

AI-Driven Non-Productive Time Reduction in Saudi Drilling Operations: Predictive Detection of Wellbore Instability, Stuck Pipe and Downhole Equipment Failures

Qamar Ur Zaman

Published October 2026

Abstract

Drilling non-productive time arises when mechanical dysfunction, unstable formations and failing equipment interrupt planned progress. Predictive analytics could shorten the interval between an emerging precursor and a suitable intervention, yet model discrimination is not equivalent to verified operational savings. This critical review examines published research from 2020 to 2025 on stuck-pipe precursors, geomechanical instability, drilling-fluid loss, harmful vibrations and downhole component degradation, with Saudi Arabian operating conditions as the principal application context. Evidence is compared through signal availability, physical plausibility, prediction horizon, dataset separation and intervention feasibility. The analysis identifies complementary rather than interchangeable functions for geomechanical models, temporal anomaly detectors, vibration classifiers and component-health estimators. It proposes a risk-gated monitoring architecture in which sensor quality checks, uncertainty estimates and mechanistic screening precede actionable recommendations. A distinction is maintained between laboratory findings, historical-well validations and prospective reductions in rig downtime. The resulting framework prioritises interpretable early warnings, crew authority, reliable telemetry and post-deployment evaluation. Its principal contribution is an integrated decision pathway connecting the forecast event, a proportionate operational response and measurable avoided disruption rather than a single universal prediction score.

Keywords: non-productive time; drilling analytics; stuck pipe; geomechanics; tool reliability; Saudi Arabia; early warning

1 Introduction

Non-productive time (NPT) is the portion of a drilling programme spent responding to events that prevent planned advancement or completion. Fishing, stuck-pipe recovery, unscheduled trips, remedial circulation, damaged bottom-hole assemblies and repeated conditioning may each appear as discrete delays, although their physical causes are frequently coupled. Consequently, predicting an isolated event does not necessarily establish that any delay has been prevented. An alarm creates value only when its warning horizon exceeds the time needed to diagnose the anomaly and implement a safe response. Recent reviews describe rapid progress in drilling machine learning but also persistent difficulties with data provenance, transferability and operational interpretation [1, 2, 17].

Saudi drilling presents a demanding setting for this distinction. Extended-reach trajectories, heterogeneous carbonate intervals, fractured or depleted sections, demanding casing programmes and variable drilling-fluid behaviour can generate competing symptoms. Increasing torque may indicate inadequate hole cleaning, differential sticking, progressive mechanical restriction or tool

deterioration. A pressure change may arise from altered hydraulics rather than a formation event. Local calibration, geological context and sound engineering judgement therefore remain essential even when algorithms detect reproducible patterns. Regional studies of stuck-pipe prediction support the feasibility of data-driven screening but do not justify assuming identical performance across Saudi fields [4, 10].

This review addresses a more operationally specific question than broad surveys of artificial intelligence in exploration. Its aim is to determine how predictive systems can distinguish emerging mechanisms, prioritise interventions and be evaluated against actual avoided NPT. Four objectives guide the argument: identify technically defensible precursors for different failure modes; compare approaches under scarce event labels and changing formations; establish safety-conscious integration with rig decisions; and define a measurable adoption pathway appropriate for Saudi operators. The review does not introduce field experiments, claim access to proprietary drilling records or calculate savings from invented datasets. It offers an evidence-grounded synthesis and a proposed implementation framework.

2 Review approach and evidence assessment

A structured critical narrative review was adopted because the literature spans heterogeneous datasets, interventions and outcome definitions that are poorly suited to pooled effect estimation. Candidate studies were identified through thematic searches for wellbore stability, stuck-pipe forecasting, lost circulation, real-time rig analytics, vibration detection, drill-bit wear, mud-motor stall and drilling equipment prognostics. Peer-reviewed articles and identifiable industry conference papers dated 2020–2025 were prioritised when they reported a relevant method, an operational measurement or a limitation transferable to drilling practice. The supplied 2026 review informed presentation and technical context but was not included in the dated evidence base.

The selected body of thirty DOI-bearing publications was organised into five evidence categories: instability mechanisms; mechanical sticking; fluid losses and circulation; drilling dynamics and tool wear; and end-to-end monitoring architecture. Studies were appraised for origin of labels, unit of model validation, operating regime, warning lead time, class balance and the relationship between predicted signal and feasible action. Particular attention was given to whether testing occurred on independent wells rather than randomly shuffled measurements from the same continuous trace [3, 6, 19, 20]. The latter arrangement may leak geological and operational similarities between training and evaluation, producing optimistic performance estimates.

Comparative findings are presented as qualified implications, not an artificial league table of accuracies. Reported success on historical logs establishes pattern recognition under the tested conditions; it does not automatically demonstrate real-time communication latency, acceptance by drillers, or economic improvement. Conference reports offer valuable field-oriented designs but may provide less reproducible datasets than journal studies. Where an approach is mechanistically plausible yet not independently validated in Saudi operations, it is labelled a candidate for prospective evaluation. Figures and tables therefore represent an original synthesis rather than measurements of a purported multi-rig experiment.

The review also distinguishes prevention, early recognition and classification after onset. These are often described with the same word, prediction, despite different temporal requirements. A stuck-pipe classifier trained on measurements captured while the pipe is already immobile cannot be assumed to warn operators before sticking. Likewise, wear models using inspection labels obtained

after a bit is retrieved require careful reconstruction of the earlier exposure history. Literature conclusions were therefore assessed against a conceptual evidence chain: independent inputs; pre-event timing; a meaningful target; usable forecast horizon; and a defined operational choice. DOI-linked bibliographic records were checked against accessible publication descriptions, but unrestricted full-text access was not available for every article. The resulting synthesis is necessarily selective, with no claim to exhaustive database coverage, independently recalculated statistics or formal quantitative meta-analysis. These boundaries limit the strength of any pooled inference but protect the interpretation from a false appearance of precision.

3 Mechanical pathways to avoidable downtime

Wellbore instability begins when the balance between in-situ stress, pore pressure, drilling-fluid support and rock strength becomes insufficient. Its manifestations include breakouts, enlargement, collapse, excessive cavings and drilling-induced fractures. These are not interchangeable binary targets: a growing caliper enlargement may precede poor hole cleaning, while fracture initiation may first present as changing fluid returns. A review of machine-learning stability analysis distinguishes models estimating intermediate physical properties from models directly forecasting unstable states [1]. Both depend on accurate formation context and should be compared with interpretable geomechanical expectations.

Mud density illustrates a critical control conflict. Insufficient support can promote shear failure or influx, whereas excessive equivalent circulating density may initiate losses, reduce fluid-column support and create a secondary stability problem. Therefore, an algorithm that classifies high sticking risk without considering the pressure window can recommend inappropriate corrective actions. The appropriate state representation combines trajectory, depth, mud properties, pore-pressure interpretation, fracture margin, cuttings loading and recent hydraulic transients. Digital stability prediction demonstrates the value of coupling physics and learned relationships, but its applicability depends on the representativeness of neighbouring wells and formation properties [2, 14].

Lost circulation can both generate NPT directly and precipitate additional complications. Predictive studies use mudlogging, fluid parameters and geological descriptors to estimate its likelihood or severity [11–13]. Mixture-density models are particularly interesting because they retain uncertainty instead of forcing a precise point estimate. Yet the labels used in many retrospective studies reflect recorded loss volumes after recognition; they may not reveal the earliest detectable precursor. Flow-back during wellbore breathing also resembles genuine influx or loss. A robust alert must therefore test alternative hydraulic explanations and avoid recommending fluid treatment solely because a statistical threshold has been exceeded.

Equipment deterioration operates on a different time scale. Drilling vibration, bit-to-rock interaction, bearing degradation, motor stalling and drillstring fatigue may develop episodically and accelerate following parameter changes. Increased rate of penetration is not always beneficial if higher weight on bit or rotation amplifies destructive oscillation. Models predicting penetration performance should be constrained by mechanical energy, torsional stability and component loading, rather than optimising drilling speed in isolation [15, 16, 25]. This distinction makes NPT a systems outcome arising from both formation response and operational control.

Drilling activity further changes the meaning of observed signals. During sliding, rotation and axial loading differ from rotary drilling, while pump transitions introduce pressure disturbances unrelated to formation deterioration. Tripping and reaming expose the drillstring to different combinations of

drag, cuttings mobility and pressure imbalance. A monitoring system should classify these states first and compare each interval only with operationally comparable conditions. Trend features expressed per depth drilled may be more informative for progressive hole deterioration, whereas high-frequency acceleration sequences are more suitable for mechanical vibration. Their sampling rates, noise structures and feasible interventions are fundamentally different. Treating every channel as an interchangeable input to one deep network obscures those distinctions and increases susceptibility to confounding. In practice, a state-aware model may contain several simpler specialised estimators joined by a coherent decision interface. Specialisation is not a weakness when it improves calibration, fault isolation and the ability of engineers to challenge a misleading signal. Table 1 summarises the main precursors, modelling roles and confounders for each pathway.

Risk-gated early-warning architecture



Model outputs are advisory; established well-control and rig-safety barriers remain authoritative.

Figure 1. Risk-gated workflow linking early warnings with auditable decisions (original conceptual synthesis).

Table 1. Mechanism-oriented early-warning evidence matrix.

Failure pathway	Primary precursors	Modelling role	Key confounder
Wellbore instability	Cavings, enlargement, density, losses	Geomechanics + learned risk	Changing lithology or mud regime
Stuck pipe	Drag, hook load, torque, flow response	Temporal anomaly + causal rules	Connections and routine conditioning
Drilling-fluid loss	Return-flow deficit, ECD, fracture context	Severity and uncertainty model	Ballooning and fluid transfers
Equipment degradation	Torsion, shock, wear, motor loading	State estimation + lifetime trend	Formation-driven mechanical changes

4 Predicting stuck pipe: from precursors to decisions

A stuck pipe event can arise through differential pressure against a permeable wall, cuttings-bed accumulation, key seating, ledges, formation collapse or mechanical deformation. These mechanisms can produce overlapping torque, drag, hook-load and pressure signatures, but imply different interventions. Consequently, binary prediction provides only the first layer of diagnosis. Regional work using particle-swarm-optimised neural networks demonstrated feasibility for estimating sticking risk from multiple drilling variables; nevertheless, its Middle Eastern field data should not be interpreted as a universal Saudi calibration [4]. Models based on operationally available variables

remain more deployable than those depending on laboratory properties that cannot be refreshed during drilling.

Early-warning indicators grounded in translational resistance, rotational response and hydraulic abnormalities offer a causal bridge between raw rig signals and sticking onset. Brankovic and colleagues constructed interpretable indicators from mudlog data and translated their evolution into a depth-dependent risk estimate [5]. Subsequent event-based modelling linked precursor occurrences through a hidden-state sequence, recognising the fact that a single anomalous sample rarely establishes imminent pipe sticking [6]. This state-transition perspective is preferable where the alarm must distinguish developing trouble from momentary control adjustments or connections.

Unsupervised approaches address the rarity of labelled stuck incidents. A recurrent autoencoder trained to represent normal operations treats sustained reconstruction error as potential deviation; field-well research demonstrates the diagnostic promise of this arrangement [3]. A later autoencoder investigation reported encouraging anomaly-detection statistics on a named offshore dataset [29]. However, normal operating regimes are themselves diverse. Changes in bit type, formation, circulation mode and operating parameters can all increase anomaly scores without identifying the underlying failure. A practical detector should condition its baseline on operational state and explain which sensors caused the departure.

Expert rules and fuzzy inference retain usefulness where mechanical knowledge is more portable than statistical event labels. A feature-engineered fuzzy expert system can represent gradual deterioration in hole cleaning rather than treating every condition as sharply normal or abnormal [7]. Combined knowledge-driven and data-driven methods further allow experts to interpret trends in friction and sticking probability together [8]. These approaches are not intrinsically safer than neural networks; rules still require revision, and fuzzy membership thresholds can be subjective. Their comparative advantage is inspectability during an urgent operational discussion.

Prediction must also be tied to a decision boundary. For suspected cuttings loading, the response may involve reviewing transport efficiency and circulation practices; for differential sticking, the required analysis differs materially. Studies addressing wellbore uncleanliness and decision-tree remedies demonstrate why the precursor mechanism matters [9, 10]. A deployment should log the observed symptoms, competing hypotheses, responsible engineer, action taken and subsequent outcome. That record allows retrospective differentiation of an accurate yet unactionable alarm from a warning that genuinely prevented recovery operations.

A warning policy must distinguish the cost of missing a developing stuck-pipe event from the cost of stopping work unnecessarily. The appropriate threshold depends on formation exposure, trajectory, available circulation capacity, planned connections and the estimated reversibility of the condition. Fixed thresholds may therefore be inferior to risk-dependent escalation bands, provided those bands are documented rather than adjusted to explain results retrospectively. The operator interface should display persistence and trend alongside probability; rapid reversals caused by a known parameter adjustment should not generate the same message as steadily deteriorating drag over a long open-hole interval. When competing sticking mechanisms receive similar likelihoods, the algorithm should recommend confirming observations rather than fabricate a single precise diagnosis. This calibrated uncertainty is operationally valuable because incorrect remedial treatment can amplify the original difficulty. Testing should report whether alerts arrived before the crew noticed trouble, not simply whether the final event label was classified correctly.

5 Wellbore deterioration and circulation risk

An instability predictor should be evaluated not only by correct classification but also by how far ahead of a consequential event it recognises deteriorating conditions. When predictive labels are derived from caliper logs acquired after drilling, apparently strong agreement may not provide advance warning during the preceding interval. Research using artificial neural networks to estimate borehole enlargement illustrates how geomechanical descriptors can support the assessment of mud-weight alternatives [14]. However, the best intervention is constrained by pore pressure, fracture resistance, hole trajectory and the hydraulic consequences of any fluid adjustment.

Mud-loss forecasting often involves severe event imbalance because stable circulation occupies most of the measured operating time. Wood and colleagues explicitly examine the impact of this imbalance on classification reliability [12]. A model reporting high overall accuracy can still miss the rare losses most likely to generate significant downtime. Reviews should therefore prioritise event recall, precision, false alarms per operating day and the measured lead time before intervention. Loss-severity regression offers a complementary perspective, as the economic consequences of a small manageable loss are not equivalent to those of a major loss requiring remedial operations [11, 13].

Where competing mechanisms share surface signatures, hybrid reasoning is desirable. A rise in standpipe pressure might indicate cuttings restriction, changing rheology or a blocked flow path, whereas a decrease could accompany washout or altered hydraulic resistance. Simultaneous torque escalation, depth-dependent drag and poor returns give a different diagnostic picture from an isolated pressure excursion. Physics-informed checks should challenge rather than merely reinforce an AI classification: the proposed physical cause must be capable of producing the observed response within the measured operating state [18, 21]. This principle helps prevent spurious correlations becoming operational instructions.

The time structure of fluid-loss events is especially consequential. A formation may absorb whole drilling fluid intermittently before producing a sustained return-flow deficit; equally, transient losses associated with pressure cycling may reverse without a damaging event. The distinction cannot be inferred reliably from one instantaneous measurement. Combining pumping history, annular pressure estimates, fluid-property changes and fracture interpretation is more physically defensible than treating pit-volume movement as an isolated label. Separate estimation of event onset, plausible loss rate and confidence in the diagnosis would support more proportionate decisions. The evidential threshold should rise when a proposed action could compromise pressure control or obscure subsequent measurements. Additional geomechanical information should be refreshed when new formations are encountered, rather than relying indefinitely on an initial drilling prognosis. An algorithm's capacity to fit historical curves must remain subordinate to the consistency of its proposed mechanism with current hydraulic constraints and approved operating limits.

6 Downhole equipment failure and drilling dynamics

The equipment-failure pathway warrants its own prognostic architecture because component health is cumulative and inspections may occur only after retrieval. Drill-bit tooth wear gradually lowers penetration efficiency and can alter torque signatures, yet formation hardening can produce similar effects. Agostini and Sampaio investigated spectral torque imaging and autoencoder-derived features for bit-wear monitoring [23]. More recent domain-adaptation work uses performance relationships to estimate wear across changing conditions [22]. Both lines of work support probabilistic mainten-

ance advice, not a guarantee that an individual bit can safely remain downhole for a predicted number of hours.

The dynamics of a bottom-hole assembly involve axial bounce, lateral whirl and torsional stick-slip. These oscillations accelerate fatigue and premature wear, potentially causing unscheduled trips, loss of tools or damaged connections. Classification using downhole acceleration shows promise for separating vibration modes, but simulation-trained models require robust assessment against field measurements [24]. Saudi-linked research has also examined whether surface variables can estimate vibrations in curved intervals [26]. A deployable model must distinguish a transient high-amplitude episode from sustained damaging exposure, retaining information about duration and repeated loading.

Optimising controllable drilling parameters introduces a competing objective. Weight on bit and rotary speed influence both instantaneous penetration and destructive vibration. A multi-model optimisation study treated torsional conditions as constraints rather than maximising rate alone [25]. This is a stronger formulation than conventional speed optimisation because unplanned equipment failure may dominate any short-term gain. Nevertheless, simulated operating improvements should never be reported as observed reductions in fishing or trip time without a prospective field comparison.

Equipment failure-time prediction is more difficult than anomaly recognition. The condition of mud motors, sensors and electronics is influenced by exposure history, operating temperature, downhole vibration, bit loading and individual component variability. Saudi industry investigations have explored neural prognostics and anomaly alarms for drilling equipment [27], while a regional study examined early warnings of mud-motor stalls [28]. These demonstrate practical relevance but have different evidential status from independently replicated, prospective reliability trials. Failure-time predictions should therefore be expressed with uncertainty and reviewed alongside maintenance records and duty-cycle information.

Reliability assessment must also recognise censored observations. Bits and motors removed during planned trips may not have failed, while components replaced after suspected deterioration can make the apparent service-life distribution shorter than the true failure distribution. Training a simple regression on only confirmed failures discards information about healthy operating exposure and can bias remaining-life forecasts. Survival methods or explicitly labelled health-state trajectories are better suited to some maintenance questions, although their suitability still depends on credible component identities and service histories. Tool-level validation should separate equipment families, duty cycles and formations; random windows from the same motor are not an independent test of generalisation. Where sensor bandwidth does not permit reliable vibration reconstruction, a model should acknowledge reduced diagnostic confidence rather than substitute surface variables as though they were identical to downhole measurements. Maintenance planning also needs clear ownership for follow-up inspections, because an unverified predicted defect can become another avoidable source of downtime.

7 Integrated real-time monitoring architecture

The proposed architecture begins by aligning surface and downhole observations to a coherent drilling timeline. Data streams may include weight on bit, torque, rotation, hook load, rate of penetration, standpipe pressure, flow returns, pit volume, mud properties, tool vibration and annular pressure. Every feed requires unit checks, sensor-health flags, timestamp reconciliation, operational-

state classification and explicit records of gaps. Research on recurrent drilling-data models shows that preprocessing choices materially affect predictive performance [20]. A model trained on cleaned retrospective sequences may behave quite differently when faced with delayed packets, rig interventions and interrupted telemetry.

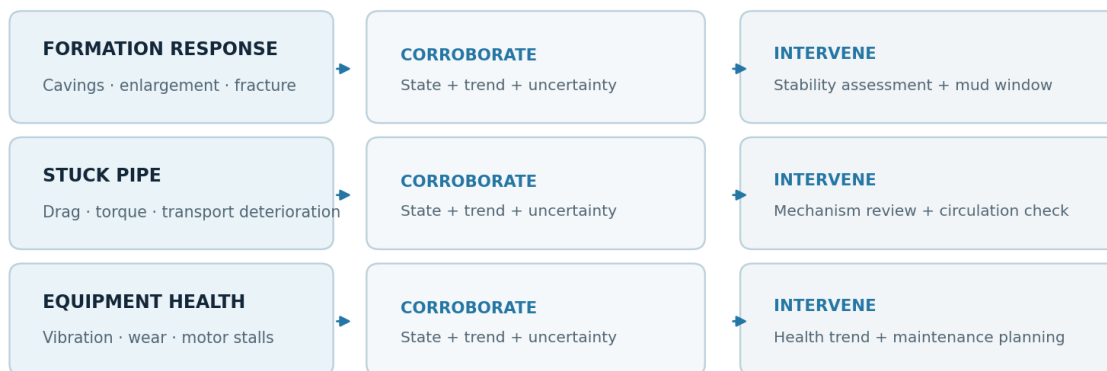
The second stage assigns signals to mechanistic families. Geomechanical context informs instability risk; movement and hydraulic indicators inform sticking; vibration and duty-cycle evidence inform equipment condition. Model outputs should be combined only after their intended targets and prediction horizons have been specified. A single composite risk score can conceal disagreement between mechanisms. Instead, the operator should see separate event hypotheses with evidence, confidence, trend, expected lead time and recommended diagnostic checks. An interpretable accident-forecasting study demonstrates the importance of making local model explanations intelligible to drilling engineers [19].

The third stage introduces graduated response. Low-confidence anomalies warrant observation and sensor verification; corroborated precursors trigger engineering review; imminent validated hazards require action under established operating and well-control procedures. AI must not bypass mandatory barriers or autonomously authorise hazardous set-point changes. Decision support should remain subordinate to the rig's approved safety systems and competent personnel, with clear escalation ownership. Early detection architectures have already been explored for multiple downhole drilling problems [18, 21]; the unresolved challenge is to connect predictions reliably to decisions without creating alarm fatigue.

Finally, every recommendation should generate a traceable event record. The system records model version, data quality, warning time, reasoning, user acknowledgement, intervention and resulting operating condition. Trend analysis can distinguish a persistent precursor from a transient operational mode change [30]. Such records are essential for measuring causal benefit: avoided downtime cannot be inferred from the number of alarms issued. Figure 1 presents the proposed decision pathway, while Figure 2 depicts the separate pathways by which formation, sticking and component hazards contribute to rig interruption.

Real-time engineering imposes deadlines that conventional machine-learning benchmarks seldom measure. If a model requires more time to process a window than the warning horizon permits, its offline accuracy is irrelevant to the immediate decision. Telemetry buffering, communications outages and version updates can introduce latency or silently change features. A practical rig deployment should specify the maximum permitted signal age, minimum reliable sensors and behaviour during partial outages. Fail-safe operation means reverting to established procedures when the advisory layer loses integrity, not extrapolating confident predictions from missing measurements. The interface should retain an audit trail for why an alert was suppressed, escalated or manually overridden. Such governance also supports incident learning: repeated false positives may arise from identifiable activities, whereas a sudden deterioration in sensitivity may indicate instrument recalibration or a new geological regime. Monitoring these failure modes is part of maintaining the predictive system itself.

Distinct precursor pathways converging on rig downtime



Only an independently audited reduction in avoidable downtime demonstrates operational benefit.

Figure 2. Different precursor families and their corresponding decision pathways (original conceptual synthesis).

8 Validation, saudi implementation and governance

Deployment should start with a retrospective baseline in which downtime is classified by verified cause, planned activity and operational context. A useful pilot then runs prospective shadow predictions without influencing crew decisions, measuring timeliness, signal availability, calibration and false alarms. Only after safety review should recommendations be introduced on selected rigs. Comparisons should account for well architecture, interval difficulty, contractor practices, formation type and changing drilling objectives. Random record-level splits cannot substitute for testing on wells or time periods withheld from model development [3, 20].

Performance should be reported as an integrated scorecard rather than one accuracy figure. Statistical discrimination includes event recall, precision, calibrated probabilities and false alerts per rig-day. Operational indicators include actionable lead time, detection-to-acknowledgement delay, intervention frequency, downhole tool replacements, stuck-pipe recovery hours and preventable NPT by root cause. Financial evaluation should subtract monitoring costs, engineering time and unnecessary intervention from credible avoided downtime.

Saudi localisation adds several requirements. Operators need formation-specific retraining or recalibration, careful controls for proprietary geological data, reliable rig-to-office connectivity and documented responsibilities across drilling, geomechanics, fluids and reliability teams. Models should be tested across carbonate and non-carbonate intervals rather than assumed transferable. Training must enable personnel to challenge erroneous predictions and recognise sensor faults. Cybersecurity, audit logs and access restrictions form part of deployment readiness because compromised information can cause unsafe recommendations even when the predictive model is statistically sound.

The literature also reveals an important evidence gap. Published models frequently predict an observable anomaly, whereas the managerial objective is avoiding expensive interruption. Demonstrating the latter requires a counterfactual design, preferably comparing similar intervals or rigs under consistent outcome definitions and recording whether a proposed action was actually taken. A mature programme will retain the ability to suspend or roll back model deployment when calibration

drifts or operational behaviour changes. The resulting system is best understood as monitored engineering decision support, not a replacement for well-control competence.

Economic evaluation can be expressed as the net difference between expected downtime and intervention expenditure across comparable conditions. However, an avoided event is intrinsically difficult to observe: after the crew acts, the counterfactual trajectory remains unknown. Simple before-and-after comparisons can be distorted by changing geology, improved crew experience, bit technology or operational schedules. A prospective design should prespecify eligible drilling intervals, matching variables, exclusions and the definition of preventable hours. Sensitivity analysis can then test whether an apparent saving persists under alternative classifications of disputed events. The audit should include hours lost through unnecessary shutdowns triggered by false warnings, engineering verification time and model support costs. It should also report unintended effects, including postponed maintenance or increased vibration exposure, rather than treating all reductions in recorded downtime as beneficial. These safeguards matter because the ultimate purpose is safer, more reliable delivery of wells, not the optimisation of a reporting metric. The proposed staged validation is summarised in Table 2.

Table 2. Proposed staged validation and allowable evidential claims.

Evidence level	Appropriate evaluation	Permitted claim
Historical records	Held-out wells, label quality, calibration	Retrospective discrimination
Shadow-mode pilot	Alarm load, latency, prospective lead time	Operational feasibility
Controlled advisory roll-out	Crew decisions, intervention logs, safety audit	Decision usefulness
Matched longitudinal assessment	Cause-coded NPT and net economics	Measured downtime impact

9 Discussion and future research

The strongest conclusion from the reviewed evidence is that NPT reduction is a sequential achievement. Reliable signals must first be acquired, an emerging mechanism identified, a feasible response selected and the result measured. Each link can fail independently. A technically sophisticated classifier adds little value if it detects an event after practical intervention has become impossible. Conversely, a moderately accurate but well-calibrated warning may be valuable where the crew can verify its explanation rapidly and intervene conservatively.

Future research should compare physics-constrained and purely data-driven detectors under identical held-out-well protocols; quantify decision lead times under realistic telemetry interruptions; and investigate robust transfer learning between Saudi geological provinces. Rare-event synthesis should retain physically meaningful precursor trajectories rather than create arbitrary positive samples. Equipment studies should connect predicted wear states with verified retrieval inspections and subsequent reliability. Research should also test how explanation formats, alarm burden and shared accountability influence human response, since a prediction that is not acted upon cannot create avoided downtime.

A further methodological priority concerns external validity. Models frequently learn rig-specific control practices as readily as physical precursors. Differences in top-drive configuration, sensor calibration, driller behaviour and casing programme can therefore appear as geological effects if the evaluation does not separate them. Multi-rig prospective studies should document these variables explicitly and measure performance decay after deployment. When an event is rare, reporting

uncertainty around detection sensitivity is more informative than quoting a seemingly exact percentage from only a handful of incidents. Negative results should also be preserved: cases in which the warning arrived too late, lacked a plausible physical explanation or generated unnecessary corrective work reveal which assumptions require revision. Transparent evaluation may initially show modest savings, but its findings are more transferable than an impressive isolated retrospective score.

10 Conclusions

AI-enabled drilling surveillance offers credible opportunities to recognise stuck-pipe precursors, unstable wellbores, circulation losses and component deterioration earlier than conventional isolated thresholds. Yet published detection performance is heterogeneous and usually insufficient to establish realised Saudi NPT savings. The proposed framework therefore couples mechanistic interpretation, calibrated uncertainty, operational-state awareness and engineering approval with explicit post-event measurement. Its practical contribution is an evidence-based basis for prioritising pilot deployments and assessing whether warnings lead to appropriate, timely interventions. Success should be defined by safer decisions and independently measured reductions in avoidable downtime rather than model accuracy alone.

References

- [1] Xu K, et al. Application of machine learning in wellbore stability prediction: A review. *Geoenergy Science and Engineering*. 2024;232:212409. <https://doi.org/10.1016/j.geoen.2023.212409>
- [2] Kamgue Lenwoue AR, et al. *Recent advances and challenges of the application of artificial intelligence to predict wellbore instabilities during drilling operations*. SPE Drilling & Completion; 2023. <https://doi.org/10.2118/215830-PA>
- [3] Mopuri KR, et al. Early sign detection for the stuck pipe scenarios using unsupervised deep learning. *Journal of Petroleum Science and Engineering*. 2022;208:109489. <https://doi.org/10.1016/j.petrol.2021.109489>
- [4] Elahifar B, Hosseini E. Machine learning algorithm for prediction of stuck pipe incidents using statistical data: Case study in Middle East oil fields. *Journal of Petroleum Exploration and Production Technology*. 2022;12:2019–2045. <https://doi.org/10.1007/s13202-021-01436-3>
- [5] Brankovic A, et al. Data-driven indicators for the detection and prediction of stuck-pipe events in oil and gas drilling operations. *Upstream Oil and Gas Technology*. 2021;7:100043. <https://doi.org/10.1016/j.upstre.2021.100043>
- [6] D'Amicis S, et al. Stuck pipe prediction from rare events in oil drilling operations. *Upstream Oil and Gas Technology*. 2023;11:100096. <https://doi.org/10.1016/j.upstre.2023.100096>
- [7] Siqueira VSM, et al. Expert system for early sign stuck pipe detection: Feature engineering and fuzzy logic approach. *Engineering Applications of Artificial Intelligence*. 2024;127:107229. <https://doi.org/10.1016/j.engappai.2023.107229>
- [8] Zhu S, et al. Intelligent prediction of stuck pipe using combined data-driven and knowledge-driven model. *Applied Sciences*. 2022;12:5282. <https://doi.org/10.3390/app12105282>
- [9] Aseel A, et al. Analysis of pipe sticking due to wellbore uncleanliness using machine learning. *Heliyon*. 2023;9:e22366. <https://doi.org/10.1016/j.heliyon.2023.e22366>
- [10] Al Dushaishi MF, et al. Data-driven stuck pipe prediction and remedies. *Upstream Oil and Gas Technology*. 2021;6:100024. <https://doi.org/10.1016/j.upstre.2020.100024>
- [11] Pang H, et al. Lost circulation prediction based on machine learning. *Journal of Petroleum Science and Engineering*. 2022;208:109364. <https://doi.org/10.1016/j.petrol.2021.109364>

- [12] Wood DA, Mardanirad S, Zakeri H. Effective prediction of lost circulation from multiple drilling variables: A class imbalance problem for machine and deep learning algorithms. *Journal of Petroleum Exploration and Production Technology*. 2022;12:83–98. <https://doi.org/10.1007/s13202-021-01411-y>
- [13] Alsaihati A, et al. Application of machine learning methods in modeling the loss of circulation rate while drilling operation. *ACS Omega*. 2022;7:20696–20709. <https://doi.org/10.1021/acsomega.2c00970>
- [14] Wu J, et al. Artificial neural network prediction of wellbore stability in offshore shallow formations. *Geoenergy Science and Engineering*. 2024;213322. <https://doi.org/10.1016/j.geoen.2024.213322>
- [15] Elkatatny S, Al-Abduljabbar A, Abdelgawad K. A new model for predicting rate of penetration using an artificial neural network. *Sensors*. 2020;20:2058. <https://doi.org/10.3390/s20072058>
- [16] Alsaihati A, Elkatatny S, Gamal H. Rate of penetration prediction while drilling vertical complex lithology using an ensemble learning model. *Journal of Petroleum Science and Engineering*. 2022;208:109335. <https://doi.org/10.1016/j.petrol.2021.109335>
- [17] Zhong R, Salehi C, Johnson R. Machine learning for drilling applications: A review. *Journal of Natural Gas Science and Engineering*. 2022;108:104807. <https://doi.org/10.1016/j.jngse.2022.104807>
- [18] Elmgerbi A, Thonhauser G. Holistic autonomous model for early detection of downhole drilling problems in real-time. *Process Safety and Environmental Protection*. 2022;164:418–434. <https://doi.org/10.1016/j.psep.2022.06.035>
- [19] Gurina E, et al. Making the Black-box brighter: Interpreting machine learning algorithm for forecasting drilling accidents. *Journal of Petroleum Science and Engineering*. 2022;218:111041. <https://doi.org/10.1016/j.petrol.2022.111041>
- [20] Tunkiel AT, Sui D, Wiktorski T. Impact of data pre-processing techniques on recurrent neural network performance in context of real-time drilling logs in an automated prediction framework. *Journal of Petroleum Science and Engineering*. 2022;208:109760. <https://doi.org/10.1016/j.petrol.2021.109760>
- [21] Wang C, et al. Downhole working conditions analysis and drilling complications detection method based on deep learning. *Journal of Natural Gas Science and Engineering*. 2020;81:103485. <https://doi.org/10.1016/j.jngse.2020.103485>
- [22] Ye S, Song X, Blunt MJ, Bijeljic B. Domain adaptation neural network based prediction of real-time drill bit tooth wear. *Geoenergy Science and Engineering*. 2024;213041. <https://doi.org/10.1016/j.geoen.2024.213041>
- [23] Agostini CE, Sampaio MA. Probabilistic neural network with Bayesian-based, spectral torque imaging and deep convolutional autoencoder for PDC bit wear monitoring. *Journal of Petroleum Science and Engineering*. 2020;193:107434. <https://doi.org/10.1016/j.petrol.2020.107434>
- [24] Wang X, et al. Automated classification of drill string vibrations using machine learning algorithms. *Geoenergy Science and Engineering*. 2024;239:212995. <https://doi.org/10.1016/j.geoen.2024.212995>
- [25] Boukredera FS, et al. Enhancing the drilling efficiency through the application of machine learning and optimization algorithm. *Engineering Applications of Artificial Intelligence*. 2023;126:107035. <https://doi.org/10.1016/j.engappai.2023.107035>
- [26] Saadeldin R, Gamal H, Elkatatny S. Machine learning solution for predicting vibrations while drilling the curve section. *ACS Omega*. 2023;8:35822–35836. <https://doi.org/10.1021/acsomega.3c03413>
- [27] Aljohar FE, Contreras Otaivora WB. *Enhancing drilling equipment reliability: Deep learning for predicting failure time and real-time anomaly detection*. ADIPEC Technical Paper, SPE-221928-MS; 2024. <https://doi.org/10.2118/221928-MS>
- [28] Zomarah B, et al. *Early warning method to detect mud motor stallings incidents in drilling operations using machine learning*. International Petroleum Technology Conference, IPTC-24183-MS; 2024. <https://doi.org/10.2523/IPTC-24183-MS>
- [29] Al-Mamoori HN, Tian J, Ma H. Stuck pipe detection in oil and gas drilling operations using deep learning autoencoder for anomaly diagnosis. *Applied Sciences*. 2025;15:5042. <https://doi.org/10.3390/app15095042>
- [30] Sui D, Sahebi H. Trend and dynamic analysis on temporal drilling data and their data-driven models. *Geoenergy Science and Engineering*. 2023;211530. <https://doi.org/10.1016/j.geoen.2023.211530>