AI-DRIVEN PREDICTIVE MAINTENANCE

Development of an AI-Driven Predictive Maintenance Framework for Enhancing Rotating Equipment Reliability in Oil and Gas Operations

Muhammad Nadeem

Published November 2026

Abstract

Rotating equipment availability governs throughput, process stability, and maintenance exposure in oil and gas operations. Yet predictive maintenance literature often separates fault classification from intervention planning and extrapolates laboratory results to industrial assets. This integrative review proposes an AI-driven framework that links condition signals, operational context, diagnostic inference, uncertainty-aware prognostics, and risk-constrained work scheduling. We critically synthesise peer-reviewed studies published primarily during 2020–2025 across pump applications, rotating machinery diagnostics, reliability modelling, and digital twins. The synthesis indicates that multisource sensing can improve fault discrimination, while regime variation, limited failure labels, maintenance-induced censoring, and weak external validation restrict operational credibility. We propose a layered architecture with sensor-quality gates, physics-informed features, calibrated model outputs, decision thresholds, and engineer approval before work execution. Evaluation focuses on lead time, calibration, false-alarm burden, reliability, and economic consequences rather than classification accuracy alone. No original industrial dataset or experimentally validated benefit is claimed. The framework offers a testable basis for oil and gas maintenance pilots and identifies evidence required before scaling.

Keywords: centrifugal pumps; compressors; condition monitoring; decision support; machine learning; oil and gas operations; predictive maintenance; reliability engineering; remaining useful life; rotating machinery

1 Introduction

Oil and gas facilities depend on a wide range of rotating equipment, including centrifugal pumps, process compressors, turbines, motors, fans, and associated bearings, seals, and couplings. A defect in one element may propagate through a process train or trigger protective shutdowns even when the component itself remains repairable. Maintenance therefore has two distinct objectives: detecting evolving damage and choosing an intervention that preserves process safety and availability. Machine learning can contribute to both, but a classifier reporting an abnormality does not automatically translate into a maintenance policy. The latter must account for failure consequences, repair windows, spare parts, duty cycles, and uncertain prognosis.

The relevant literature spans several partially disconnected traditions. Optimisation research treats inspection intervals, replacement policies, and cost trade-offs rigorously [1]. Digital twin reviews emphasise continuously updated representations but reveal inconsistent definitions of operational fidelity [2]. Pump-focused research synthesises artificial intelligence across diagnosis and prognosis [3]. These bodies of work together suggest that useful predictive maintenance requires a defensible transition from signals to decisions, not simply another high-performing model.Oil and gas processing facilities create consequential conditions for such translation. Pump cavitation can be

confused with bearing damage when vibration is evaluated without suction or flow context; compressors experience changing thermodynamic states; maintenance records describe repaired assets rather than uninterrupted run-to-failure histories. Hazardous-area controls, change management, operational technology segregation, and turnaround schedules additionally constrain deployment. The proposed framework therefore prioritises interpretability, auditable decisions, and conservative fallback rules. Its applicability is defined through these operational requirements and a proposed site-specific validation agenda, not through unsupported claims that different facilities share identical failure distributions.

2 Aim of the study

The study aims to develop a technically credible, evidence-informed framework that enables oil and gas operators to convert heterogeneous rotating-equipment condition data into validated, risk-sensitive maintenance decisions, while distinguishing analytical plausibility from demonstrable operational benefit.

3 Research objectives

1. Objective one is to compare sensing and feature-engineering approaches against the physical failure mechanisms of pumps, turbines, compressors, and their auxiliary components: bearings, seals and couplings.
2. Objective two is to assess fault detection, diagnosis, and remaining useful life modelling under scarce labels and shifting operating regimes.
3. Objective three is to specify an integrated, auditable decision architecture that combines prediction, maintenance criticality, and schedule feasibility.
4. Objective four is to define prospective evaluation measures for reliability, safety, workload, and lifecycle economics.
5. Objective five is to identify the governance arrangements and research evidence needed to implement multi-asset solutions across oil and gas operations.

4 Research questions

- **RQ1:** Which measurements and contextual variables provide physically meaningful evidence for rotating-equipment degradation?
- **RQ2:** Under what data conditions do statistical, machine-learning, deep-learning, and hybrid prognostic approaches deliver credible advantages?
- **RQ3:** How should uncertain predictions be translated into maintenance priorities without weakening engineering safeguards?
- **RQ4:** What field-validation design can establish improvement beyond model accuracy?
- **RQ5:** Which organisational controls are necessary for responsible deployment across oil and gas facilities?

5 Review methodology

The study adopted an integrative critical-review design because the research problem spans heterogeneous diagnostic experiments, implementation studies, and maintenance optimisation theory

rather than a single comparable intervention. Search concepts were structured for Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, and publisher databases, with supplementary bibliographic chaining. Representative expressions combined ("predictive maintenance" OR "condition monitoring" OR "remaining useful life") AND ("rotating machinery" OR pump OR compressor OR bearing) AND ("machine learning" OR "deep learning" OR "digital twin"), supplemented by oil and gas processing, fault diagnosis, sensor fusion, domain adaptation, and maintenance scheduling terms.

The primary time window was 2020–2025; a small number of earlier articles were retained for specific methodological context. Eligible sources were peer-reviewed journal articles with identifiable equipment, models, diagnostic tasks, reliability reasoning, or deployment implications. Exclusions comprised promotional reports, unverifiable bibliographic records, nontechnical opinion pieces, and results without a discernible methodological basis. We assessed records by title and abstract relevance, then reviewed the full text where accessible. The review did not maintain a reproducible screening log or exhaustive database export; accordingly, no search yield, exclusion count, or PRISMA-compliant comprehensiveness claim is made.

Quality appraisal prioritised data origin and diversity, leakage-resistant train–test partitioning, comparison baselines, operating-regime coverage, error costs, external validation, and transparency about limitations. We grouped evidence thematically into sensing, fault inference, prognostics, digital integration, and operational optimisation. We compared divergent results by task definition and evaluation conditions rather than pooling them into a misleading common accuracy estimate. Thirty published articles form the cited evidence base, while the framework itself is an original conceptual synthesis. Limitations of the search include restricted access to some complete texts, possible publisher coverage bias, English-language emphasis, and inability to claim exhaustive retrieval.

During synthesis, studies were treated as evidence for narrowly specified claims. Pump field studies informed sensor and labelling challenges; digital-twin publications informed architecture and simulation boundaries; optimisation research informed work-scheduling trade-offs; broader rotating-machine research informed model families rather than site-specific effectiveness. This separation reduces inappropriate transfer of laboratory success to safety-critical operations. Where a publication reported strong diagnostic performance without prospective maintenance outcomes, we categorised its findings as predictive feasibility rather than proven reliability benefit.

6 Thematic literature review and critical analysis

The literature supports a layered view of the maintenance problem. Failure mechanisms are physical, signatures are measured, algorithms infer hidden states, and operators decide on action. Treating these layers as interchangeable overstates what an artificial intelligence model can prove. Table 1 contrasts the main evidence families and identifies decision-relevant caveats that should guide industrial experimentation.

Table 1. Comparative evidence synthesis for proposed oil and gas deployment.

Evidence stream	Strength for oil and gas operations	Limitation requiring mitigation
Process-aware pump diagnostics [4–6]	Real operating signals; hydraulic context	Task-specific labels and shifts across duty cycles
Machine learning and spectral analysis [14,20–25]	Complex fault features; multiple model options	Laboratory conditions; weak external transfer

Evidence stream	Strength for oil and gas operations	Limitation requiring mitigation
Hybrid and explainable inference [7–8]	Scarce-label support; inspectable rules	Simulation discrepancy; rule calibration
Prognostics and digital twins [12–13,16,27]	Horizon estimates; virtual scenarios	Censoring and twin-fidelity uncertainty
RAM and maintenance policy [1,17]	Links decisions to availability and cost	Sensitive to failure distributions and constraints

6.1 Sensing and failure-mechanism specificity

Pressure transients can reveal hydraulic phenomena not uniquely observable in broadband vibration. Tiwari and colleagues demonstrated deep-learning discrimination of blockage and cavitation from dynamic pressure signals [4]. Mohammed investigated field pump measurements and highlighted temperature and pressure as influential predictors [5]. The two studies support process-aware monitoring, but neither shows that the same feature ranking holds across all industrial pump services. Process liquid, operating envelope, suction conditions, and sensor placement influence how signals map to failure mechanisms.

Real water-injection pump research by e Souza and colleagues used operational data, compared neural and conventional classifiers, and identified labelling as a material challenge [6]. Their reported precision-recall trade-off illustrates why seemingly encouraging classification results can still miss operationally important events. For industrial equipment, therefore, interpret vibration spectra alongside speed, load, suction and discharge pressure, flow, temperature, lubricant state, and work history. Document sensor provenance and acquisition frequency explicitly; otherwise, model drift may reflect instrumentation changes rather than mechanical deterioration.Signals also require fault-specific interpretation. Bearing outer-race defects, looseness, misalignment, and cavitation produce different spectral structures but can co-occur. A compressor approaching surge demands another monitoring logic than a centrifugal pump operating at low flow. The framework therefore avoids a universal health score inferred from arbitrary raw channels. Instead, equipment-family models use a common governance architecture while retaining physically distinct feature definitions and alarm envelopes.Sensor fusion is valuable only when channels contribute complementary evidence. A temperature rise may indicate lubrication distress, yet also follow a higher process load. An elevated vibration statistic may reflect imbalance, transient cavitation, loose supports, or a sensor mounting issue. Interpretability consequently begins with operating-regime conditioning and a documented causal hypothesis. For variable-speed machinery, evaluate order tracking and speed-normalised features alongside conventional spectral peaks. Analysts should determine whether each channel is continuous, measured periodically, or sampled only after an abnormal observation. These acquisition differences affect what the model can legitimately learn and whether a retrospective evaluation resembles real-time detection.

6.2 Diagnostic models: capability versus transportability

Domain adaptation combined with digital-twin representations offers one strategy for limited labelled pump faults [7]. However, synthetic-to-real transfer depends on whether the simulation reproduces the defect, noise, boundary conditions, and instrumentation of actual service. Explainable fuzzy systems show an alternative, interpretable path for uncertain pumping-system assessments [8]. These approaches are complementary: simulation can broaden learning coverage, whereas rule or fuzzy components can make escalation criteria inspectable by maintenance engineers.

Broad industrial reviews document a proliferation of classifiers and sensing arrangements [9], while connected asset-management studies show that information interoperability is a deployment prerequisite [10]. Oil and gas predictive analytics reviews also identify diverse tasks and incompatible performance conventions [11]. Thus, claims that one architecture outperforms another require matching data, prediction horizon, fault taxonomy, and cost assumptions. A field model with modest balanced accuracy but stable alert lead time may be more valuable than a benchmark model trained on narrowly separated healthy and damaged examples.

In the recommended hierarchy, establish mechanically derived thresholds and simple statistical baselines before trialling gradient boosting, support vector machines, or convolutional networks. Deep models are justified when large, representative waveform histories make learned representations advantageous. Reject them when their gains disappear under equipment-wise or forward-time holdout. Every classification result must retain a mapping to a specific actionable fault hypothesis, not merely an opaque abnormality label.

Fault labels should also distinguish the initiating mechanism from the observed symptom and the completed repair. For example, a bearing replacement recorded after severe vibration is not proof that the bearing initiated the event. Maintenance logs may reflect precautionary work performed during broader shutdowns. Training directly on such labels can teach scheduling habits instead of physical degradation. A defensible adjudication process combines vibration analysis, inspection photographs, oil analysis, dismantling reports, and technician review. Disagreements should be retained as uncertainty rather than silently converted into hard classes.

6.3 Prognostics and uncertainty

Remaining useful life estimates differ fundamentally from fault labels because they concern time under future operating and maintenance conditions. Wu and colleagues survey deep prognostic methods and their data requirements [12]. Contemporary digital-twin research extends the discussion to model updating and virtual–physical alignment [13], while broader fault-diagnosis synthesis highlights generalisation challenges [14]. These studies motivate intervals and confidence statements rather than deterministic replacement dates. A single point estimate can mislead planners when assets are maintained before observed failure, resulting in censored survival times.

Deep convolutional models have demonstrated feasibility for rotating assemblies in other demanding environments [15], yet transferring them to oil and gas compressors requires regime stratification and equipment-specific calibration. Maintenance optimisation within digital twins formally links state estimates to policies [16]. Reliability, availability, and maintainability analysis supplies complementary system-level criticality measures [17]. Critically, expected remaining life and the economically optimal replacement time are not equivalent. The preferred action depends on failure consequence, repair duration, inspection options, and availability of redundancy.

Prospective models should estimate survival probability over specified decision horizons, measure interval coverage, and report how often failures occur earlier than lower-bound predictions. For small failure samples, hierarchical sharing across similar equipment may help, provided component differences and plant boundaries are represented explicitly. Whenever an asset operates outside validated conditions, the system should withdraw precise prognoses and revert to conservative engineering rules. This fallback is a reliability feature, not evidence of analytical failure.

An explicit decision horizon prevents the common confusion between near-term alerts and long-term life prediction. Forecasting failure within the next operational week supports standby activation and

urgent inspection; estimating a broad lifetime distribution supports spares planning and turnaround coordination. These tasks require different loss functions and validation designs. When maintenance replaces a component before failure, the observation is right-censored rather than a failure. Treating replacement time as the failure time systematically distorts hazard estimates and can make schedules look artificially precise.

The interpretation of predictive intervals deserves particular caution. Statistical coverage established on historical assets does not guarantee coverage after equipment upgrades or process changes. Stratify prediction errors by duty regime and remaining horizon, and evaluate coverage of high-risk lower bounds separately from central estimates. A survival forecast may be broad yet still useful if it distinguishes assets needing immediate review from those eligible for planned inspection. Narrow but miscalibrated intervals create false confidence and can be more dangerous than transparently indicating that evidence is insufficient for quantitative forecasting.

6.4 Data engineering and digital integration

Preprocessing choices materially influence predictive performance. Bekar and colleagues describe industrial data cleaning and analysis as integral to maintenance modelling [18]. Coupled simulation and artificial intelligence likewise expose dependence on changing operating conditions [19]. Vibration-analysis synthesis shows how filtering, decomposition, frequency features, and subsequent classification form a connected methodological chain [20]. Consequently, the proposed system stores raw measurements and transformation metadata, maintains sensor calibration history, aligns timestamps, and preserves process context before generating derived features.

Researchers have demonstrated advanced signal decomposition and neural fault diagnosis on rotating machinery [21]. A related convolutional long short-term memory approach tackles prognostic sequences [22]. Recurrence plots and lightweight convolutional models provide another representation strategy [23], whereas variational mode decomposition paired with probabilistic networks exemplifies structured feature extraction [24]. These investigations expand methodological options, but no single representation establishes field readiness. Tests should examine processing latency, missingness, noise sensitivity, variation in rotating speed, and robustness during start-up and shutdown.

Feature ranking with support vector machines offers a comparatively transparent benchmark [25]. Evidence syntheses of artificial intelligence applications recognise organisational data scarcity as a recurring implementation constraint [26]. Reviews of predictive maintenance with digital twins identify the challenge of generating credible degradation data where preventive action interrupts observed failure trajectories [27]. Earlier deep-learning prognostic surveys show the historical development of representation methods [28]. These findings point toward an interoperable design with distinct historian, analytic, recommendation, and enterprise-workflow layers rather than a monolithic black-box application.

Industrial asset data commonly originate from systems built for distinct purposes. A vibration monitor may register high-frequency waveforms, a distributed control system may store aggregated process tags, and an enterprise maintenance platform may preserve technician narratives. Joining these records demands common equipment hierarchies, synchronised clocks, and explicit event-window definitions. Model developers must prevent future maintenance information from entering predictors intended for earlier decision times. Every data split should respect temporal order and, where possible, hold out complete equipment units to estimate transportability.

Maintenance records often omit the reason why measurements were taken. A vibration reading collected after an operator noticed unusual noise cannot be treated as an unbiased routine sample. When abnormal states are disproportionately recorded, estimated defect prevalence and alert rates become distorted. Deployment studies should document collection policies and consider selection bias before model training. Similarly, an apparent improvement from adding a sensor channel may disappear after correcting for differences in sampling frequency or availability. These considerations argue for a prospective data dictionary and measurement plan before accepting any retrospective machine-learning benchmark as the basis for investment.

6.5 Structural health and legacy evidence

Machine-learning relationships between dynamic response and physical damage reinforce the importance of interpretable features and careful material-specific calibration [29]. Earlier motor-vibration forecasting experiments demonstrate the promise of neural predictions yet are not substitutes for prospective field evidence [30]. This distinction matters because rotating equipment includes coupled mechanical, electrical, and process failure mechanisms. Laboratory fault signatures may be cleaner, fault prevalence may be artificially balanced, and simulated deterioration may not reproduce maintenance interruptions. Therefore, model selection must be governed by the decision context and its uncertainty rather than by publication novelty or isolated numerical scores.

7 Proposed AI-driven predictive maintenance framework

Figure 1 presents a proposed information and decision flow. First, field sensors and historians contribute vibration, acoustic, pressure, temperature, speed, flow, and maintenance-event records. Second, a quality-control layer checks timestamps, operating status, sensor drift, and invalid observations. Third, engineering features and equipment-family models estimate fault likelihood, condition progression, and survival probability over a defined planning horizon. Fourth, a decision engine combines model uncertainty with failure criticality, available redundancy, production constraints, and repair readiness. Fifth, approved work orders and actual inspection findings feed back into the evidence base.

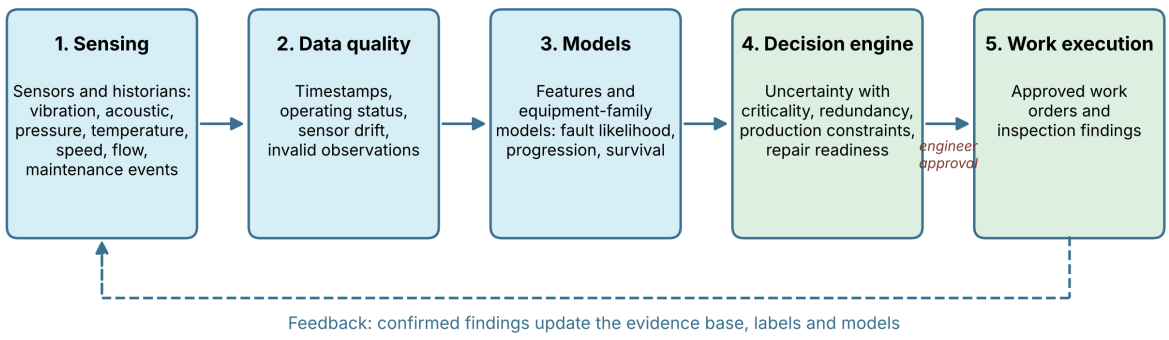


Figure 1. Proposed five-stage AI maintenance architecture with engineer-controlled feedback.

The distinction between inference and authorisation is deliberate. A high model score does not authorise shutting down a train or deferring a mandated inspection. Designated operations and reliability personnel must explicitly approve proposed recommendations; safety-critical alarms retain priority over algorithmic optimisation. Each decision should record the model version,

evidence window, operating regime, uncertainty interval, applicable constraints, approving role, and resulting action. This allows investigation when a recommendation later proves inappropriate.

A staged deployment is preferred. The initial phase observes equipment and produces retrospective alerts without altering work orders. The second phase validates prospective alerts against inspections and failure records. The third phase allows advisory scheduling within defined limits; organisations should only consider automated workflow generation later. A staged approach lets organisations distinguish model improvement from changes in maintenance practice and prevents unwarranted transfers between facility units.

The governance layer should require periodic review of false alarms, missed events, regime drift, and unapproved model changes. A better headline score does not automatically justify recalibration if the revised model changes alert timing or disproportionately increases the burden on a critical maintenance crew. A versioned model registry, rollback capability, threshold-change log, and minimum evidence requirements create a controlled engineering lifecycle. When a sensor becomes unreliable, outputs should be explicitly downgraded to unknown rather than silently imputed into a reassuring condition state.

8 Discussion: translating prediction into reliability

Figure 2 expresses the governing causal proposition: predictive signal quality can lead to earlier recognition, which may improve planning and reduce preventable failures, but only when interventions are feasible and correctly executed. Failure to close the action loop leaves diagnostic accuracy disconnected from uptime. This matters most in continuous process production, where a spare pump, compressor bypass, inspection opportunity, or scheduled turnaround can determine whether an alert has practical value.

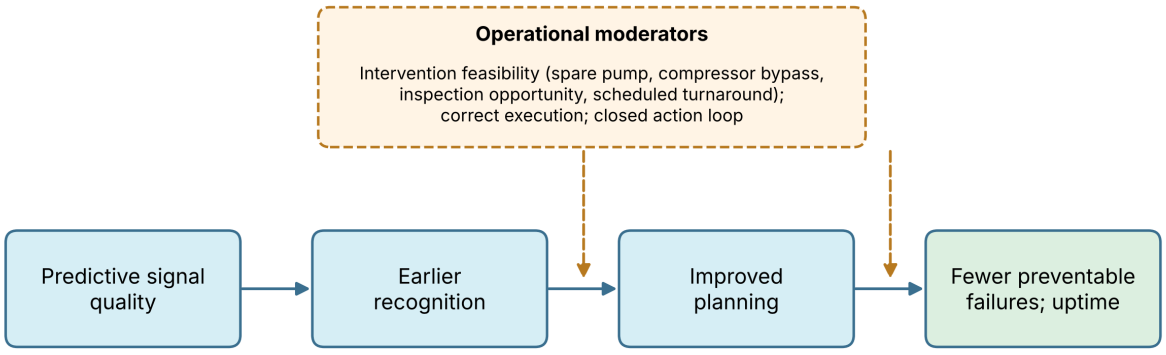


Figure 2. Conditional pathway from measurements to plant outcomes, with operational moderators.

Maintenance thresholds should reflect asymmetric consequences. A false negative on a critical compressor bearing may have major safety and production consequences, while repeated false positives add labour, spares, and unnecessary intervention risks. Thresholds therefore differ by asset criticality and can vary with failure mode. A useful decision policy can minimise expected lifecycle loss subject to hard safety requirements. It must also account for intervention-introduced hazards, such as incorrect assembly, contamination, or infant mortality of replacement parts.

The evaluation should distinguish model, decision, and plant outcomes. Model measures include time-dependent precision, recall, horizon-specific discrimination, survival calibration, and interval coverage. Decision measures include actionable lead time, work-order acceptance, inspection yield, alarm burden, and recommendation reversals. Plant measures include unplanned downtime, mean time between functional failures, maintenance cost, and production loss, adjusted for operating intensity and shutdown exposure. Comparisons against a clearly specified existing policy are essential a favourable before–after difference alone does not establish causation when throughput, equipment age, or overhaul strategy changes concurrently.

Scheduling logic should respect the difference between asset-level and train-level reliability. Replacing an individual bearing early can improve its apparent failure probability yet impose a shutdown that increases total plant exposure. Conversely, grouping work may reduce shutdown frequency while leaving some components in service for longer. The optimum depends on redundancy, isolation arrangements, permissible work scopes, and the uncertainty of component condition. A transparent decision record should show which constraints drove the recommendation and whether a less intrusive inspection could provide enough information to postpone replacement safely.

An alert should be evaluated against the earliest moment at which corrective action would have been technically possible. A classifier that recognises damage only minutes before an automatic trip provides little planning value despite potentially excellent event classification. Conversely, an earlier warning may support lubricant analysis, standby equipment preparation, or targeted inspection without immediate replacement. Researchers must document how they computed alert lead times, whether work order durations were realistic, and which detected events they could have prevented. This operational definition of usefulness also clarifies why you should not compare different prediction horizons using a single aggregate score. Evaluation reports should identify the denominator used for each failure and alarm rate. Maintenance actions require traceable links between detected symptoms and verified physical causes. Interpret condition indicators within documented speed, load, and process regimes. Independent validation should measure missed defects alongside unnecessary inspections and repair costs. A technically sophisticated forecast remains provisional until engineers demonstrate its operational relevance. Comparative evaluations need consistent horizons, equipment populations, and intervention definitions.

Table 2 summarises prospective outcome measures that separate prediction from impact.

Table 2. Prospective outcome measures separating prediction from impact.

Evaluation level	Proposed measures	Interpretation safeguard
Analytical	Horizon recall, precision, calibrated risk, RUL interval coverage	Time-forward and asset-holdout validation
Decision	Useful alert lead time, inspection yield; alert workload, override rate	Record rejected as well as accepted recommendations
Reliability	Unplanned outage hours, functional failure frequency, repair duration	Adjust for operating hours, duty, and turnaround periods
Economic and safety	Avoided disruption, intervention cost, risk-constraint violations	Compare with existing policy, never monetise unsafe delay

9 Research gaps and future research

The foremost gap is external validation under diverse oil and gas operating conditions. Available reports often evaluate a single pump, bearing dataset, or laboratory fault family, leaving uncertainty about transferability across lubricants, process streams, climates, and machine configurations. Future studies should prospectively compare multiple units and test entirely unseen equipment. Cross-site evidence should preserve commercial confidentiality while publishing enough feature definitions, split rules, and evaluation logic for meaningful replication.

A second gap is uncertainty calibration under intervention-driven missing data. Researchers often remove failed components, while preventive repairs never reveal their counterfactual failure dates. Future work should combine censored survival analysis, regime-aware representation learning, and transparent uncertainty quantification. A third gap concerns linking diagnostic outcomes with scheduling and process safety. Researchers should measure whether predictions alter maintenance decisions and whether those decisions improve reliability after adjusting for confounding process conditions.

A fourth gap is cyber-physical assurance. Data pathways joining operational technology and enterprise systems create integrity and availability concerns. Research should test how the model responds to sensor dropout, spoofed values, clock desynchronisation, and interrupted historian access. Finally, human-factors studies should examine how reliability engineers interpret competing risk measures, challenge model recommendations, and document overrides. Human monitoring cannot compensate for systematic technical errors unless the interface makes limitations visible.

Future trials should preregister primary outcomes and evaluation windows before operators see the results. A shadow deployment can first establish whether alerts precede independently verified defects. A later phased rollout, potentially across comparable equipment groups, can examine work-order conversion and operational outcomes. Reporting should include cancelled alerts, non-actionable detections, missing data episodes, and interventions performed without model prompting. These less flattering cases are central to estimating true field value and avoiding selective presentation of successful predictions.

Intervention effects should be included when interpreting reported gains. If a team acts on a warning and finds no defect, the observation might represent a successful early intervention, a false alarm, or an inspection with limited sensitivity. Independent engineering adjudication is needed to distinguish these explanations. Future field protocols could use anonymous reviews of condition traces and maintenance photographs to separate model influence from ground-truth assessment. Such designs would improve evidence quality without requiring deliberate exposure of critical equipment to failure.

10 Practical, managerial and policy implications

Reliability and maintenance managers should begin with an asset criticality register and select pilot equipment groups that combine consequential failure modes with adequate condition history. Choose sensors for fault mechanisms rather than installing them to maximise data volume. Reliability and maintenance teams need consistent failure codes, symptom definitions, repair timestamps, and links between equipment identifiers and operating histories. This foundational work often determines whether subsequent modelling is meaningful.

Procurement requirements should demand access to model input definitions, validation results, retraining controls, uncertainty outputs, and audit trails. Reliability and operations groups should jointly define allowable recommendations, escalation routes, and conditions requiring deterministic shutdown rules. Workforce development should focus on interpreting calibrated probabilities and false-alarm trade-offs, not merely operating dashboards. Management reporting should use absolute impacts against agreed baselines and should not convert hypothetical simulations into claimed realised savings.

At the policy level, industrial digitalisation initiatives can encourage common equipment taxonomies, secure data-sharing arrangements, and independent evaluation protocols without prescribing a single proprietary algorithm. Oil and gas operators can align experimental governance with their own process-safety, cybersecurity, data-residency, and corporate engineering requirements. The proposed framework is not a substitute for applicable standards or site approvals. Its contribution is to provide a structure for generating evidence that may support maintenance policy revisions.

A realistic business case should separate implementation expense from recurring operating costs. Instrument calibration, historian retention, network segmentation, software support, specialist review, and retraining all affect lifecycle economics. Calculate benefits from verified disruption avoidance rather than multiplying predicted failures by an assumed shutdown value. Sensitivity analyses should vary fault prevalence, labour availability, false-alarm cost, and planned maintenance opportunities. This approach lets management identify when inexpensive rule-based monitoring remains preferable to complex models.

Implementation responsibilities should be unambiguous. Instrumentation specialists own sensor integrity; data engineers own timestamp reconciliation and lineage; model custodians own versioning and drift checks; reliability engineers own fault interpretation; operations personnel own safe plant decisions. If responsibilities overlap without clear escalation, a technically correct prediction may never yield an executable work order. An operating procedure should prescribe decision response times and identify the authority empowered to stop, defer, or modify proposed action. Regular joint reviews can reveal whether observed shortcomings originate in sensing, modelling, planning, or execution rather than treating every problem as an algorithm defect.

11 Limitations

This article develops a conceptual framework through integrative evidence synthesis; it does not present field measurements, a trained site-specific model, or independently observed reductions in downtime. Different studies use incompatible failure labels and evaluation horizons, preventing defensible meta-analysis. The cited set is deliberately thematic rather than an exhaustive systematic sample. In addition, access limitations and the lack of a retained search screening database prevent quantification of publication-selection bias. Local teams must establish equipment-specific reliability thresholds, cybersecurity architectures, and cost parameters before deployment. These restrictions prevent treating the proposed design as a validated implementation in any particular oil and gas facility.

The operating context is a proposed application environment rather than a measured explanatory variable. Oil and gas facilities differ in age, vendor mix, maintenance maturity, data architecture, and process configuration. Consequently, a uniform performance claim would be inappropriate. The paper also does not assess the regulatory adequacy of a particular installation, and references to

engineering safeguards describe proposed governance principles rather than certification of a specific software product.

12 Conclusion

The review establishes that an AI-driven maintenance programme for rotating equipment in oil and gas operations must integrate physical failure reasoning, trustworthy sensing, calibrated diagnosis and prognosis, and safety-constrained work execution. Pump-specific studies support process-aware classification, while digital-twin and deep-learning research offers routes for richer modelling; neither substitutes for prospective field validation. The proposed architecture makes uncertainty, asset criticality, engineer approval, and feedback from completed maintenance explicit design requirements. Its central contribution is an evaluation-ready connection between analytical inference and auditable decisions, rather than a claim of measured performance. Future multi-site oil and gas pilots should determine whether the framework reduces actionable risk, alarm burden, and production losses compared with existing practices while preserving process-safety obligations.

References

[1] de Jonge B, Scarf PA. A review on maintenance optimisation. *European Journal of Operational Research*. 2020;285:805–824. https://doi.org/10.1016/j.ejor.2019.09.047

[2] Errandonea I, Beltrán S, Arrizabalaga S. Digital twin for maintenance: a literature review. *Computers in Industry*. 2020;123:103316. https://doi.org/10.1016/j.compind.2020.103316

[3] Aliyu R, Mokhtar AA, Hussin H. Prognostic health management of pumps using artificial intelligence in the oil and gas sector: a review. *Applied Sciences*. 2022;12:11691. https://doi.org/10.3390/app122211691

[4] Tiwari R, Bordoloi DJ, Dewangan A. Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm. *Measurement*. 2021;173:108676. https://doi.org/10.1016/j.measurement.2020.108676

[5] Mohammed A. Data driven-based model for predicting pump failures in the oil and gas industry. *Engineering Failure Analysis*. 2023;145:107019. https://doi.org/10.1016/j.engfailanal.2022.107019

[6] e Souza ACO, de Souza MB, da Silva FV. Development of a CNN-based fault detection system for a real water injection centrifugal pump. *Expert Systems with Applications*. 2024;244:122947. https://doi.org/10.1016/j.eswa.2023.122947

[7] Kumar A, Kumar R, Xiang J, Qiao Z, Zhou Y, Shao H. Digital twin-assisted AI framework based on domain adaptation for bearing defect diagnosis in the centrifugal pump. *Measurement*. 2024;235:115013. https://doi.org/10.1016/j.measurement.2024.115013

[8] Upasane SJ, Hagras H, Anisi MH, Savill S, Taylor I, Manousakis K. A type-2 fuzzy-based explainable AI system for predictive maintenance within the water pumping industry. *IEEE Transactions on Artificial Intelligence*. 2024;5:490–504. https://doi.org/10.1109/TAI.2023.3279808

[9] Çınar ZM, Nuhu AA, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. *Sustainability*. 2020;12:8211. https://doi.org/10.3390/su12198211

[10] Cheng JCP, Chen W, Chen K, Wang Q. Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*. 2020;112:103087. https://doi.org/10.1016/j.autcon.2020.103087

[11] R Azmi PA, Yusoff M, Mohd Sallehud-din MT. A review of predictive analytics models in the oil and gas industries. *Sensors*. 2024;24:4013. https://doi.org/10.3390/s24124013

[12] Wu F, Wu Q, Tan Y, Xu X. Remaining useful life prediction based on deep learning: a survey. *Sensors*. 2024;24:3454. https://doi.org/10.3390/s24113454

[13] Ma H, Fan C, Zhang Y, Wang Q, Yu K, Ma Z. Digital twin-inspired methods for rotating machinery intelligent fault diagnosis and remaining useful life prediction: a state-of-the-art review and future challenges. *Mechanical Systems and Signal Processing*. 2025;232:112770. https://doi.org/10.1016/j.ymssp.2025.112770

[14] Vashishtha G, Chauhan S, Sehri M, Zimroz R, Dumond P, Kumar R, Gupta MK. A roadmap to fault diagnosis of industrial machines via machine learning: a brief review. *Measurement*. 2025;242:116216. https://doi.org/10.1016/j.measurement.2024.116216

[15] Apeiranthitis S, Zacharia P, Chatzopoulos A, Papoutsidakis M. Predictive maintenance of machinery with rotating parts using convolutional neural networks. *Electronics*. 2024;13:460. https://doi.org/10.3390/electronics13020460

[16] Gosavi A, Le VK. Maintenance optimization in a digital twin for Industry 4.0. *Annals of Operations Research*. 2024;340:245–269. https://doi.org/10.1007/s10479-022-05089-1

[17] Clavijo Mesa MV, Patino-Rodriguez CE, Guevara Carazas FJ, Gunawan I, Droguett EL. Asset management strategies using reliability, availability, and maintainability (RAM) analysis. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. 2021;43:495. https://doi.org/10.1007/s40430-021-03222-y

[18] Bekar ET, Nyqvist P, Skoogh A. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. *Advances in Mechanical Engineering*. 2020;12:1687814020919207. https://doi.org/10.1177/1687814020919207

[19] Crespo Márquez A, Crespo Del Castillo A, Gómez Fernández JF. Integrating artificial intelligent techniques and continuous time simulation modelling: practical predictive analytics for energy efficiency and failure detection. *Computers in Industry*. 2020;115:103164. https://doi.org/10.1016/j.compind.2019.103164

[20] Bagri I, Tahiry K, Hraiba A, Touil A, Mousrij A. Vibration signal analysis for intelligent rotating machinery diagnosis and prognosis: a comprehensive systematic literature review. *Vibration*. 2024;7:1013–1062. https://doi.org/10.3390/vibration7040054

[21] Gao Y, Ahmad Z, Kim JM. Fault diagnosis of rotating machinery using an optimal blind deconvolution method and hybrid invertible neural network. *Sensors*. 2024;24:256. https://doi.org/10.3390/s24010256

[22] Gao Y, Ahmad Z, Kim JM. The prediction of the remaining useful life of rotating machinery based on an adaptive maximum second-order cyclostationarity blind deconvolution and a convolutional LSTM autoencoder. *Sensors*. 2024;24:2382. https://doi.org/10.3390/s24082382

[23] Shi Y, Wang H, Sun W, Bai R. Intelligent fault diagnosis method for rotating machinery based on recurrence binary plot and DSD-CNN. *Entropy*. 2024;26:675. https://doi.org/10.3390/e26080675

[24] Li Z, Zou C, Chen Z, Lu H, Xie S, Zhang W, He J. Research on the fault diagnosis method of rotating machinery based on improved variational modal decomposition and probabilistic neural network algorithm. *Applied Sciences*. 2024;14:7380. https://doi.org/10.3390/app14167380

[25] Huynh HH, Min CH. Rotating machinery fault detection using support vector machine via feature ranking. *Algorithms*. 2024;17:441. https://doi.org/10.3390/a17100441

[26] Scaife AD. Improve predictive maintenance through the application of artificial intelligence: a systematic review. *Results in Engineering*. 2024;21:101645. https://doi.org/10.1016/j.rineng.2023.101645

[27] van Dinter R, Tekinerdogan B, Catal C. Predictive maintenance using digital twins: a systematic literature review. *Information and Software Technology*. 2022;151:107008. https://doi.org/10.1016/j.infsof.2022.107008

[28] Wang Y, Zhao Y, Addepalli S. Remaining useful life prediction using deep learning approaches: a review. *Procedia Manufacturing*. 2020;49:81–88. https://doi.org/10.1016/j.promfg.2020.06.015

[29] Fleet T, Kamei K, He F, Khan MA, Khan KA, Starr A. A machine learning approach to model interdependencies between dynamic response and crack propagation. *Sensors*. 2020;20:6847. https://doi.org/10.3390/s20236847

[30] Scalabrini Sampaio G, Vallim Filho ARA, Santos da Silva L, Augusto da Silva L. Prediction of motor failure time using an artificial neural network. *Sensors*. 2019;19:4342. https://doi.org/10.3390/s19194342