ARTIFICIAL INTELLIGENCE

AI-Driven Predictive Maintenance of Industrial Fire Protection Machinery: A Reliability-Based Engineering Framework for Saudi Manufacturing Facilities

Talha Umar Khan, Waheed Mahboob Jilani

Published November 2026

Abstract

Industrial fire protection machinery has a very particular reliability requirement in that it has to be capable of performing when required, even though it will be in standby for the vast majority of its life. The conventional approach of scheduled inspection is still absolutely vital, though periodic testing may well miss the gradual degradation that occurs between inspections. This paper is a critical review of the potential for AI to augment rather than supplant the existing inspection and reliability engineering practices for fire pumps, drivers, valves, controllers and the associated suppression system in the Saudi manufacturing environment. It is a structured, interpretive synthesis of the peer-reviewed literature published between 2020 and 2025 on predictive maintenance, fire protection equipment, fault diagnosis, physics-informed prognostics and risk based decision making. The proposed contribution is a demand aware engineering framework which links operational signals, proof test results, component degradation, uncertainty and consequence severity to auditable maintenance decisions. Particular emphasis is placed on low frequency failures, environmental exposure, the unequal consequences of various types of failure and the difference between diagnostic accuracy and demonstrated system availability. The review identifies tangible opportunities in the area of multi-modal monitoring, hybrid reliability models and explainable maintenance prioritisation, whilst also highlighting the continuing problems with data portability, causal attribution, model calibration and field validation. A staged evaluation programme is recommended prior to any safety critical deployment. The key point is that prediction only becomes of operational value when it improves the evidence of readiness on demand and does not in any way compromise the requirement for mandatory proof testing or human responsibility.

Keywords: predictive maintenance; fire protection systems; fire pumps; reliability engineering; artificial intelligence; Saudi Arabia

1 Introduction

Manufacturing plants are reliant on protective equipment whose value is not really apparent in the normal production metrics. You have fire pumps, diesel drivers, motor controllers, pressure maintenance assemblies and distribution valves which may be completely idle for months and yet are absolutely critical at the point of ignition or when the suppression system is called upon. The engineering objective of this equipment is entirely different to that of the production equipment. A failed conveyor belt will be obvious in that it stops the output, a dormant protective piece of equipment may degrade without any observable service impact at all. The defect may not be apparent until you do a test or you actually have an emergency. The maintenance policy therefore has

to take account of the degradation that can occur whilst the equipment is in a stand by state and the probability that the entire protective chain will work when required.

Industrial predictive maintenance research has moved on from threshold alarms to statistical learning, degradation modelling and digitally connected decisions. Reviews of the field identify progress in sensing and fault classification but also point to inconsistent evaluation methodologies, limited industrial transferability and a lack of consideration of the actual deployment environment [1–6]. Digital twin work offers a route to integrate equipment data with virtual models but this is hampered by issues of fidelity, computational cost and incomplete data histories [7–9]. The problems become considerably more acute when the predictive output is being used to make decisions about safety systems rather than simply about production availability.

The key characteristic of fire protection machinery is the interplay between demand, consequence and evidence. A pump may produce perfectly acceptable vibration signatures during a brief test, yet be unable to meet the required hydraulic duty when the full demand envelope is applied. An abnormal acoustic reading is not necessarily indicative of a loss of emergency capability. Diagnosis, future degradation and system level protection adequacy all have to be assessed in an analytical manner. Fire pump testing research illustrates why direct hydraulic performance evidence is so important [16] and fault tree analysis shows how sprinkler reliability is a function of the individual components and their interdependencies [15]. Machine learning results in no way replace these physical relationships.

Saudi manufacturing is a relevant case in which to consider the application of the approach as the environment can be expected to encompass extreme ambient temperatures, airborne dust, variable water chemistry, a mix of equipment generations and a geographically dispersed maintenance team. These are legitimate engineering stressors rather than actual national failure rates as such (the literature reviewed doesn't provide any Saudi specific failure rates). The local implementation would therefore require site specific measurements, documented duty cycles and validation across the range of installation types rather than simply importing model scores from another industry.

This paper is about how AI derived indicators can be turned into defensible decisions for machines that are required to be available in the face of rare events. It proposes an architecture which maintains the required inspection, explicitly quantifies the uncertainty and ranks the failures in terms of the safety function they impact. It's a critical engineering synthesis and a testable conceptual framework, not a claim of having done any field experiments.

2 Aim, objectives and review methodology

The study is aimed at producing an evidence-based interpretation of AI-enabled predictive maintenance for industrial fire protection machinery in the Saudi manufacturing environment with readiness on demand as the key engineering outcome. The analysis is driven by four objectives: to be able to distinguish between hidden failures and visible degradation; to assess what can be learned from the intermittent and continuous data streams using current learning methods; to be able to combine equipment level predictions with system level reliability and consequence; and to be able to define the validation requirements prior to any maintenance decision being made on the basis of a predictive alert. These are very different objectives from the usual model comparison studies where the primary metric of success is the accuracy of the model in isolation.

A critical review methodology (rather than a quantitative meta-analysis) was adopted. The evidence base comprises publications from 2020 to 2025 and was gathered via targeted combinations of keywords relating to fire pumps, sprinkler reliability, industrial predictive maintenance, machine learning, remaining useful life, condition monitoring, digital twins and risk based maintenance. Publisher records and DOI linked scholarly metadata were used to verify the identity and date of the articles. The three provided review papers were used to guide the structure of the review, the way in which the comparisons were made and the level of explanation required; they were not in any way used as a substitute for the subject specific evidence. There was no full export from a bibliographic database and so no claim is made to the extent of the retrieval or to a statistically pooled treatment effect.

Eligibility was based on conceptual relevance and methodological contribution. Direct evidence was that relating to fire pump performance or suppression system reliability. Transferable engineering evidence was that relating to industrial rotating machinery, hydraulic equipment, uncertainty aware prognosis, maintenance optimisation or intelligent safety management. Papers which were solely about wildfire spread, general image based flame recognition or unrelated production optimisation were not considered as evidence of fire pump readiness. The sources were assessed in terms of asset class, measured signals, target outcome, validation environment, known limitations and the plausibility of the transfer pathway. This is a qualitative evidence mapping which deliberately draws the line between direct evidence and analogies which still require validation.

The synthesis was guided by four appraisal questions. First, does the paper actually assess protective function or merely a surrogate (e.g. vibration or fault class accuracy)? Second, are the operating states, test conditions and failure definitions adequately defined? Third, does the validation approach prevent leakage between adjacent time windows, units or maintenance periods? And fourth, does the proposed decision rule take into account uncertainty, inspection obligations and the differing consequences of a false reassurance and an unnecessary intervention? The last of these is of particular importance as a model which fails to detect a rare dormant failure will appear to be highly accurate on a dataset which is overwhelmingly composed of healthy equipment [2,3,21].

The review is of course bound by constraints. We have a focused set of 30 DOI identifiable papers to work from, and they are all of very different designs and populations (they are not a census of all the work). The results reported for bearings, oilfield injection pumps and general industrial assets are used to inform the methodological choices rather than to suggest that the same level of effectiveness would be achieved in the Saudi fire protection installations. The recommendations that follow are all engineering propositions which would require prospective evaluation.

3 Reliability characteristics of protective machinery

A fire protection installation is a coupled system rather than a single pump. The success of the system in responding to a fire can be dependent upon the availability of stored water, the suction conditions, the position of the isolation valves, the starting energy, the controller logic, the health of the prime mover, the pump hydraulics and the downstream distribution. Common cause disturbances can affect more than one component at the same time. The fault tree is therefore a useful tool for identifying the various combinations of failures which can result in the failure of a required protective function. Bahrami and Givehchi [15] show how the uncertainty in the individual sprinkler system failure contributions can be made explicit rather than simply being absorbed into a single reliability figure.

Failure mechanisms are also state dependent. In standby you can have corrosion, battery degradation, seal degradation, fouling, inadvertent isolation or latent control faults all of which will have very limited dynamic signatures. During start-up you can see electrical transients, insufficient cranking energy and switching faults. During the period of sustained running you have cavitation, misalignment, bearing wear, overheating and inadequate flow/pressure performance all of which are much more accessible to the sensors. The same parameter can therefore have very different diagnostic significance in different states. Time stamped state labels must be attached to every learning input otherwise the classifier will be liable to confuse a test condition with an actual fault.

Inspection frequency presents another major analytical problem. Testing makes latent faults visible but each scheduled test also alters the population under test through mechanical and thermal cycling. An apparent increase in faults with increased test frequency may simply be a result of improved detection rather than an actual increase in degradation. On the other hand, longer test intervals can lead to very 'healthy' databases as the unseen faults are not present in the labels. The evidence models need to be able to separate the date on which a fault arose, the date on which it became detectable and the date on which it was actually discovered. Without that separation the predicted failure horizons will be systematically influenced by the inspection policy rather than the actual state of the asset.

Reliability on demand is not the same as MTBF. Failure probability on demand is about whether the protection will operate when it's required to do so (i.e. will it trip when it should?). Operating reliability is about the performance over the duration of the required response. Availability also takes into account planned maintenance, repair and equipment isolation. These are all distinct concepts when you're looking at a proposed alert. A pump may have an excellent predicted bearing life and yet be unavailable simply because the controller is in a disabled state. A good driver can do nothing about inadequate suction pressure. Risk based maintenance research supports the principle of prioritising maintenance according to the consequences of failure rather than treating all the abnormalities in the same way [14].

The recent fire pump investigation has moved a bit closer to the measurement issue. Zhang et al. [16] present a number of linked methods for the assessment of fire pump performance; Nguyen [17] looks at intelligent fault classification for firefighting pump maintenance. These are all positive steps in terms of technical feasibility but don't actually establish that the learned warning scores will in fact reduce emergency demand failures in a range of manufacturing contexts. Work on industrial pump prognosis also points to the potential of sensor based diagnostics and the difficulties that can arise from the paucity of failure data and the variable nature of the operating environment [18,19].

4 Diagnostic evidence and model selection

Selection of an AI method should be driven by the failure mechanism and the observations available rather than by the architecture itself. Pressure, differential pressure, flow, motor current, speed, bearing temperature and vibration all describe different facets of the fire pump's duty. Acoustic measurements can be used to supplement the mechanical monitoring where direct sensor installation is not feasible (Nguyen and Huang [19] present an approach to water pump fault classification from sound). The presence of an acoustic signature cannot in itself be used to confirm that firefighting flow has been delivered. Hydraulic proof test records must be kept in association with the mechanical data when developing or interpreting a readiness model.

Simple statistical baselines should be given the go-ahead in the initial phase. Control charts, residual analysis, logistic regression and tree ensembles are all capable of flagging deviations from the norm and are reasonably easy for an engineer to understand. Supervised deep learning is of course appealing in situations where you have a lot of labelled faults and well defined duty cycles (as has been demonstrated in the wider field of industrial diagnostics [3,6]). The reality though is that actual protection system failures are rare and scheduled maintenance will break the deterioration trend. Your training data will thus contain a lot of normal starts and very few instances of a failed emergency duty. In that situation high overall accuracy is a very weak indicator of a reliable detection. You need to look at precision, event level recall, calibration and missed demand risk.

The appropriate training objective is similarly asymmetric. False alarms are costly in terms of manpower and testing, whereas a missed critical impairment in a real fire can have far more serious consequences. Cost-sensitive learning or the use of constrained decision thresholds may be the way to deal with this imbalance but the cost weights must be derived from a hazard analysis rather than simply chosen to give the best ROC curve. Reliability diagrams, probability calibration and threshold sensitivity should be presented alongside the standard metrics. For rare events, precision-recall analysis is generally far more informative than a single AUC value; prevalence and the inspection scenario must be stated in conjunction with the performance.

Unsupervised detection is a viable stop-gap solution in the absence of failure labels. The model can learn the normal relationships between speed, pressure, current and temperature during similar test states and then flag any inconsistent patterns for investigation. The anomaly scores are not failure probabilities though. Commissioning changes, seasonal temperatures, sensor replacement and test valve configurations can all lead to distribution shifts without any actual damage having occurred. A shift detection layer, engineering operating bounds and technician feedback are all required before any anomalies can be turned into a maintenance instruction [2,13].

Remaining useful life prediction introduces a number of further assumptions. You need a defined end of life point, observations of progressive degradation and a fairly stable failure mechanism. Modern physics-informed approaches combine data driven estimation with constraints or degradation knowledge in order to improve the degree of conceptual consistency when you would otherwise be making an unphysical prediction [10-12]. For fire pumps such an approach may be quite reasonable for bearings, batteries and seals but would be much less defensible in the case of intermittent control faults. A component level forecast should never be presented as a hard deadline for the complete protection system to fail.

Explainability is relevant when an alert is going to change the maintenance priorities. Multisensor experiments have demonstrated how feature attribution can be used to support fault interpretation [24] and operational decision support research has shown the link between interpretable predictions and spares planning [25,30]. The explanation should tell you the operating state, the baseline deviation, the sensor quality and the recommended verification test. It should not be presented as proof of causality. A correlated vibration and pressure anomaly may both be the result of a single underlying hydraulic disturbance. Tamascelli and co-workers [21] make the same point with respect to confidence assessment when machine learning is being applied to safety and reliability questions.

5 A demand-aware engineering framework

Figure 1 shows the proposed architecture. The first layer is an asset and function register. Each item under monitoring is linked to its protected zone, duty specification, driver config, standby state,

inspection schedule and failure consequences. Equipment genealogy is important as an identical model can behave very differently after an impeller has been replaced, a controller reprogrammed or a sensor moved. This layer is deliberately engineering led; a model trained without any functional context will be likely to prioritise the easily observed defects and will tend to ignore the failures which result in the safety function being lost.

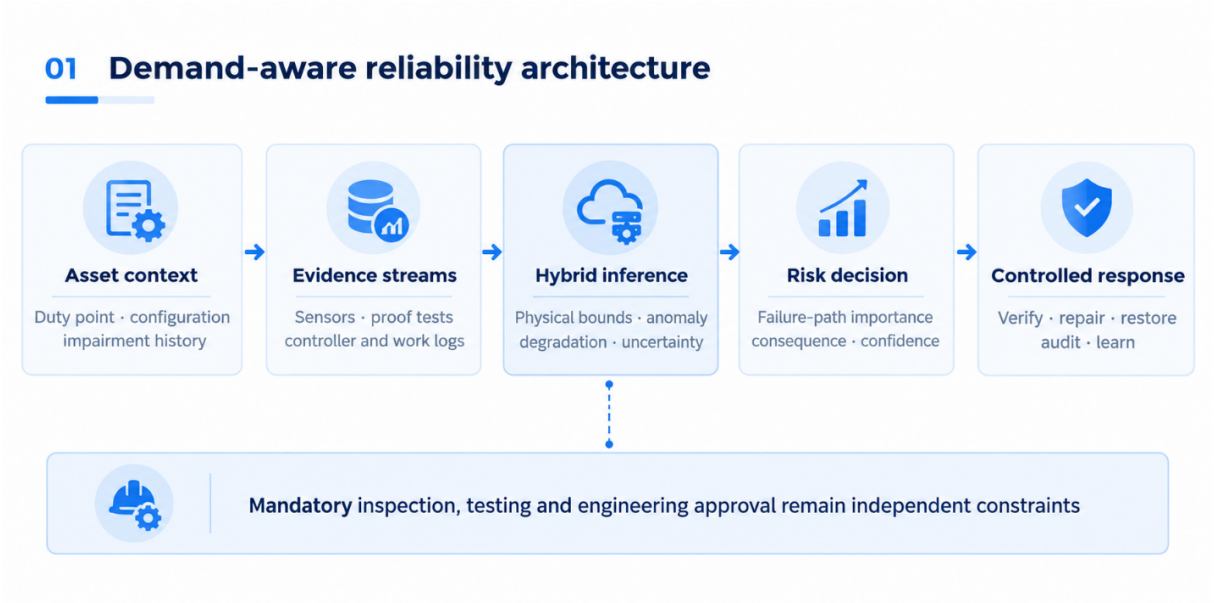


Figure 1. Proposed demand-aware reliability architecture. The diagram represents a conceptual integration, not an empirically validated control system.

The acquisition layer brings together three classes of evidence. The continuous signals relate to environmental exposure and the standby state; the event records relate to start attempts, alarms, trips and maintenance; and the controlled tests provide a relatively short period of hydraulic and electrical performance. The sensor sampling has to be appropriate for the phenomenon in question (slow temperature trends are very different from the transients associated with a motor start or bearing frequency). Time synchronisation, instrument calibration and missing data indicators all form part of the dataset and should not be silently corrected. Connected fire pump testing provides a good real world example of this [16].

The analysis layer consists of a number of complementary modules. Rule based tests look for hard engineering violations (e.g. abnormal pressure, failure to start within the applicable duty criterion). Condition monitoring models assess deviations from expected performance under comparable operating conditions. Mechanistic residuals compare the actual pump performance against a validated reference curve which has been suitably adjusted for the test conditions. Where appropriate a probabilistic degradation model will be used to estimate the probability of component failure over a given planning horizon. Digital representations can be used to support scenario analysis but must reflect the actual instrument coverage and uncertainty rather than suggest perfect synchronisation [7–9,28,29].

The decision layer is where you turn the evidence into a maintenance priority. A simple example of the score would be expected consequence x probability of failure over the relevant interval, with a penalty for low confidence or for incomplete observation. It's a proposed ranking mechanism rather than a calibrated equation in any sense and it's not in any way a substitute for a site specific safety analysis. The consequences to be included would be life safety, operational impact, repair exposure

and potential for escalation. You have to be very careful not to double count failures that are correlated and you mustn't assume that a high machine health score in any way implies that the downstream sprinkler will perform as expected. RCA can identify the dominant failure paths and barriers which in turn would require direct verification [14,15].

There's a separate governance layer in place with hard and fast rules. The statutory and code-based inspection, testing and maintenance activities are all still on the schedule regardless of what the model says. A forecast may well lead to work being carried forward, may result in more diagnostic effort or may assist in the preparation of spares, but it should not in any way be used to defer the required tests. The operator is still in complete control to isolate, record impairment, put in place compensatory measures and restore service in accordance with the approved procedures. The model should in no way be able to initiate a suppression system shutdown or change any of the protective controls. That separation is there to protect against both software errors and from the danger of simply placing too much faith in the probabilistic output.

Model outputs should also contain an evidence provenance record. An alert generated by a recently calibrated pressure transducer during a properly executed full flow test is going to be of very different confidence than one generated by a short no flow start with very incomplete metadata. A succinct evidence record might simply state the relevant duty point, observation window, sensor health, model version and any known limitations. This level of traceability allows a reviewer to reconstruct exactly why an intervention was recommended and whether the supporting evidence actually related to the protected function. It also gives the maintenance team some basis upon which to question a dubious recommendation rather than just accept the algorithm as infallible.

Figure 2 shows the resulting maintenance logic. If an actual functional test fails then the immediate corrective actions take precedence over a positive AI prediction. If there is a verified abnormal trend and high consequence then enhanced testing and prioritised action are warranted. If the indicators are in conflict or the confidence is low then the appropriate response is diagnostic investigation rather than silent acceptance. The only situation in which the prescribed baseline schedule can be maintained is when the situation is stable and well observed. The policy is asymmetric in that unnecessary investigation is a waste of resources but false reassurance can leave a vital protective barrier unattended.

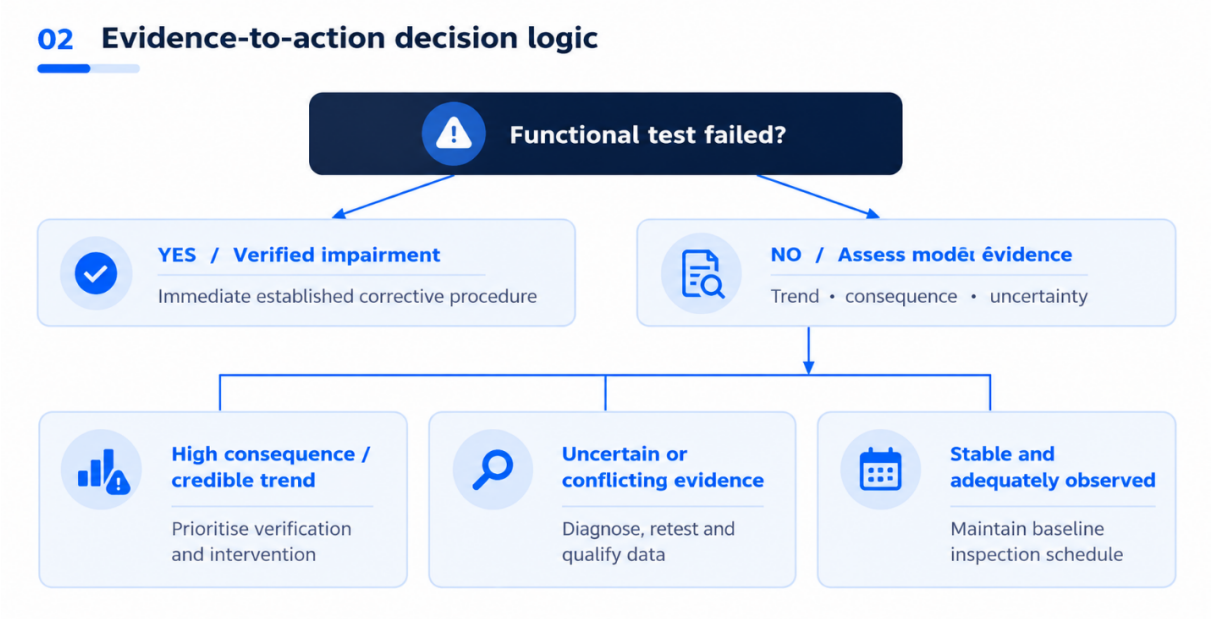


Figure 2. Conceptual evidence-to-action logic. A confirmed functional failure overrides model reassurance; all actions remain subject to established procedures.

Edge processing can be of value in situations where the facilities have connectivity issues or where there is a requirement for local alarm continuity. Bala and co-workers [26] go on to discuss the pros and cons of moving intelligent maintenance functions closer to the equipment. For safety related applications local calculation would need to be coupled with buffered data, secure updates, defined fail safe behaviour and a clearly defined boundary between the advisory analytics and the certified control logic. Model maintenance itself would then have to become a controlled engineering activity with versioning, performance monitoring and the ability to rollback.

6 Validation, uncertainty and performance measures

A proper validation programme must start with a credible reference standard. For fire pumps a simple labelled mechanical fault is not good enough. You need to look at documented starting performance, pressure/flow duty, sustained operation and identified component faults. A technician confirmed cause and a verified test after repair is far better than a label generated by the system itself. Table 1 shows the representative failure modes with respect to observability and the type of evidence that would be appropriate. It also highlights the fact that data from other industries will have to be treated with caution.

Table 1. Function-oriented evidence map for representative fire-protection machinery failure conditions.

Representative condition	Potential observations	Engineering interpretation	Minimum confirmation
Failed start	Start attempt; current; battery voltage; controller events	Driver energy, control circuit, or starting sequence	Witnessed start test; fault isolation and cause verification
Reduced hydraulic duty	Suction/discharge pressure, flow, speed, test valve state	Pump wear, obstruction or changing test configuration	Verified pressure–flow duty test against approved baseline

Representative condition	Potential observations	Engineering interpretation	Minimum confirmation
Bearing degradation	Vibration spectrum; acoustic envelope; temperature	Mechanical damage or changing loading conditions	Physical inspection and repeat measurement under comparable duty
Cavitation or suction loss	Pressure fluctuations; noise; flow and reservoir status	Insufficient inlet conditions; air ingress; obstruction	Check supply, hydraulic conditions and qualified test results
Latent valve impairment	Valve supervision; position records; line pressure	Inadvertent isolation or actuator malfunction	Physical valve inspection and functional confirmation
Controller anomaly	Trip/event logs; supply voltage; command response	Faulty transfer, wiring, sensors or configuration	Qualified control-system inspection and approved functional test
Standby deterioration	Battery, cooling, leakage and environmental history	Corrosion, seal damage, discharged battery or fouling	Targeted maintenance examination and documented restoration

Training and test data must be split by physical unit, maintenance episode and time. Simply randomly distributing adjacent seconds of the same weekly test across training and test sets will result in excessively optimistic estimates as the records will have very similar operating conditions and latent faults. A sensible approach would be to use historical development data, a later time-based holdout and at least one further site or equipment family. Inspection interventions need to be time-stamped so that post-repair measurements are not erroneously treated as part of the continuous degradation. Fernandes and co-workers [3] point out the same sorts of practical problems that exist when moving from fault identification in the lab to actual deployment in the plant.

Evaluation should be based on four distinct types of evidence. Analytical performance covers event level sensitivity, precision and calibrated risk prediction. Temporal value is about whether the alert provides a useful lead time for intervention. Engineering validity is about whether the alert is consistent with pump laws, known failure modes and observed hydraulic duty. Operational benefit is about actual readiness, work order quality, the number of unnecessary interventions and the effect on duration of impairment. Table 2 provides an evaluation matrix which avoids the simplistic reporting of isolated classifier accuracy as a safety outcome.

Table 2. Evaluation matrix for advisory AI maintenance systems; metrics are proposed, not observed results.

Performance dimension	Required measure	Critical risk addressed	Validation requirement
Fault detection	Event-level recall and precision	Missed critical defects; alerts per independent fault	Held-out failure episodes and technician-adjudicated labels
Probability quality	Calibration and prediction intervals	Underconfidence or false reassurance	Reliability diagram and calibration assessment
Lead time	Time from actionable alert to verified defect	Insufficient interval for safe intervention	Prospective work-order and alert chronology
Functional readiness	Successful start and approved hydraulic duty	Failure on demand or insufficient delivery	Witnessed controlled tests with instrument traceability
Maintenance burden	Unnecessary actions and repeat call-outs	Excess labour, testing or equipment disturbance	Work-order audit against baseline regime

Performance dimension	Required measure	Critical risk addressed	Validation requirement
System protection	Impairment duration and failure-path coverage	Unrecognised common causes or downstream unavailability	Asset linkage, fault-tree review and restoration evidence
External transfer	Performance after site or configuration change	Model drift, missing sensors, poor portability	Unit-disjoint and independent-facility validation

Uncertainty needs to be explicitly dealt with at both sensor and prediction levels. Sensor calibration, limited observation windows, ambiguity of operating state, estimation of rare events and site differences are all factors which can erode confidence. Bayesian or interval based predictions can make the uncertainty visible rather than just presenting a point estimate [27]. Physics informed augmentation may be of use in the modelling of degradation when data is sparse, though synthetic examples are not the same as real emergency failures [12]. The proposed risk thresholds will therefore need to be commissioned, have conservative escalation rules and be subject to periodic recalibration against new test data.

Prospective evaluation should start in shadow mode. The system would generate recommendations without altering any existing work orders and engineers would be in a position to assess the relevance of the alerts, the response burden and the rate of failure discovery. Only after independent review would a facility consider a controlled advisory pilot. The primary endpoints should be performance on real test failures, the number of avoidable missed defects and time to verified restoration with follow-up long enough to look at seasonal and operational variation. An increase in the number of anomalies detected is not in itself evidence of improved protection reliability.

Statistical comparisons need to take into account the fact that actual system failures are going to be very rare and that the observations within a facility are going to be correlated. Confidence intervals around sensitivity will still be wide even after the accumulation of thousands of normal test records. Facility level clustering, interval censoring and changing inspection schedules all need to be explicitly addressed when you're attempting to estimate effects. In a controlled pilot you would need prespecified stopping rules to protect against the possibility of repeated false reassurance, the build up of unexplained alert volumes or evidence of software interfering with the normal operation of the process. External reviewers would need to look at the adverse cases in detail, including the alerts that were missed prior to the actual test failure. A robust safety claim will require a sustained body of evidence of non-inferiority in terms of compliance and improvement in verified readiness.

7 Application to Saudi manufacturing facilities

Implementation should begin with a facility level hazard and asset assessment rather than with a one size fits all national model. Petrochemicals, food processing, metal fabrication and industrial warehousing will all have very different water demands, different environmental loads, different staffing arrangements and different consequences of an interruption. The proposed framework is therefore based on configurable functional requirements and a common data dictionary rather than a single predictive threshold being applied across the sites. Facility operators are expected to record the source and approval status of their design and maintenance criteria.

Phase 1 can be focused on electrically driven and diesel driven fire pumps as the controlled test environment will give you repeated, structured observations. The instrumentation should start with pressure (verified), motor/engine state, current/speed, battery condition and test result. Additional vibration or acoustic channels should only be included if there is a specific failure mode that you're expecting to see them in. Sensor placement must not in any way compromise equipment integrity or protection certification. You'll want to have specific data fields for fuel quality, cooling systems, environmental dust and heat exposure where applicable but again the focus should be on testing the causality rather than just collecting data.

A second phase would be the addition of controller events, valve supervision, impairment logs and the relevant water supply conditions. This is a necessary expansion at the system level as a component can be perfectly healthy and yet have demand performance completely defeated by a closed valve or a supply limitation. Any recurrent proof test failures would have to be investigated via fault tree and causal analysis prior to any algorithmic ranking being applied. There is some useful work on the reliability and false signals of fire protection sensors [22,23], though one must be careful not to confuse fire detection algorithms with algorithms for predicting mechanical readiness.

Governance is just as important. There needs to be data stewardship, appropriate access permissions, cybersecurity segregation and where necessary bilingual maintenance instructions. The alert must identify the asset, the symptom, the confidence level, the safety implication and the need for a human check. Every decision must be accompanied by an audit trail which records the technician disposition, the root cause findings and the verification of restoration. Work on human reliability in the context of emergency pump operation makes it very clear that the maintenance processes and the actions of the operator are themselves part of the risk picture [20]. Data governance has to protect both the industrial operations and the integrity of the data that will be used in future model evaluation.

The economic case should be looked at on a full lifecycle basis rather than just the sensor purchase price.

Workforce capability is another factor which will have a bearing on the quality of implementation. The instrument technicians have to be able to spot sensor artefacts, the maintenance planners have to be able to distinguish between an anomaly and a failure forecast and the responsible engineers have to be able to assess the consequence of any proposed isolation. The training exercises should include contradictory evidence, missing sensors and apparently reassuring scores in the face of documented test failures. The escalation paths should be in place for the night shift and for the contracted maintenance staff as much as they are for the centrally staffed sites. In this scenario an intelligible workflow and good record keeping may be worth more in the short term than the deployment of a more computationally complex classifier. The potential benefits would be earlier detection of degrading equipment, fewer repeat visits, better spare preparation and fewer prolonged impairments. The costs would be in the area of calibration, instrumentation, cyber security, model revalidation, false alarms and staff training. There is no defensible national savings figure to be derived from the evidence. Any comparative pilot studies should report on total ownership cost and on the reliability outcomes with uncertainty intervals.

8 Critical synthesis and future research

The three major evidence gaps are: first, the disconnect between the readily available machine learning datasets and actual protection in demand outcomes. A lot of the data available is based on

simulated faults, lab wear or equipment operating under fairly normal duty cycles. Fire protection equipment is very much the opposite in that it has long periods of stand-by, censored deterioration and very high performance requirements which only really come into play during the periodic proof tests. Without the paired test and sensor histories you can't actually establish the link between the predicted component health and the actual state of suppression readiness. The work on intelligent firefighting pump diagnostics is very promising in this area but still very much limited to a narrow scope of system level field validation [16,17].

The second gap is one of reliability integration. Fault prediction gives you an estimate for a given item under a given model, protection is achieved via a network of interdependent functions. Work on risk based maintenance and sprinkler fault trees provides complementary analytical tools [14,15]. Future work would be to combine the learned component states with dynamic fault trees, Bayesian dependency models or reliability block structures and assess the sensitivity of the results to uncertain failure probabilities. Common cause hazards, testing induced wear and maintenance errors all need to be explicitly addressed. A model that appears to be very accurate is of little use if the reliability structure is wrong.

The third gap is in accountability under distribution shift. Differences in ambient conditions, duty configuration, water supply, instrumentation and maintenance practice can all render an imported model invalid. Reporting should therefore include calibration drift, out-of-distribution detection, missing channel tolerance and justifications that engineers can check. Transfer learning and hybrid physics constraints are potential research directions [10–13] but external validation is a prerequisite. The development effort should be focused on simple models that beat the transparent baseline on the real events rather than on ever more complex architectures evaluated on artificial splits.

A practical research agenda arises from the above gaps. Multi-site consortia could define a minimum event schema (in a privacy conscious way) that links asset ID's with event type, start time, flow pressure tests, defects, replacements, impairment duration and consequence class. Prospective studies could compare a fixed inspection regime with advisory prediction against the same inspection regime without any analytics (a study which would not be ethically acceptable if it involved a reduction in the amount of testing). The accumulation of sequential evidence would allow the engineers to update their estimates as more real faults are recorded. An independent assessment should also consider human factors, response times, cyber resilience and the cost of false reassurance.

This agenda also defines what is meant by novelty. The proposed framework is not claiming to have invented condition monitoring, fault trees or digital twins. Its contribution is in the integration of those three elements in answer to a very particular and demanding question of readiness in the face of infrequent emergency demand. You wouldn't demonstrate that integration by just getting a good machine learning benchmark. You'd have to show a demonstrable improvement in verified system capability. That would be a valid research direction to support industrial modernisation without in any way compromising the primacy of safety engineering.

A common benchmark to define component classification, functional duty verification, system level availability and maintenance response so that we can have independent comparative evidence and not have the convenient statistical labels get confused with actual protection outcomes across sites.

9 Conclusions

AI has real potential to enhance the observation and prioritisation of industrial fire protection machinery, especially in the case where connected proof test records can be combined with mechanical and electrical condition data. The best supported application is in the area of fault diagnosis, structured monitoring and improved decision support. There is very limited direct evidence to support a reduction in emergency demand failure probability across a range of manufacturing sites. That said, this is the key distinction when it comes to interpreting the literature appropriately.

The review puts forward a demand aware framework that links functional registers, operating state aware sensing, hybrid diagnostics, uncertainty, reliability structures and human controlled interventions. It views the need for inspection and testing as a continuing assurance mechanism rather than something that can be taken over by an algorithm. Predictions should lead to appropriate verification and confirmed functional failures should trigger the appropriate corrective action irrespective of the model score.

For Saudi manufacturing the successful application will be highly dependent on the specific hydraulic duties at site, the environmental conditions, the heterogeneous nature of the assets and the existence of auditable governance. The next step in the research is a prospective, multi-site evaluation with independent test evidence, calibrated risk estimates and meaningful measures of protective readiness. In the absence of such evidence the framework should be viewed as a very robust design proposition for advisory purposes only and not as a validated autonomous safety system.

References

[1] Zonta T, da Costa CA, Righi RR, de Lima MJ, Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*. 2020;150:106889. https://doi.org/10.1016/j.cie.2020.106889

[2] Leukel J, González J, Riekert M. Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. *Journal of Manufacturing Systems*. 2021;61:87–96. https://doi.org/10.1016/j.jmsy.2021.08.012

[3] Fernandes M, Corchado JM, Marreiros G. Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: A systematic literature review. *Applied Intelligence*. 2022;52:14246–14280. https://doi.org/10.1007/s10489-022-03344-3

[4] Nacchia M, Fruggiero F, Lambiase A, Bruton K. A systematic mapping of the advancing use of machine learning techniques for predictive maintenance in the manufacturing sector. *Applied Sciences*. 2021;11:2546. https://doi.org/10.3390/app11062546

[5] Phan TLJ, Gehrhardt I, Heik D, Bahrpeyma F, Reichelt D. A systematic mapping study on machine learning techniques applied for condition monitoring and predictive maintenance in the manufacturing sector. *Logistics*. 2022;6:35. https://doi.org/10.3390/logistics6020035

[6] Polverino L, Abbate R, Manco P, Perfetto D, Caputo F, et al. Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review. *International Journal of Engineering Business Management*. 2023; 15, article 18479790231186848. https://doi.org/10.1177/18479790231186848

[7] Chen C, Fu H, Zheng Y, Tao F, Liu Y. The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*. 2023;71:581–594. https://doi.org/10.1016/j.jmsy.2023.10.010

[8] 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

[9] Zhong D, Xia Z, Zhu Y, Duan J. Overview of predictive maintenance based on digital twin technology. *Heliyon*. 2023;9:e14534. https://doi.org/10.1016/j.heliyon.2023.e14534

[10] Deng W, Nguyen KTP, Medjaher K, Gogu C, Morio J. Physics-informed machine learning in prognostics and health management: State of the art and challenges. *Applied Mathematical Modelling*. 2023;124:325–352. https://doi.org/10.1016/j.apm.2023.07.011

[11] Li H, Zhang Z, Li T, Si X. A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. *Mechanical Systems and Signal Processing*. 2024;209:111120. https://doi.org/10.1016/j.ymssp.2024.111120

[12] de Beaulieu MH, Jha MS, Garnier H, Cerbah F. Remaining Useful Life prediction based on physics-informed data augmentation. *Reliability Engineering & System Safety*. 2024;252:110451. https://doi.org/10.1016/j.ress.2024.110451

[13] Wu Y, Sicard B, Gadsden SA. Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications*. 2024;255:124678. https://doi.org/10.1016/j.eswa.2024.124678

[14] da Silva RF, Melani AHA, Michalski MAC, de Souza GFM. Reliability and risk centered maintenance: A novel method for supporting maintenance management. *Applied Sciences*. 2023;13:10605. https://doi.org/10.3390/app131910605

[15] Bahrami M, Givehchi S. A novel approach for determining the reliability of sprinkler systems: A case study. *Results in Engineering*. 2023;17:100843. https://doi.org/10.1016/j.rineng.2022.100843

[16] Zhang S, Li Y, Chen X, Zhou R, Wu Z, Zarhmouti T. A novel IoT-based performance testing method and system for fire pumps. *Water*. 2024;16:792. https://doi.org/10.3390/w16050792

[17] Nguyen TP. AIoT fault detection for firefighting pump maintenance services based metaheuristics and combined deep learning methodologies. *Computational Intelligence*. 2025;41:e70071. https://doi.org/10.1111/coin.70071

[18] Almazrouei SM, Dweiri F, Aydin R, Alnaqbi A. A review on the advancements and challenges of artificial intelligence based models for predictive maintenance of water injection pumps in the oil and gas industry. *Discover Applied Sciences*. 2023;5:391. https://doi.org/10.1007/s42452-023-05618-y

[19] Nguyen MT, Huang JH. *Fault detection in water pumps based on sound analysis using a deep learning technique*. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2022;236. https://doi.org/10.1177/09544089211039304

[20] Aydin M, Sezer SI, Arici SS, Akyuz E. Predicting human reliability for emergency fire pump operational process on tanker ships utilising fuzzy Bayesian Network CREAM modelling. *Ocean Engineering*. 2024;314:119717. https://doi.org/10.1016/j.oceaneng.2024.119717

[21] Tamascelli N, Campari A, Parhizkar T, Paltrinieri N. Artificial intelligence for safety and reliability: A descriptive, bibliometric and interpretative review on machine learning. *Journal of Loss Prevention in the Process Industries*. 2024;90:105343. https://doi.org/10.1016/j.jlp.2024.105343

[22] Khan F, Xu Z, Sun J, Khan FM, Ahmed A, Zhao Y. Recent advances in sensors for fire detection. *Sensors*. 2022;22:3310. https://doi.org/10.3390/s22093310

[23] Diaconu BM. Recent advances and emerging directions in fire detection systems based on machine learning algorithms. *Fire*. 2023;6:441. https://doi.org/10.3390/fire6110441

[24] Gawde S, Patil S, Kumar S, Kamat P, Kotecha K. An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion. *Decision Analytics Journal*. 2024;10:100425. https://doi.org/10.1016/j.dajour.2024.100425

[25] Dereci U, Tuzkaya G. An explainable artificial intelligence model for predictive maintenance and spare parts optimization. *Supply Chain Analytics*. 2024;8:100078. https://doi.org/10.1016/j.sca.2024.100078

[26] Bala A, Rashid RZJA, Ismail I, et al. Artificial intelligence and edge computing for machine maintenance—review. *Artificial Intelligence Review*. 2024;57:119. https://doi.org/10.1007/s10462-024-10748-9

[27] Jiang GJ, Yang JS, Cheng TC, Sun HH. Remaining useful life prediction of rolling bearings based on Bayesian neural network and uncertainty quantification. *Quality and Reliability Engineering International*. 2023;39:1756–1774. https://doi.org/10.1002/qre.3308

[28] Fuller A, Fan Z, Day C, Barlow C. Digital twin: Enabling technologies, challenges and open research. *IEEE Access*. 2020;8:108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358

[29] Rasheed A, San O, Kvamsdal T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*. 2020;8:21980–22012. https://doi.org/10.1109/ACCESS.2020.2970143

[30] Rajaoarisoa L, Randrianandraina R, Nalepa GJ, Gama J. Decision-making systems improvement based on explainable artificial intelligence approaches for predictive maintenance. *Engineering Applications of Artificial Intelligence*. 2025;139:109601. https://doi.org/10.1016/j.engappai.2024.109601