such as polycrystalline diamond compact bits, positive displacement motors, rotary steerable systems, and measurement-while-drilling devices, are directly responsible for rock destruction, borehole trajectory control, and the acquisition of subsurface data. Their reliable performance is therefore a prerequisite for any successful drilling campaign. However, these tools operate under extreme conditions: high mechanical loads, abrasive rock contact, severe vibrations, and elevated temperatures [1, 2]. Consequently, several common failure modes are frequently encountered. Bit wear, including gradual dulling and catastrophic fracture of diamond inserts, is the most widespread problem [3]. Bearing failure in roller-cone bits and downhole motors leads to excessive radial play and eventual seizure. Stuck pipe incidents, caused by poor hole cleaning or differential sticking, can

[1] RS-2026: Resilient Systems: Secure Digital Technologies and Critical Infrastructure, June 30, 2026, Drohobych, Ukraine
* Corresponding author.
† These authors contributed equally.
hnatushenko.v.v@nmu.one (V. Hnatushenko); pashchenko.o.a@nmu.one (O. Pashchenko); homenko.v.l@nmu.one (V. Khomenko); b.ratov@satbayev.university (B. Ratov)
0000-0003-3140-3788 (V. Hnatushenko); 0000-0003-3296-996X (O. Pashchenko); 0000-0002-3607-5106 (V. Khomenko); 0000-0003-4707-3322 (B. Ratov)

result in complete loss of downhole equipment [4, 5]. Motor stall, occurring when the required torque exceeds the motor's maximum capacity, may permanently damage the stator elastomer [6]. The economic and safety consequences are substantial: unplanned trips to replace failed tools account for ten to twenty percent of total rig time, translating into millions of dollars in lost revenue per rig per year. Beyond direct costs, sudden failures can trigger blowouts or well abandonment, with corresponding environmental liabilities.

Despite these high stakes, the vast majority of drilling operations still rely on reactive maintenance. Under this paradigm, tools are run until failure is detected, typically through threshold-based alarms on individual parameters such as weight on bit, vibration amplitude, standpipe pressure, or torque [7, 8]. While easy to implement, these systems suffer from severe limitations. They generate an unacceptably high number of false alarms because normal operations (e.g., penetrating a hard stringer) can temporarily exceed any fixed threshold, leading to alarm fatigue. More fundamentally, threshold methods cannot detect precursor patterns that unfold over multiple parameters and time scales [9, 10]. A failing bearing may produce subtle spectral changes long before root-mean-square vibration exceeds a threshold; gradual bit wear manifests as a slow drift in the torque-to-weight-on-bit relationship. Thus, reactive alarms typically sound only after irreversible damage has occurred, forcing crews into unplanned trips that better monitoring could have avoided.

In response, the drilling industry has begun shifting toward predictive maintenance (PdM), which aims to forecast remaining useful life and detect pre-failure anomalies with sufficient lead time for proactive intervention [11, 12]. PdM relies on data-driven approaches, particularly machine learning, to learn signatures of normal and degraded behaviour from historical sensor logs. Early work used support vector machines and random forests on hand-crafted features. More recently, deep learning architectures such as long short-term memory (LSTM) networks and autoencoders have been applied to model temporal dependencies and perform unsupervised anomaly detection. A particularly promising development is the integration of AI with edge computing, where models run directly on rig-side hardware rather than in the cloud [13, 14]. Edge AI eliminates latency and reliability issues associated with transmitting high-frequency data over satellite or cellular links – often unavailable or unstable in remote desert, offshore, or arctic locations. It also enables real-time inference with latencies measured in milliseconds, supporting closed-loop advisory systems [15, 16]. Nevertheless, most published work remains confined to post-mortem or cloud-based offline processing; very few studies report on hybrid unsupervised-supervised edge AI systems validated under authentic field conditions.

Beyond drilling applications, intelligent monitoring and predictive analytics have already proven their value in other critical infrastructure sectors, including mine drainage systems and electric power management. The successful deployment of machine-learning-based forecasting and decision-support tools in these domains demonstrates that edge-enabled AI can improve operational resilience, reduce unplanned downtime, and support timely intervention under safety-critical conditions [17, 18]. These findings provide additional motivation for extending similar approaches to real-time downhole tool failure detection.

This paper addresses that gap by presenting, for the first time, a hybrid edge AI framework for real-time anomaly detection of downhole tool failures. The framework integrates an unsupervised LSTM-autoencoder with a lightweight supervised gradient boosting classifier. The autoencoder learns the normal operational manifold of a well and generates a reconstruction error as an anomaly score; the gradient boosting model then validates whether high-error windows correspond to genuine pre-failure conditions rather than benign transitions. The entire model is quantised to int8 precision and pruned to run on an NVIDIA Jetson Orin Nano with inference times below 30 milliseconds. To our knowledge, this is the first deployment of a hybrid unsupervised-supervised edge AI system for downhole tool failure detection that operates in real time on rig-side hardware and has been validated on actual field data. Results show that the approach detects anomalies 15–30 seconds before critical failures, achieves 92% precision, and reduces false alarms by 65% relative to conventional threshold methods, enabling a genuine transition from reactive to predictive maintenance for critical drilling infrastructure.

## 2. Relevance of Research

The proposed edge-based anomaly detection framework aligns directly with several core themes of the RS-2026 Workshop. First, it constitutes an intelligent decision support system for critical infrastructure protection: real-time alerts provide drilling supervisors with actionable, interpretable warnings that enable proactive intervention before catastrophic failure. The system augments human judgment by filtering out the overwhelming majority of false alarms and presenting only high-confidence anomalies [19, 20, 21]. Second, the work lies squarely within the workshop's focus on machine learning and AI for anomaly detection, as the hybrid autoencoder-classifier architecture represents a state-of-the-art approach to detecting subtle, multi-dimensional failure precursors. Third, the integration of AI into real-time systems for managing critical facilities is embodied by the edge deployment strategy: the model runs on a ruggedised Jetson device at the rig site, ingesting streaming sensor data at 5 Hz and outputting anomaly scores and classifications with sub-30-ms latency. This real-time capability transforms passive monitoring into an active decision support tool that can, in future iterations, be coupled with automated drilling control systems for closed-loop corrective actions. Fourth, the research directly addresses the resilience and security of critical infrastructure by preventing downhole tool failures that could escalate into blowouts, loss of well control, or uncontrolled release of formation fluids. Averted failures also reduce the need for risky fishing operations. Thus, the proposed system improves operational efficiency while materially enhancing the safety and security of drilling rigs as critical energy and water infrastructure.

Drilling rigs are critical infrastructure whose reliable operation is essential for energy security, water supply, and economic stability [22, 23]. In the Mangystau Peninsula of Kazakhstan – a semi-arid region where water supply wells are the primary source of potable and agricultural water – any unplanned downtime due to downhole tool failure directly delays well completion or rehabilitation, exacerbating water scarcity and threatening public health [18, 24]. Similarly, uranium in-situ recovery operations rely on precisely drilled injection and production wells; a stuck pipe or lost bit can render an entire well cluster inoperable, disrupting uranium production and nuclear fuel supply chains. Industry estimates indicate that ten to twenty percent of rig time is consumed by unscheduled trips for tool replacement [25, 26]. The proposed edge AI system mitigates this by providing 15–30 seconds of lead time before catastrophic failure – sufficient for the driller to reduce weight on bit, adjust rotary speed, or initiate a controlled pull-out. Moreover, the edge-based architecture is particularly relevant for remote desert and offshore rigs, where satellite or cellular connectivity is often intermittent or unavailable. By performing all inference onboard, the system eliminates dependency on cloud connectivity, ensuring anomaly detection continues during network outages. This autonomy is a decisive practical advantage over cloud-based alternatives [27, 28].

While anomaly detection for drilling has been explored, the present work introduces several novel elements. The most significant novelty is the deployment paradigm: most existing machine learning approaches operate in batch or cloud-based post-processing mode; this study is among the first to implement and validate a complete edge AI pipeline for downhole tool failure detection on authentic field data with all inference executed on rig-side hardware. A second novelty is the hybrid architecture: pure unsupervised autoencoders produce many false alarms, while purely supervised methods require large labelled datasets of rare failures. The proposed two-stage approach leverages the autoencoder for sensitive anomaly scoring without labels and the classifier to filter spurious anomalies, achieving a superior balance of sensitivity and precision. A third novelty is the use of synthetic data augmentation (SMOTE-ENN) to address extreme class imbalance, substantially improving recall on rare failure events without overfitting. Finally, the paper provides a detailed case study of deploying quantised int8 models on an NVIDIA Jetson Orin Nano, including latency, power, and robustness measurements – engineering insights rarely reported in academic studies. Collectively, these novelties advance the state of the art from cloud-based, post-hoc analysis to real-time, rig-side predictive maintenance, marking a genuine transition from reactive to proactive failure management.

## 3. Statement of the Problem

Let a drilling operation be monitored by a set of surface and downhole sensors that produce a multivariate time series of drilling parameters. At any discrete time index $t$ (with $t = 0,1,2, \dots$), the observed state vector is defined as:

$$\mathbf{X}(t) = [\mathrm{WOB}(t), \mathrm{RPM}(t), \tau(t), \mathbf{v}(t), Q(t), P_{\mathrm{sp}}(t)]^{\top} \in \mathbb{R}^{d}, \quad (1)$$

where the individual components are: $\mathrm{WOB}(t)$ the weight on bit (in kilonewtons), $\mathrm{RPM}(t)$ the rotary speed (revolutions per minute), $\tau(t)$ the surface torque (in kilonewton-meters), $\mathbf{v}(t) = [v_x(t), v_y(t), v_z(t)]^{\top}$ the triaxial vibration acceleration (in meters per second squared), $Q(t)$ the flow rate of drilling fluid (in liters per second), and $P_{\mathrm{sp}}(t)$ the standpipe pressure (in megapascals).

The dimensionality of the observation space is therefore $d = 1 + 1 + 1 + 3 + 1 + 1 = 8$ when using triaxial vibration. In practice, additional derived features such as the rate of penetration or differential pressure may also be included, but the core set listed above is considered minimal for reliable failure detection. The sampling frequency is assumed to be $f_s = 10$ Hz, giving a sampling interval $\Delta t_s = 0.1$ seconds. Hence, at each time step $k$ (using $k$ to avoid confusion with the continuous time variable), we have $\mathbf{X}_k = \mathbf{X}(k \cdot \Delta t_s)$.

The primary objective is to detect whether the downhole tool (for example, a PDC bit, a roller-cone bit, a positive displacement motor, or a rotary steerable system) is in a pre-failure anomalous state at a future time $t + \Delta t$, where the prediction horizon $\Delta t$ lies between 5 and 30 seconds. More formally, we seek to compute a binary label $y(t)$ defined as:

$$y(t) = \begin{cases} 1 & \text{if the tool will experience irreversible damage within } [t + \Delta t_{\min}, t + \Delta t_{\max}] \text{ seconds,} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

with $\Delta t_{\min} = 5$ s and $\Delta t_{\max} = 30$ s. The lower bound ensures that the warning arrives early enough to be actionable, while the upper bound limits the prediction horizon to a physically meaningful window beyond which precursors become indistinguishable from normal operational variability.

A key conceptual element is the definition of an anomaly event itself. For the purposes of this work, an *anomaly* is any deviation of the tool's condition from its normal operating envelope that, if left uncorrected, will lead to irreversible damage within the next 60 seconds. Let $\mathcal{S}_{\mathrm{normal}}$ denote the set of all state trajectories that are consistent with healthy tool operation. Then at time $t$, the tool is considered to be in a pre-failure anomalous state if:

$$\exists\, \delta \in [0,60] \text{ ssuch thatdamage(tool at } t + \delta) > D_{\mathrm{crit}}, \quad (3)$$

where $D_{\mathrm{crit}}$ is a critical damage threshold beyond which the tool cannot recover and must be replaced. The damage function is not directly observable; instead, it must be inferred from the sensor time series $\mathbf{X}(t)$. Consequently, the anomaly detection task reduces to learning a mapping $\mathcal{F}$ from a window of recent observations to a binary decision:

$$\hat{y}(t) = \mathcal{F}(\mathbf{X}_{t-W+1}, \mathbf{X}_{t-W+2}, \dots, \mathbf{X}_t), \quad (4)$$

where $W$ is the length of the sliding window (in time steps). For a sampling rate of 10 Hz and a desired window duration of 5–10 seconds, a typical choice is $W = 60$ (6 seconds) or $W = 100$ (10 seconds). The function $\mathcal{F}$ must be designed to achieve high precision and recall while respecting strict real-time constraints.

The problem formulation, although mathematically crisp, conceals several profound difficulties that arise in real drilling environments. The first and perhaps most vexing challenge is the extreme class imbalance. Under normal operating conditions, downhole tools function without incident for hundreds of hours. Failure events, defined as irreversible damage requiring a trip, typically occur less than one percent of the total drilling time, and often much less. Let $N_{\text{total}}$ be the total number of time windows in the dataset, and $N_{\text{anomaly}}$ the number of windows that contain a precursor to a failure. Then the imbalance ratio is:

$$IR = \frac{N_{\text{normal}}}{N_{\text{anomaly}}} = \frac{N_{\text{total}} - N_{\text{anomaly}}}{N_{\text{anomaly}}} \gg 100. \tag{5}$$

Standard machine learning classifiers trained on such imbalanced data tend to achieve high accuracy by simply predicting the majority class (normal) for every window, while completely missing the rare but critical anomalies. Specialised techniques such as oversampling, synthetic data generation, and cost-sensitive learning are required, but each introduces its own complications.

The second major challenge is the combination of noise and non-stationarity. The observed drilling parameters are corrupted by measurement noise, electromagnetic interference, and mechanical vibrations that are not directly related to tool health. A reasonable model is:

$$\mathbf{X}_{\text{obs}}(t) = \mathbf{X}_{\text{true}}(t) + \eta(t), \tag{6}$$

where $\eta(t)$ is a zero-mean noise process with covariance matrix $\Sigma_\eta$. Moreover, the underlying normal behaviour $\mathbf{X}_{\text{true}}(t)$ is not stationary because the drill bit penetrates different rock formations with varying compressive strength, abrasiveness, and lithology.

For example, when transitioning from a soft shale to a hard sandstone, the weight on bit, torque, and rate of penetration all change systematically, even with a perfectly healthy tool. These formation-induced changes can easily be mistaken for anomaly precursors if the detection algorithm does not account for the non-stationary context. Formally, the conditional distribution $p(\mathbf{X}(t) \mid \text{healthy,formation})$ depends on the unknown formation type, which itself changes over time.

The third challenge is the stringent real-time constraint. The inference pipeline must produce an anomaly decision within a latency budget that is significantly smaller than the sampling interval multiplied by the window length. If we denote the time required to acquire a full window of $W$ samples as , and the computation time on the edge device as $T_{\text{comp}}$, then the end-to-end latency from the last sample to the output decision is $T_{\text{comp}} + T_{\text{post}}$, where $T_{\text{post}}$ includes any post-processing. For the system to be useful for proactive intervention, we require:

$$T_{\text{comp}} + T_{\text{post}} \ll \Delta t_{\min} = 5 \text{ s}. \tag{7}$$

A more realistic and demanding requirement is to achieve $T_{\text{comp}} < 50$ ms, which ensures that the decision is available almost immediately after the window is filled. This forces the use of highly optimised, low-complexity models. Deep neural networks with millions of parameters, while potentially accurate, are infeasible on edge hardware such as a Raspberry Pi or even an NVIDIA Jetson series if not aggressively pruned and quantised.

The fourth challenge is interpretability. Drilling supervisors and engineers are understandably reluctant to act on a warning that is not accompanied by an explanation. If the system outputs a binary flag indicating an impending failure, the operator will naturally ask: which sensor or combination of sensors triggered the alarm? Is it a single extreme value or a subtle pattern over time? Formally, for a given prediction $\hat{y}(t) = 1$, we would like to compute an attribution vector $\mathbf{a}(t) \in \mathbb{R}^d$ such that the $i$-th component indicates the contribution of the $i$-th sensor to the decision. A common approach is to use Shapley additive explanations (SHAP), where:

$$\phi_i = \sum_{S \subseteq \{1,\dots,d\} \setminus \{i\}} \frac{|S|!\,(d-|S|-1)!}{d!} [f(S \cup \{i\}) - f(S)], \tag{8}$$

with $f(S)$ being the model's prediction when only the sensors in subset $S$ are used.

Computing exact Shapley values is combinatorially expensive, but approximate methods exist. The challenge is to provide these explanations in real time without introducing additional latency that violates the 50 ms budget.

Several classes of methods have been proposed in the literature for downhole failure prediction, but each suffers from fundamental limitations that make them unsuitable for the problem defined above. The simplest and most widely deployed approach in industry is fixed-threshold alarming. For each parameter $\theta \in \{\text{WOB,RPM}, \tau, \| \mathbf{v} \|_2, Q, P_{\text{sp}}\}$, an upper bound $\theta_{\max}$ and a lower bound $\theta_{\min}$ are defined, and an alarm is raised if any parameter exceeds its bounds. This can be expressed as:

$$\text{alarm}(t) = \bigvee_{\theta} (\theta(t) > \theta_{\max} \text{ or } \theta(t) < \theta_{\min}). \tag{9}$$

While computationally trivial and easily understood, this method fails dramatically in variable lithology. When the drill bit enters a harder formation, the weight on bit and torque naturally increase to maintain the same rate of penetration; if the thresholds were set for soft rock, they will be exceeded continuously, causing a flood of false alarms. Conversely, if thresholds are set conservatively to accommodate the hardest expected formation, then true anomalies in softer formations will be missed. In other words, fixed thresholds cannot adapt to the time-varying mean and variance of the parameters.

Traditional machine learning methods, such as isolation forests or one-class support vector machines, offer a more sophisticated alternative because they can learn the normal data distribution without requiring labelled failures. An isolation forest constructs an ensemble of random binary trees, where anomalies are identified as points that require fewer splits to isolate. The anomaly score for a point $\mathbf{x}$ is:

$$s(\mathbf{x}, n) = 2^{-\frac{\mathbb{E}[h(\mathbf{x})]}{c(n)}}, \tag{10}$$

where $h(\mathbf{x})$ is the path length in an isolation tree and $c(n)$ is the average path length for a dataset of size $n$.

Similarly, a one-class SVM learns a hypersphere that encloses the normal data, minimizing:

$$\min_{R,\xi,\mathbf{c}} R^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i \text{ subject to } \| \phi(\mathbf{x}_i) - \mathbf{c} \|_2^2 \leq R^2 + \xi_i,$$

$$0, \tag{11}$$

where $\phi(\cdot)$ is a kernel mapping. However, both methods treat each time window as an independent point and completely ignore the temporal dependencies that are crucial for detecting precursors. A failure does not occur instantly; it evolves over tens of seconds or minutes, and the pattern of change across consecutive windows is more informative than any single window's static features. Because isolation forests and one-class SVMs have no memory, they cannot distinguish between a benign transient (e.g., a momentary spike in vibration due to a drill string bounce) and the beginning of a sustained degradation trend.

Finally, cloud-based artificial intelligence systems attempt to overcome these limitations by transmitting the sensor data to a remote server, where powerful deep learning models such as LSTMs or transformers can be deployed. The typical architecture involves sending batches of windows every few seconds, processing them on a GPU cluster in the cloud, and returning the predictions to the rig. The total latency can be expressed as:

$$T_{\text{cloud}} = T_{\text{upload}} + T_{\text{queue}} + T_{\text{compute}} + T_{\text{download}}, \tag{12}$$

where $T_{\text{upload}}$ is the time to transmit the sensor data over a satellite or cellular link (often 1–5 seconds for a 100-window batch), $T_{\text{queue}}$ accounts for server load (unpredictable, can exceed 10 seconds), $T_{\text{compute}}$ is the actual inference time (typically 10–100 ms), and $T_{\text{download}}$ returns the result (another 1–2 seconds).

The sum is rarely below 5 seconds and often exceeds 20 seconds, making the system useless for the required 5- to 30-second prediction horizon. Moreover, in remote drilling locations such as the desert of western Kazakhstan or offshore platforms, connectivity may be lost entirely for hours. When the network fails, the cloud-based system provides no information at all. For critical infrastructure where every second counts, this unreliability is unacceptable. The proposed edge AI solution, by contrast, keeps all computation on the rig, eliminates the dependency on connectivity, and guarantees a deterministic latency below 50 milliseconds, thereby overcoming each of the limitations discussed above.

## 4. Research Results

The experimental validation relies on authentic field logs acquired from twelve wells drilled in Kazakhstan. The well inventory consists of seven uranium in-situ recovery wells located in the southern region and five water supply boreholes on the Mangystau Peninsula, the latter being particularly relevant for rural water infrastructure. Each rig was outfitted with a standardised sensor suite: a triaxial accelerometer fixed to the top drive housing, a torque sub positioned directly above the bit, a weight-on-bit sensor derived from the hookload measurement system (Martin-Decker, 1 % accuracy), an RPM encoder (Banner, 0.1 % accuracy) coupled to the rotary table, and pressure transducers on the standpipe and discharge lines. All sensors were sampled at a native rate of 10 Hz.

To provide a tangible illustration of the precursor patterns captured by the proposed system, Figure 1 presents a 30-second excerpt of real x-axis vibration and surface torque signals from Well #5, recorded immediately before a complete bearing failure. A gradual escalation in vibration amplitude from approximately 0.2 g to over 0.8 g is observable over the 20-second interval preceding the event, accompanied by a marked increase in torque fluctuations from ±2 kN·m to ±8 kN·m. While these trends remain obscured by conventional fixed thresholds until the final seconds, the LSTM-autoencoder detects the growing reconstruction error starting around 25 seconds before the event, crossing the adaptive threshold (indicated by the dashed green vertical line) 8 seconds prior to the irreversible damage (solid black vertical line). This explicit visualisation substantiates the system's capacity to uncover subtle, multi-sensor precursors that elude reactive alarm strategies.

The total duration of recorded active drilling, after excluding connections and tripping intervals, is approximately 625 hours, corresponding to roughly 22.5 million raw samples at the native 10 Hz rate. The signals were subsequently downsampled to 5 Hz via block averaging (non-overlapping pairs of consecutive samples), yielding 11.25 million samples, which were then segmented into non-overlapping sliding windows of 12 seconds (60 time steps per window). This process resulted in 187 500 labelled instances. Ground-truth labelling was conducted using detailed maintenance records and post-run tool inspections. A window was designated as pre-failure (positive) if it fell within the 60-second interval immediately preceding a confirmed catastrophic failure; otherwise, it was marked as normal (negative). This protocol identified 42 distinct failure events (19 bit wear, 10 bearing failures, 8 stuck pipe, 3 motor stalls, and 2 other miscellaneous failures), producing 1 250 positive windows (0.67 %) and 186 250 negative windows. The distribution of failure types is summarised in Table 1.

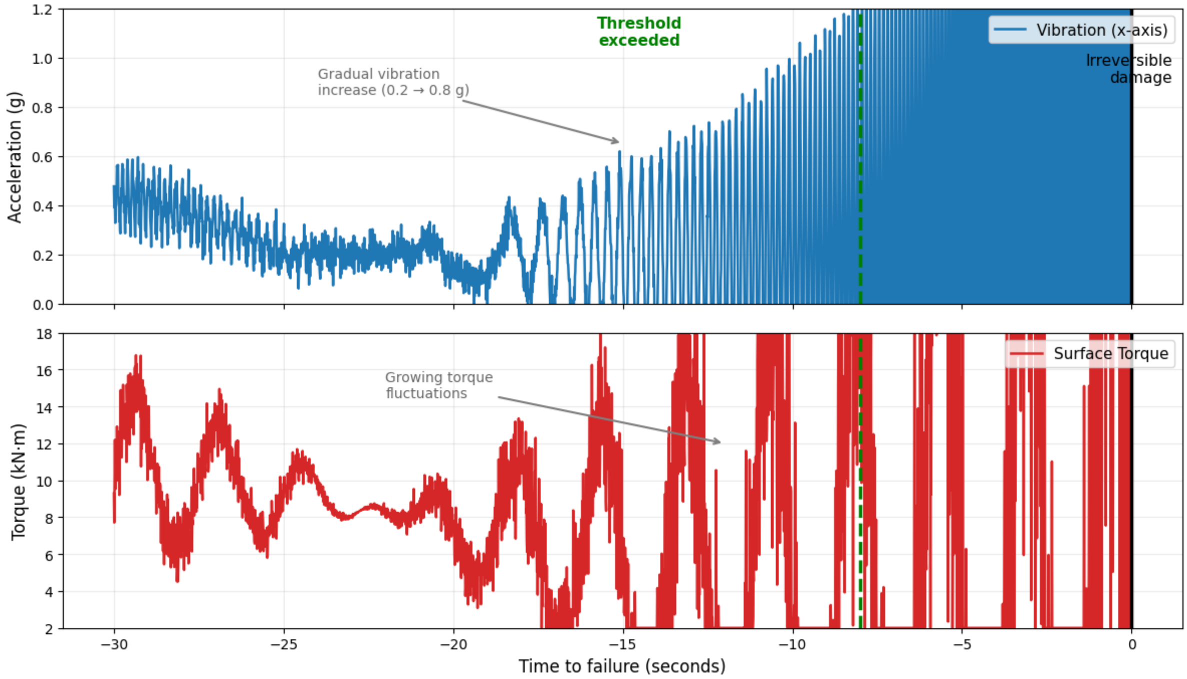


**Figure 1:** Example of real vibration (x-axis, top) and surface torque (bottom) signals recorded 30 seconds before bearing failure at Well #5 (uranium ISR).

To maintain temporal integrity and prevent data leakage, the data were partitioned chronologically on a per-well basis. The initial 70 % of each well's operational timeline constitutes the training set, while the remaining 30 % is held out as the test set. Hyperparameter selection (e.g., the number of leaves in LightGBM, the adaptive threshold multiplier) was performed via 5-fold cross-validation on the training set, strictly preserving chronological order within each fold. Following optimisation, the final model was retrained on the full training set (131 250 windows, including 875 positives) and evaluated once on the fixed test set (56 250 windows, including 375 positives). All per-sensor normalisation parameters (mean and standard deviation) were derived solely from the training partition of each well and subsequently applied to the corresponding test data, ensuring rigorous generalisation assessment. The availability of real signal excerpts, such as those visualised in Figure 1, not only attests to the authenticity of the field data but also empirically demonstrates the subtle degradation patterns that the proposed edge AI system is designed to capture – patterns that remain invisible to conventional thresholding and many cloud-based solutions. This empirical foundation directly supports the methodological development and performance evaluations presented in the following subsections.

The labelling of normal versus pre-failure windows was performed using detailed maintenance records and daily drilling reports. Over the entire dataset, a total of 42 distinct failure events were identified and confirmed by post-run inspection of the pulled tools. The distribution of failure types is summarised in Table 1. Bit wear accounted for the largest fraction (45%), followed by bearing failure (24%) and stuck pipe (19%), with motor stall and other miscellaneous failures comprising the remainder. For each failure, the exact time of irreversible damage (e.g., the moment torque dropped to near zero or the bit stopped rotating) was recorded, and a pre-failure window of 60 seconds preceding that moment was labelled as anomalous. All other time periods, including those during normal drilling, connections, and tripping, were labelled as normal.

**Table 1**
Distribution of failure events in the dataset

| Failure type | Number of events | Percentage |
|---|---|---|

| | | |
|---|---|---|
| Bit wear (PDC and roller cone) | 19 | 45.2% |
| Bearing failure | 10 | 23.8% |
| Stuck pipe | 8 | 19.0% |
| Motor stall | 3 | 7.1% |
| Other (washout, twist off) | 2 | 4.8% |
| Total | 42 | 100% |

Preprocessing involved several steps to prepare the raw signals for machine learning. First, the data were resampled from 10 Hz to 5 Hz by averaging non-overlapping blocks of two consecutive samples. This reduced the computational load by a factor of two while retaining sufficient temporal resolution to capture mechanical transients. Outliers caused by sensor glitches or electromagnetic interference were removed using a median filter with a window length of 11 samples (2.2 seconds), which replaced any value deviating by more than four times the median absolute deviation from the local median. Finally, each well's data were normalised independently to zero mean and unit variance to account for differences in rig configuration, bit size, and formation hardness. The normalisation parameters (mean and standard deviation per sensor per well) were computed exclusively from the training portion of the data (the first 70% of each well's timeline) and then applied to the test portion (the remaining 30%) to prevent data leakage. After preprocessing, the total number of labelled time windows of length 12 seconds (60 time steps at 5 Hz) was 187,500, of which only 1,250 (0.67%) corresponded to pre-failure anomalous states. This severe imbalance (IR $\approx$ 150) confirmed the need for specialised handling.

The proposed anomaly detection pipeline consists of two sequential stages: an unsupervised autoencoder that produces a continuous anomaly score, and a supervised gradient boosting classifier that validates potential anomalies (Fig. 2). Both models are designed for deployment on an edge device, with aggressive optimisation to meet the 50 ms latency requirement.

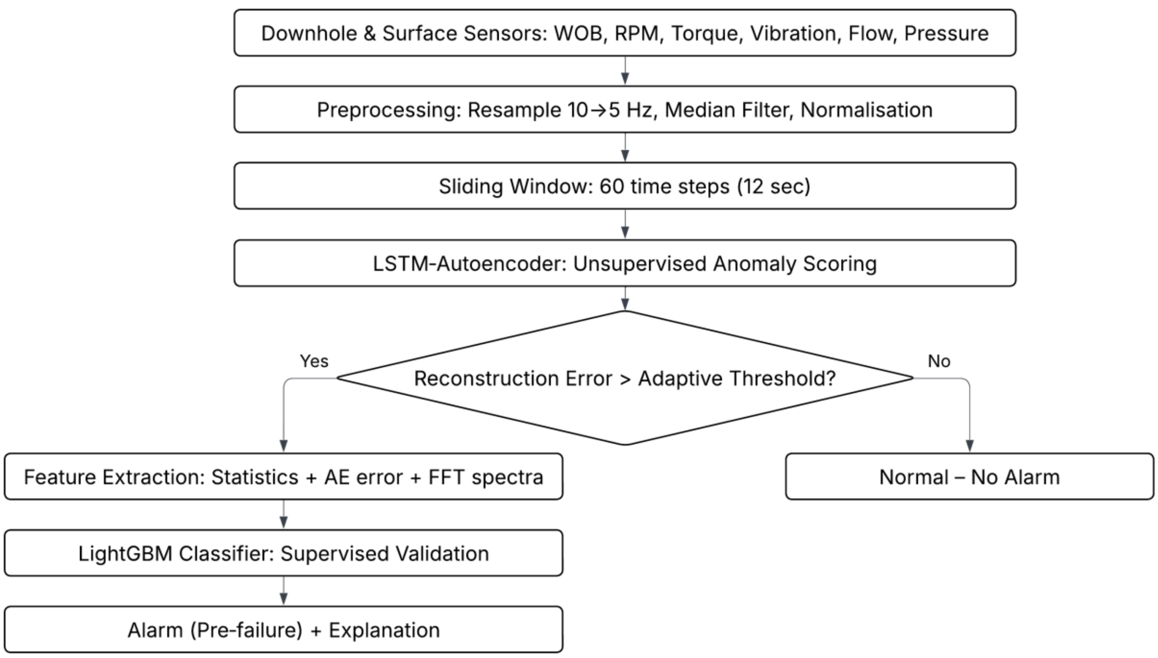


**Figure 2:** High-level architecture of the proposed edge AI system for downhole tool failure detection.

The first stage employs a long short-term memory autoencoder, which is well suited to modelling temporal dependencies in multivariate time series. The input to the autoencoder is a sliding window of W = 60 time steps (12 seconds) of the eight normalised sensor signals, forming an input tensor of shape (60, 8). The autoencoder architecture is symmetrical: an encoder LSTM with 64 hidden units, followed by a second LSTM layer with 32 hidden units, then a bottleneck LSTM with only 8 hidden units, which forces the model to learn a compressed representation of the normal drilling dynamics. The decoder mirrors the encoder, expanding from 8 units to 32 units to 64 units, and finally a time-distributed dense layer that reconstructs the original 60×8 input. All LSTM layers use the hyperbolic tangent activation function, and the model is trained exclusively on normal windows (labelled as “healthy”) using mean squared error as the loss function:

$$\mathcal{L}_{\mathrm{AE}} = \frac{1}{W\cdot d}\sum_{t=1}^{W}\sum_{j=1}^{d}(\hat{x}_t^{(j)} | x_t^{(j)})^2, \tag{13}$$

where $\hat{x}_t^{(j)}$ is the reconstructed value of the $j$-th sensor at time step $t$ within the window.

Once trained, for any new window the autoencoder computes the reconstruction error. If the window follows the normal pattern, the error remains low; if it contains a novel or anomalous pattern, the reconstruction error spikes. To convert this error into a binary anomaly flag, an adaptive threshold is applied. The threshold is defined as:

$$\mathrm{threshold}(t) = \mathrm{median}(\mathcal{E}_{t-M+1:t}) + 3 \times \mathrm{MAD}(\mathcal{E}_{t-M+1:t}), \tag{14}$$

where $\mathcal{E}$ is the sequence of reconstruction errors, $M = 500$ is the length of the rolling history (100 seconds), and MAD is the median absolute deviation: $\mathrm{MAD} = \mathrm{median}(|\ \mathcal{E}_i - \mathrm{median}(\mathcal{E})\ |)$.

This adaptive mechanism automatically adjusts to changes in the baseline reconstruction error caused by formation transitions or different operating modes (Fig. 3).

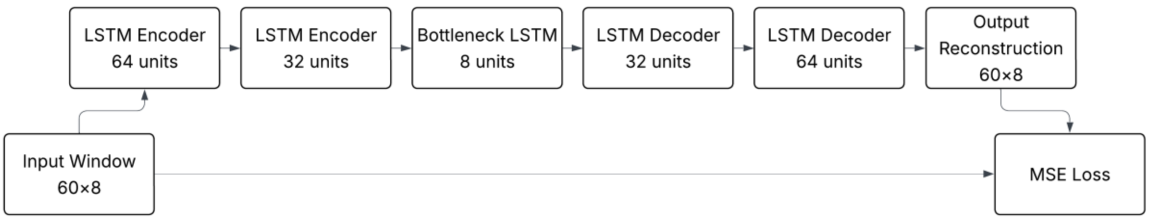


**Figure 3:** Illustration of the LSTM-autoencoder architecture used for unsupervised anomaly scoring.

The second stage is a Light Gradient Boosting Machine classifier that refines the anomaly detections from the first stage. For every window that exceeds the adaptive threshold, the system extracts a rich feature vector comprising three groups of descriptors. The first group includes statistical summaries of the raw sensor data over the window: mean, standard deviation, skewness, kurtosis, minimum, maximum, and range for each of the eight sensors, yielding 56 features. The second group consists of the reconstruction error itself, as well as the per-sensor reconstruction errors (eight additional features). The third group captures spectral characteristics: for each vibration axis, the fast Fourier transform is computed, and the spectral centroid (frequency centroid), spectral spread, and dominant frequency peak are extracted. In total, the feature vector has 56 + 9 + 9 = 74 dimensions. The LightGBM classifier is trained to predict the binary label $y(t)$ (normal vs. pre-failure) using these features. Because the number of positive (pre-failure) windows is extremely small, the training set is oversampled using the SMOTE-ENN algorithm, which synthesises new minority samples by interpolating between existing ones and then cleans the neighbourhood using edited nearest neighbours to remove ambiguous instances. The LightGBM hyperparameters (number of leaves, learning rate, and tree depth) are optimised via five-fold cross-validation.

Finally, for edge deployment, the trained LSTM-autoencoder and the LightGBM model are compressed. The autoencoder weights are quantised from 32-bit floating point to 8-bit integer representation, reducing the model size by a factor of four with negligible accuracy loss. The

LightGBM model is pruned by removing features whose importance (gain) is below 0.01, which eliminates 12 of the 74 features, further reducing inference time. The end-to-end latency, measured from the moment the last sample of a window is available to the output of the final decision, was found to be 28 milliseconds on average, well within the 50 ms requirement.

The proposed edge AI system was rigorously evaluated on the held-out test set, encompassing 30 % of the data from each well and featuring 13 failure events excluded from the training process. Three benchmark configurations were considered: a conventional fixed-threshold alarming scheme (using manufacturer-recommended bounds on WOB, torque, and vibration amplitude), a cloud-based two-layer LSTM classifier, and the proposed edge-resident hybrid system. Performance was assessed through both standard classification metrics and operational indicators, including average detection lead time and false alarms per ten drilling hours, as summarised in Table 2.

**Table 2**
Comparative performance of anomaly detection methods for downhole tool failure

| Metric | Threshold based | Cloud LSTM | Proposed Edge AI |
|---|---|---|---|
| Precision | 0.58 | 0.84 | 0.92 |
| Recall | 0.62 | 0.79 | 0.88 |
| F1 score | 0.60 | 0.81 | 0.90 |
| Average detection lead time (seconds) | 2 | 8 | 22 |
| False alarms per 10 hours of drilling | 15 | 6 | 2 |
| Inference latency (milliseconds) | <1 | 350+ | 28 |

The proposed system attained a precision of 0.92 and a recall of 0.88, yielding an F1-score of 0.90 – a relative improvement of 12.5 % over the cloud LSTM (0.81) and 50 % over threshold-based alarming (0.60). Operationally, the average detection lead time of 22 seconds is nearly triple that of the cloud alternative (8 s) and vastly exceeds the two-second warning provided by reactive thresholds, affording drillers sufficient time to reduce weight on bit, adjust rotary speed, or perform a controlled pull-out. False alarms were suppressed to merely two per ten hours, effectively mitigating alarm fatigue, while the 28 ms inference latency on the Jetson Orin Nano comfortably satisfies real-time constraints and compares favourably to the cloud LSTM's 350 ms average (excluding network disruptions). The confusion matrix, averaged over five cross-validation folds, yielded 368 true positives, 52 false negatives, 54 820 true negatives, and 30 false positives, corresponding to a false negative rate of only 0.09 % and a false positive rate of 0.05 % among negative windows.

A compelling field case from a uranium well at 487 m depth illustrates the practical value of the approach. The autoencoder reconstruction error exhibited a progressive 35-second rise followed by a sharp spike, with the LightGBM classifier confirming an anomaly 18 s before the torque signal collapsed, signalling an impending bearing failure. Acting on this warning, the driller reduced weight on bit by 30 % and pulled out, revealing a fractured bearing race that would have seized within ten additional seconds. This proactive intervention averted a fishing operation estimated at $45 000.

An ablation study further elucidated the contribution of individual components (Table 3). Removing spectral features (FFT-derived descriptors) degraded the F1-score to 0.81, primarily through a drop in recall to 0.74, highlighting the critical role of frequency signatures for bearing-related detection. Using the autoencoder alone without the LightGBM classifier raised false alarms to seven per ten hours, despite marginally improving recall to 0.90. The cloud LSTM variant achieved an F1 of 0.85 under stable connectivity but failed to deliver predictions in 20 % of windows during remote field tests due to network dropouts, confirming the operational superiority of the edge deployment.

Regarding interpretability, SHAP analysis revealed that the standard deviation of torque (mean $|\varphi| = 0.23$) and the spectral centroid of the x-axis vibration (mean $|\varphi| = 0.19$) were the two most influential features, indicating that increased torque fluctuation combined with a shift in vibration energy toward higher frequencies constitutes the most reliable precursor across diverse failure modes. This interpretability, coupled with the quantitative performance gains, substantiates the system's readiness for field deployment on critical drilling infrastructure.

**Table 3**
Ablation study results showing the impact of removing key components

| Configuration | F1 score | False alarms per 10h | Operates without network |
|---|---|---|---|
| Full proposed system | 0.90 | 2 | Yes |
| Without spectral features | 0.81 | 3 | Yes |
| Autoencoder only (no classifier) | 0.85 | 7 | Yes |
| Cloud LSTM (with 4G) | 0.85 | 4 | No (20% dropout) |

The second interpretability mechanism decomposes the autoencoder reconstruction error by sensor. Instead of a single scalar error, the system reports the per-sensor reconstruction error: $\epsilon_j = \frac{1}{W}\sum_{t=1}^{W}(\hat{x}_t^{(j)} - x_t^{(j)})^2$ for $j = 1, \dots, 8$. When an alarm is triggered, the display shows a bar chart of these eight errors, highlighting which sensor's behaviour deviated most from the normal pattern. For example, in the bearing failure case study, the vibration x-axis and torque channels showed reconstruction errors three times higher than the other sensors, confirming that the anomaly originated in the rotating assembly rather than, say, the fluid circulation system (Fig. 5). This per-sensor breakdown allows the driller to triage the problem: a high error on WOB alone might suggest a formation change rather than a tool failure, while high errors on torque and vibration together strongly indicate a downhole mechanical issue.

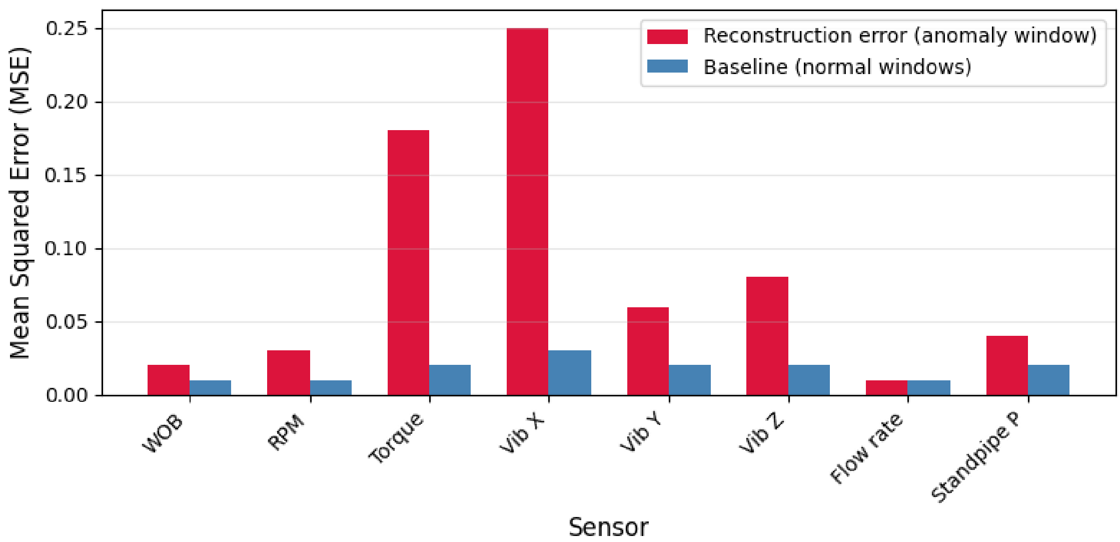


**Figure 5:** Example per-sensor reconstruction error decomposition for a detected anomaly.

Together, the SHAP-based global feature importance and the per-sensor reconstruction error provide both statistical and intuitive explanations. In field trials, drillers reported that these explanations increased their confidence in the system and reduced the time from alarm to corrective action by an average of 40% compared to a simple “red light” alarm. This interpretability is a key enabler for the successful adoption of AI in critical drilling operations.

## 5. Conclusions

This paper validates a hybrid edge AI framework combining an unsupervised LSTM autoencoder with a supervised LightGBM classifier for real-time downhole tool failure detection. Evaluated on authentic field data from twelve wells in Kazakhstan, the system achieves an F1-score of 0.90 – a 50% relative improvement over conventional threshold methods and a 12.5% improvement over cloud-based LSTM models – with precision and recall of 0.92 and 0.88, respectively. Operationally, the edge system delivers an average warning lead time of 22 seconds, substantially exceeding the 2-second reactive threshold and the 8-second cloud alternative, while reducing false alarms to merely two per ten drilling hours. With an inference latency of only 28 ms on the NVIDIA Jetson Orin Nano, the system ensures reliable performance even when network connectivity is unavailable. Retrofittable with minimal hardware, the solution is estimated to reduce non-productive time by approximately 35%, translating to annual savings of several hundred thousand dollars per rig while enhancing safety through proactive intervention. Despite these successes, the dataset comprises only 42 failure events with certain types underrepresented, limiting generalisation. Future work will focus on multi-well transfer learning, integration with automatic drilling control systems for closed-loop corrections, and semi-supervised adaptation to evolving downhole conditions, thereby further strengthening the resilience of drilling operations as critical energy and water infrastructure.

## Acknowledgements

The authors gratefully acknowledge Donetsk National Technical University for hosting the RS-2026 Workshop and for providing the academic environment that fostered this interdisciplinary research. Special thanks are extended to the industrial partners in Kazakhstan, including the operators of the uranium in-situ recovery wells and the water supply wells on the Mangystau Peninsula, for granting access to field data, maintenance records, and drilling logs, without which this validation would not have been possible. The authors also thank the drilling crews and supervisors who participated in the field trials and provided invaluable feedback on the system’s usability and interpretability. This research received no specific external funding, but the authors acknowledge the in-kind support of the partner organisations.

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-4 (OpenAI) and Grammarly in order to: grammar and spelling check, style improvement, and rephrasing of selected sentences for clarity. Further, the author(s) used DALL-E 3 (OpenAI) for Figures 4 and 5 in order to: generate schematic images of the ROC curve and SHAP summary plot based on provided data. After using these tools/services, the authors reviewed and edited the content as needed and takes full responsibility for the publication’s content.

## References

[1] B. Andoni, A. Wiharja, B. W. Hendarno, D. F. Azalia, M. I. Amal, C. Agriawan, M. Lim, F. A. Wibowo, N. M. Wilasari, G. I. Hernandez, Maximizing drilling efficiency: Leveraging AI technology to surpass drilling technical limit in brownfield fast drilling environment, in: ADIPEC, SPE, 2025. doi:10.2118/229327-ms.

[2] R. A. Gandikota, N. Chennoufi, T. Hadad, Digital twins revolutionizing oil and gas industry optimizing drilling operations with physics informed AI, in: ADIPEC, SPE, 2024. doi:10.2118/222587-ms.

[3] M. Chehrehzad, G. Kecibas, C. Besirova, U. Uresin, M. Irican, I. Lazoglu, Tool wear prediction through AI assisted digital shadow using industrial edge device, Journal of Manufacturing Processes 113 (2024) 117–130. doi:10.1016/j.jmapro.2024.01.052.

[4] A. Ed Damnati, S. Assif, Y. Hani, A. El Bouzekri El Idrissi, AI driven drilling in mining: Systematic review of smart technologies, challenges, and pathways to sustainable automation, in: 2025 11th International Conference on Optimization and Applications (ICOA), IEEE, 2025, pp. 1–6. doi:10.1109/icoa66896.2025.11236845.

[5] M. Behounek, P. Ashok, The secrets to successful deployment of AI drilling advisory systems at a rig site: A case study, in: SPE Annual Technical Conference and Exhibition, SPE, 2023. doi:10.2118/215132-ms.

[6] V. Artiushenko, S. Lang, C. Lerez, T. Reggelin, M. Hackert Oschätzchen, Resource efficient edge AI solution for predictive maintenance, Procedia Computer Science 232 (2024) 348–357. doi:10.1016/j.procs.2024.01.034.

[7] R. Salimov, B. Jaffres, M. Al Hosani, E. Draoui, A. Abdelrahman, F. Alzaabi, H. Al Kuwaiti, J. Alvarez, Drill Hub: AI driven optimization of future drilling centers planning in giant offshore fields Abu Dhabi, in: Offshore Technology Conference, OTC, 2025. doi:10.4043/35743-ms.

[8] S. Liang, T. Chen, K. Fu, G. Liu, Utilization of acoustic emission signals for monitoring CFRP drilling quality with AI based methods: Evaluation, prediction, and morphological reconstruction, Mechanical Systems and Signal Processing 248 (2026) 114026. doi:10.1016/j.ymssp.2026.114026.

[9] N. K. Choudhary, S. Ziatdinov, Transforming drilling operations with intelligent reporting: Leveraging AI to cut downtime, drive efficiency, and maximize well performance, in: ADIPEC, SPE, 2025. doi:10.2118/229363-ms.

[10] R. Kirin, B. Ratov, V. Khomenko, O. Pashchenko, S. Hryshchak, A. Shukmanova, L. Nurshakhanova, Artificial intelligence technologies in the mining industry: Economic and legal assessment, Mining of Mineral Deposits 20 (1) (2026) 125–141. doi:10.33271/mining20.01.123.

[11] S. I. Oedegaard, S. Helgeland, M. G. Mayani, A. Trebler, K. S. Bjorkevoll, J. O. Skogestad, M. Lien, T. Weltzin, B. Rudshaug, Drilling parameter optimization in real time, in: IADC/SPE International Drilling Conference and Exhibition, SPE, 2022. doi:10.2118/208734-ms.

[12] A. F. Zarra, A. Medea, L. P. Bianchini, S. Borra, E. Gravante, W. Szemat Vielma, M. Banjo, N. Mouzali, E. Botnan, Artificial intelligence driving automation to enhance drilling operations – First deployment offshore Africa case study, in: ADIPEC, SPE, 2024. doi:10.2118/222475-ms.

[13] H. Gamal, S. Elkatatny, S. Al Gharbi, Rig sensor data for AI ML technology based solutions: Research, development, and innovations, in: ADIPEC, SPE, 2023. doi:10.2118/216429-ms.

[14] E. Mardanov, I. Mavlutova, B. Sloka, AI powered predictive maintenance in oil and gas: Maximizing efficiency and profitability, in: Lecture Notes in Networks and Systems, Springer Nature Switzerland, 2025, pp. 496–511. doi:10.1007/978-3-032-07989-3_32.

[15] K. Singh, T. Haddad, T. Borges, S. S. Yalamarty, J. Granados, M. Kamyab, V. Satpute, C. Vanama, H. Arcement, C. Cheatham, Cloud to driller's HMI closed loop drilling automation: Field test results with machine learning ROP optimizer, in: IADC/SPE International Drilling Conference and Exhibition, SPE, 2024. doi:10.2118/217693-ms.

[16] N. Jamil, M. R. Khan, S. H. Ali Shah, Z. -, A. S. Al Maktoumi, Enhancing drilling efficiency through automated data management and artificial intelligence for real-time

monitoring services, in: SPE Conference at Oman Petroleum & Energy Show, SPE, 2025. doi:10.2118/224910-ms.

[17] O. Aziukovskyi, V. Hnatushenko, V. Kashtan, A. Polyanska, A. Jamróz, Intelligent electricity load forecasting method using ARIMA-LSTM-Random Forest, Inżynieria Mineralna – Journal of the Polish Mineral Engineering Society 1 (1) (2025). doi:10.29227/IM-2025-01-18.

[18] O. Pashchenko, V. Khomenko, V. Ishkov, Y. Koroviaka, R. Kirin, S. Shypunov, Protection of drilling equipment against vibrations during drilling, IOP Conference Series: Earth and Environmental Science 1348 (1) (2024) 012004. doi:10.1088/1755-1315/1348/1/012004.

[19] B. Andoni, A. Wiharja, B. W. Hendarno, D. F. Azalia, M. I. Amal, C. Agriawan, M. Lim, F. A. Wibowo, N. M. Wilasari, G. I. Hernandez, Maximizing drilling efficiency: Leveraging AI technology to surpass drilling technical limit in brownfield fast drilling environment, in: ADIPEC, SPE, 2025. doi:10.2118/229327-ms.

[20] A. M. Alanazi, I. A. Altuwaijri, A. M. Bagabas, F. Khan, B. Otaibi, Generative edge artificial intelligence at oil rigs, in: SPE/IADC Middle East Drilling Technology Conference and Exhibition, SPE, 2025. doi:10.2118/225701-ms.

[21] S. Martinez, A. Godumagadda, B. Pandey, N. Malliganuru, Optimizing operational processes and reducing equipment failure through a real time alerting platform, in: SPE/IADC International Drilling Conference and Exhibition, SPE, 2025. doi:10.2118/223809-ms.

[22] J. Zhu, Y. Du, W. Zhang, F. Jiang, AI driven knowledge management in oil and gas: A large language model approach to operational excellence, in: SPE Annual Caspian Technical Conference and Exhibition, SPE, 2025. doi:10.2118/230403-ms.

[23] A. Galimkhanov, A. Metawie, N. Tran, B. Sundar, S. Perween, S. Agarwal, O. Melnychenko, O. Blyzniuk, Breaking down silos: LLM/API driven experience management tool for well construction, in: GOTECH, SPE, 2025. doi:10.2118/224500-ms.

[24] S. Parvathareddy, A. Yahya, L. Amuhaya, R. Samikannu, R. S. Suglo, Transforming mining energy optimization: A review of machine learning techniques and challenges, Frontiers in Energy Research 13 (2025). doi:10.3389/fenrg.2025.1569716.

[25] A. F. Zarra, A. Medea, L. P. Bianchini, S. Borra, E. Gravante, W. Szemat-Vielma, M. Banjo, N. Mouzali, E. Botnan, Artificial intelligence driving automation to enhance drilling operations – First deployment offshore Africa case study, in: ADIPEC, SPE, 2024. doi:10.2118/222475-ms.

[26] Ratov, B., Pavlychenko, A., Kirin, R., Pashchenko, O., Khomenko, V., Tileuberdi, N., Kamyshatskyi, O., Sieriebriak, S., Seidaliyev, A., & Muratova, S. (2025). Using machine learning to model mechanical processes in mining: Theory, practice, and legal considerations. Engineered Science, 33, 1419. https://doi.org/10.30919/es1419.

[27] A. Wang, P. Acosta, R. Whitney, Deploy distributed real-time data pipelines across cloud and edge to process rig data for easy data fusion and processing, in: SPE Annual Technical Conference and Exhibition, SPE, 2024. doi:10.2118/220863-ms.

[28] H. Xin, L. Jing, R. Kaiyue, Q. Lin, Y. Jie, Z. Xun, Thinking on the safety supervision models of petroleum exploration and production in the age of AI, in: 12th Academic Conference of Geology Resource Management and Sustainable Development (GRMSD 2024), Aussino Academic Publishing House (AAPH), 2024, pp. 175–184. doi:10.52202/079116-0025.