

Original Research

Received 07 June 2026

Revised 18 July 2026

Accepted 01 August 2026

Natural Resources for Human Health 2026, Vol. 6 (14s): 846–857

KEYWORDS:

Reinforcement learning; predictive maintenance; healthcare equipment replacement; optimization; risk stratification; Vision 2030.

Optimization Models for Healthcare Equipment Replacement Policies Using Reinforcement Learning

Waleed Mohammed Alzuman¹, Mohamed Yacin Sikkandar^{1*}, Sabarunisha Begum S², Jackson Daniel³, Mohammed Rafiq⁴, Ahmed Farag Salem Babetat⁵, Faleh Ibrahim Alrashdi¹, Majed Atallah Almutairi¹, Fahad Sattan Alrashdi⁶, Khalid Munawir Alharbi⁶, Mashhor Shalwan Nasser Alotaibi⁷

¹ Department of Medical Equipment Technology, College of Applied Medical Sciences, Majmaah University, Al Majmaah 11952, Saudi Arabia.

² Department of Biotechnology, P.S.R. Engineering College, Sivakasi 626140, India.

³ Department of Biomedical Engineering, KIT-Kalaignarkaranidhi Institute of Technology, Coimbatore, Tamilnadu 641402, India

⁴ Public Health Department, College of Applied Medical Sciences, Majmaah University Al Majmaah 11952, Saudi Arabia

⁵ Dr. Mohammad Alfagih Hospital, Riyadh 13223, Saudi Arabia.

⁶ Outpatient Clinic, College of Dental, Majmaah University, Al Majmaah 11952, Saudi Arabia.

⁷ Ministry of Health, Riyadh 14511, Saudi Arabia.

*Corresponding author Email: m.sikkandar@mu.edu.sa

ABSTRACT: Healthcare organizations must reconcile cost, reliability, and patient safety when replacing clinical equipment, a task made harder in Saudi Arabia by Vision 2030's rapid expansion of healthcare infrastructure and by the periodic patient-volume surges of Hajj and Umrah. Fixed-interval and threshold-based replacement rules assume stable operating conditions and cannot adapt to this variability. This study proposes a two-stage decision-support framework that couples supervised failure prediction with a reward-driven reinforcement-learning (RL) replacement policy. A representative predictive-maintenance dataset (124,494 records from 1,169 devices) was pre-processed into 99 engineered features. Logistic Regression, Random Forest, and Gradient Boosting classifiers were trained to estimate failure probability; the best-performing classifier informed a risk-stratification layer and an RL policy weighing maintenance, replacement, and failure costs against equipment uptime. Random Forest achieved the strongest discriminative performance (ROC-AUC = 0.849), and device-level failure history was the most influential predictor of future failure. A moderate risk-based intervention threshold (probability ≥ 0.5) minimized total lifecycle cost (US\$860 per cycle), a 38.6% reduction relative to a reactive baseline (US\$1,400), while an overly aggressive, low-bar threshold (≥ 0.3) increased cost more than eleven-fold by over-triggering replacements. Under the optimal policy, 70.6% of interventions were timely replacements, 17.6% were preventive maintenance, and 11.8% ended in unplanned failure. Coupling predictive analytics with reinforcement learning shifts equipment-replacement decisions from reactive, calendar-driven practice toward adaptive, evidence-based policies, offering measurable cost savings

and reliability gains for Healthcare Technology Management units pursuing Vision 2030's digital-health objectives.

1. INTRODUCTION

Healthcare providers are under mounting pressure to align the replacement of clinical equipment with cost efficiency, operational reliability, and patient safety [1]. Reinforcement learning (RL) has emerged as a data-driven alternative to fixed-interval or threshold-based replacement rules, learning optimal maintenance and replacement policies directly from operational feedback rather than from static assumptions [2,3]. RL agents interact repeatedly with an evolving operating environment and, through reward-guided trial and error, learn when equipment should continue in service, be repaired, or be retired [4]. The approach has already been used to schedule patient care, allocate hospital resources, plan equipment maintenance, and analyze service-utilization patterns to guide policy [1,3,6,7]. By re-expressing the trade-off between maintenance expenditure and operational risk as a reward signal, RL can surface replacement strategies that static heuristics miss [5]. Coupling RL with formal optimization further helps organizations respond to constrained budgets and fluctuating demand [24,25], and deep RL extends this capacity to high-dimensional operational data such as usage frequency, degradation trajectories, and failure probability [17,18].

Historically, hospitals have relied on periodic maintenance schedules or simple threshold rules, both of which assume a largely stationary operating environment [19,20]. These assumptions break down as clinical technology evolves and usage patterns become more variable, leaving conventional lifecycle models increasingly unable to keep pace [13]. Machine learning (ML) models can forecast failures with reasonable accuracy but stop short of prescribing what to do about them, since they do not learn sequential decision policies [8,9]. RL closes that gap: algorithms such as Q-learning, deep Q-networks, and policy-gradient methods have already been applied to predictive maintenance and operational optimization in industrial settings [10,11,23]. In Saudi Arabia specifically, the healthcare sector is undergoing rapid transformation under Vision 2030, with sustained investment in infrastructure, digital health, and advanced clinical technology [22,30]. Many hospitals have established Healthcare Technology Management (HTM) units and adopted asset-tracking and preventive-maintenance programs, yet strategic, data-driven replacement planning remains limited; most current practice is reactive, centered on inspection and repair of existing assets rather than optimized lifecycle decisions [29]. Wide variation in equipment utilization and recalibration cycles, periodic patient-volume surges during Hajj and Umrah, and the compliance requirements of the Saudi Food and Drug Authority (SFDA) [28] make reactive strategies increasingly inadequate, motivating an adaptive, regulation-aware replacement framework that can sustain equipment availability, safety, and cost-effectiveness in support of Vision 2030's healthcare goals.

Prior RL applications in industrial and manufacturing maintenance [12,23,24,25], and the broader theoretical foundations of sequential decision-making [26,27], have not been extended to the specific constraints of hospital asset lifecycles - regulatory compliance, patient safety, and the heterogeneity of medical devices - and reviews of equipment-replacement optimization point to a persistent lack of learning-based, regulation-aware frameworks in this domain [13]. RL in healthcare remains concentrated on treatment optimization and predictive maintenance, with comparatively little work addressing lifecycle replacement, multi-objective trade-offs between cost, reliability, and safety, or deployment-ready decision support aligned with regional regulatory frameworks such as those in Saudi Arabia [14,15,16]. To address this gap, the present study develops and evaluates a decision-support framework that couples supervised failure prediction with a reward-driven RL replacement policy for healthcare equipment. Specifically, the study (i) builds a predictive model of device failure from operational sensor data, (ii) formulates a reinforcement-learning environment that weighs maintenance, replacement, and failure costs against equipment uptime, and (iii) compares the resulting adaptive policy against static intervention thresholds to quantify the cost and reliability benefits of an RL-based approach relative to conventional practice..

2. MATERIALS AND METHODS

2.1 Database Description

Because direct access to hospital maintenance records is restricted by privacy and operational constraints, this study used a publicly available predictive-maintenance dataset whose device-operation characteristics are broadly comparable to hospital-grade imaging systems, ventilators, and laboratory analyzers [21]. The dataset comprises 124,494 daily records from 1,169

devices operated between January and November 2015, each carrying nine sensor metrics and a binary failure indicator. Although drawn from an industrial rather than a clinical setting, its continuous-usage, environmental-stress, and failure patterns make it a suitable proxy for developing and testing the proposed framework ahead of deployment on hospital-specific data.

2.2 Data Pre-processing and Feature Engineering

Missing values were imputed and statistically anomalous sensor readings were removed. Continuous features were scaled to a 0-1 range with min-max normalization to stabilize training, and the data were partitioned into training (70%), validation (15%), and test (15%) subsets. Beyond the nine raw sensor metrics, 90 additional features were engineered - lagged values ($n = 27$), rolling statistics ($n = 20$), first-order differences ($n = 5$), and device-level aggregates ($n = 38$) - for a total of 99 predictors (Table 1, Fig. 1) capturing both the instantaneous state and the progressive degradation trajectory of each device. Class imbalance in the failure label (0.09% positive) was addressed through class weighting during model training.

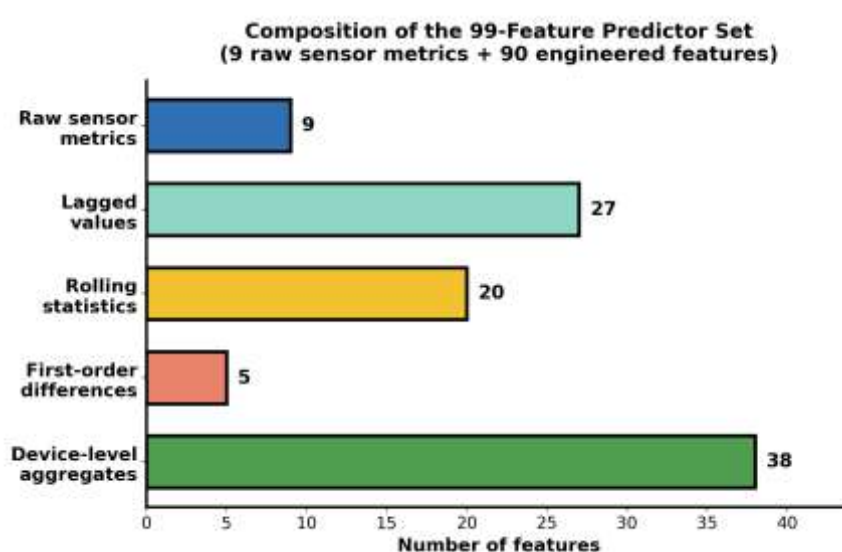


Figure 1. Composition of the 99-feature predictor set.

After cleaning, the working dataset comprised 64,486 records across 1,169 devices, retaining 106 failure events (a 0.09% failure rate) (Table 1). Extensive pre-processing and feature engineering reduced raw record count while substantially enriching informational content, and temporal features (day of week, month, day of year) were added to capture operational patterns.

Table 1. Dataset Summary After Pre-processing

Attribute	Value
Total samples (raw)	124,494
Samples after pre-processing	64,486
Number of devices	1,169
Sensor metrics	9
Engineered features	99
Failure events	106
Failure rate	0.09%

2.3 Predictive Failure Model

Three supervised classifiers spanning increasing structural complexity - Logistic Regression, Random Forest, and Gradient Boosting - were trained on the engineered feature set to estimate the probability that a given device-day record would end in failure. Logistic Regression served as an interpretable linear baseline, expressing failure probability as a sigmoid function of a weighted combination of the 99 predictors. Random Forest, a bagging ensemble of decision trees trained on bootstrapped samples with randomized feature subsets, was included for its ability to capture nonlinear interactions among sensor and temporal features while remaining comparatively robust to overfitting and providing a natural, aggregate measure of feature importance. Gradient Boosting, which builds trees sequentially so that each new tree corrects the residual errors of its predecessors, was included as a higher-capacity alternative capable of modelling more complex decision boundaries, at some cost to interpretability and stability under severe class imbalance.

All three models were trained on the 70% training partition and tuned on the 15% validation partition described in Section 2.2, with the held-out 15% test partition reserved for final, unbiased evaluation. Because failures constituted only 0.09% of records, class weighting was applied during training to penalize misclassification of the minority (failure) class more heavily than the majority (non-failure) class; synthetic oversampling of failure cases was also explored as an alternative to weighting during preliminary experimentation. Model performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC (Section 2.7). Because a missed failure carries direct patient-safety and operational consequences, recall and ROC-AUC - which are less distorted by extreme class imbalance than raw accuracy - were treated as the primary criteria for model selection rather than accuracy alone.

Random Forest was carried forward as the failure-probability estimator supplying the reinforcement-learning environment. This choice reflected its favorable balance of discriminative performance (Section 3.2) and interpretability: unlike Gradient Boosting, Random Forest exposes a transparent, aggregate measure of feature importance (Section 3.3) that can be communicated to HTM staff, which is an important practical consideration for a decision-support tool intended to be trusted and acted upon by non-specialist hospital personnel.

2.4 Reinforcement-learning Replacement Framework

The failure probabilities produced by the Random Forest model were embedded in a Markov decision process (MDP) formulation of the device-replacement problem, in which an agent observes the current condition of a device, selects an action, and receives a reward reflecting the economic and operational consequences of that action, with the objective of learning a policy that maximizes cumulative reward - equivalently, minimizes cumulative lifecycle cost - over a device's operating horizon (Fig. 2). The MDP was specified as follows:

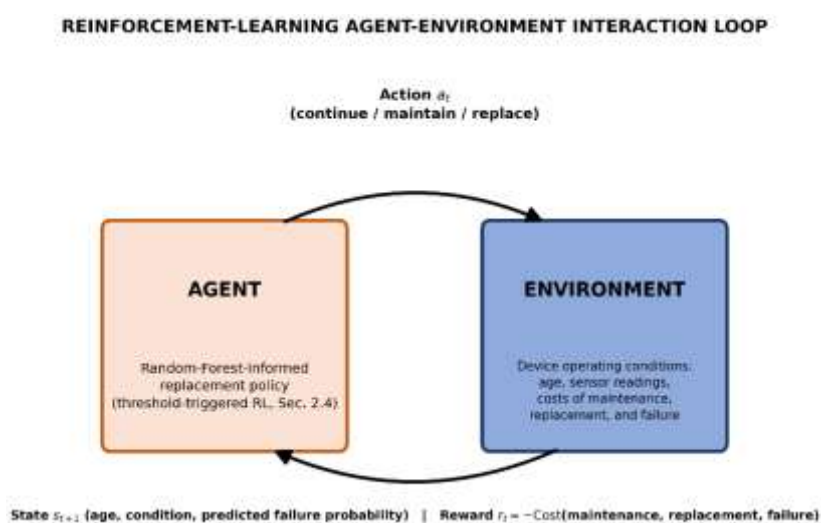


Figure 2. Reinforcement-learning agent-environment interaction loop.

State: the device's current age, its recent operating condition as summarized by the engineered sensor features, and the failure probability output by the Random Forest model at that time step, so that the agent's decisions are informed jointly by operating history and model-estimated risk.

Actions: at each decision point the agent selects one of three actions - continue routine operation, schedule preventive maintenance, or replace the device - each carrying a distinct cost and a distinct effect on the device's subsequent failure risk.

Reward: defined at each step as the negative of the total cost incurred, combining a fixed maintenance cost, a fixed replacement cost, and a substantially larger failure cost incurred only if the device was allowed to fail (set at US\$100, US\$500, and US\$1,000 per event, respectively, reflecting the relative severity of routine upkeep, planned replacement, and unplanned breakdown), together with a small credit for continued uptime. This structure penalizes both replacing equipment prematurely - forfeiting remaining useful life for no operational benefit - and delaying replacement past the point of elevated failure risk.

While the system architecture (Section 2.5) is designed to accommodate value-based and policy-gradient deep-RL algorithms such as deep Q-networks (DQN) and proximal policy optimization (PPO) for learning the replacement policy directly from continuous state inputs, the present evaluation used a simplified, risk-threshold formulation of the RL policy to keep the analysis tractable and transparent given the limited number of observed failure events ($n = 106$) available for policy learning and validation. Under this formulation, the replacement action was triggered whenever the Random-Forest-predicted failure probability crossed a fixed threshold, and four such policies were compared: an "always continue" (fully reactive) baseline that takes no proactive action, and three risk-based thresholds - replace when predicted failure probability is ≥ 0.7 , ≥ 0.5 , or ≥ 0.3 - representing progressively more conservative (earlier) intervention points. Each policy was applied to the same held-out device trajectories, and cumulative reward (equivalently, total realized cost), together with the resulting counts of failures, replacements, and maintenance actions, was recorded to assess policy performance, cost-efficiency, and stability (Sections 3.5-3.6). This threshold-policy evaluation can be viewed as a discretized, interpretable approximation of the value function that a fully trained DQN or PPO agent would be expected to learn, and it provides a transparent baseline against which future deep-RL implementations can be benchmarked once larger volumes of hospital-specific operational data become available.

2.5 System Architecture

The proposed decision-support system was organized into five functional layers (Fig. 3, Fig. 4), moving from raw operational data through prediction and decision-making to outputs intended for human review. Each layer consumes the output of the layer before it, so that the accuracy of the prediction layer bounds the quality of the decisions made downstream, while the evaluation and reporting layers provide the feedback and transparency needed to audit and refine the system over time.

Data layer: Stores and manages the raw sensor telemetry, device metadata (identifiers, installation dates, and equipment class), and maintenance histories that underlie the analysis. This layer is responsible for consolidating heterogeneous inputs into a structured, versioned form and supplying clean, time-ordered records to the prediction layer for pre-processing and feature extraction, consistent with the pipeline described in Section 2.2.

Prediction layer: Applies the trained classifier (Section 2.3) to the engineered feature set to estimate, for each device at each time step, the probability of failure. This layer encapsulates the model-selection process described above - Logistic Regression, Random Forest, and Gradient Boosting were each implemented here during development - and outputs a single failure-likelihood score to the decision layer, rather than exposing the RL agent directly to raw, high-dimensional sensor data.

Decision layer (reinforcement-learning agent): Receives the device's state - age, operating condition, and predicted failure probability - from the prediction layer and selects one of the three actions defined in Section 2.4 (continue, maintain, or replace). As implemented and evaluated in this study, the agent applies the risk-threshold policies described in Section 2.4; the layer is nonetheless designed so that a value-based or policy-gradient algorithm such as DQN or PPO could be substituted in to learn the mapping from state to action directly, once sufficient failure data are available to support such training.

Optimization and evaluation layer: Quantify the trade-off between replacing a device too early - forfeiting remaining useful life - and delaying replacement until failure risk is high, by weighing the cost of early intervention against the expected cost of a delayed failure for each candidate policy. This layer runs the cost-effectiveness analysis and policy simulations reported in Sections 3.5-3.6, and computes the classification and reinforcement-learning evaluation metrics (accuracy, precision, recall, F1-

score, ROC-AUC, cumulative reward, and total lifecycle cost) used to compare policies and validate the system's decision-making before any recommendation reaches a human reviewer.

Visualization and reporting layer: Translate the outputs of the layers above into a form usable by hospital administrators and HTM staff, in the form of dashboards intended to display real-time failure-risk predictions, upcoming maintenance and replacement schedules, and summary performance indicators. This layer is the system's principal human interface: it does not alter the underlying predictions or policy but is intended to make the model's recommendations auditable and actionable, supporting the interpretability and clinical-workflow integration goals discussed in Section 4.

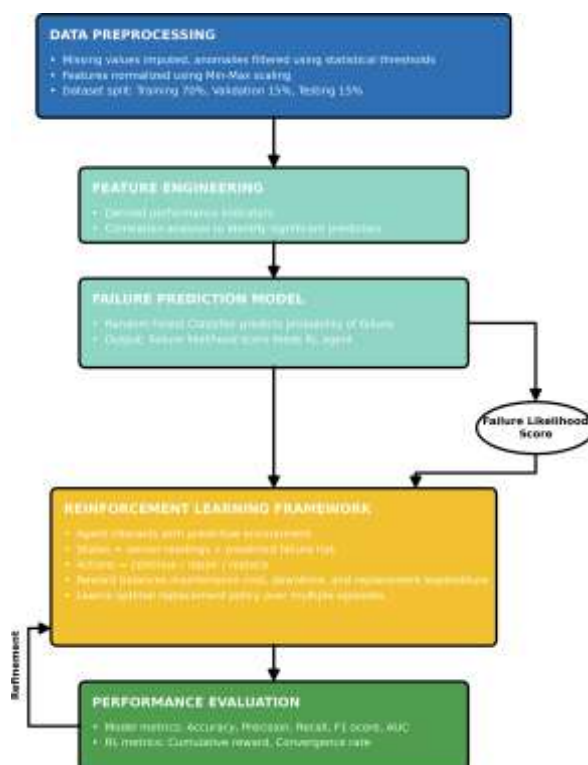


Figure 3. Flowchart of the proposed methodology

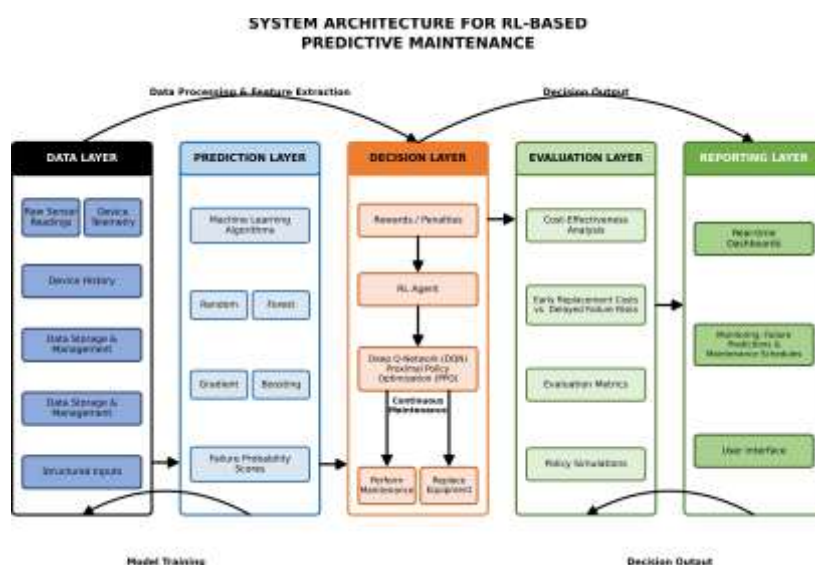


Figure 4. Proposed five-layer model architecture.

3. RESULTS AND DISCUSSION

3.1 Predictive model performance

All three classifiers achieved high raw accuracy (0.969-0.995, Table 2), but this figure is close to uninformative given the extreme rarity of failures in the dataset: a trivial classifier that always predicted "no failure" would already score approximately 99.91% accuracy, so the near-identical, near-ceiling accuracy values across models mainly reflect class imbalance rather than genuine discriminative skill. Precision was correspondingly low for all three models (0.012-0.040), an expected consequence of the roughly 1,100-to-1 ratio of non-failure to failure records rather than evidence that the models are operationally unreliable, since even a well-calibrated classifier will generate many false positives per true positive when positives are this rare. Recall and ROC-AUC therefore provide a more meaningful basis for comparison. Logistic Regression achieved the highest recall at the default 0.5 threshold (0.45) but the lowest ROC-AUC (0.684, Fig. 5b), indicating that, while it identified a relatively large share of true failures at that specific threshold, its ability to rank devices by risk across the full range of possible thresholds was comparatively weak. Gradient Boosting achieved the highest accuracy, precision, and F1-score of the three (0.995, 0.040, and 0.066, respectively) but a lower ROC-AUC (0.826) than Random Forest (0.849, Table 2, Fig. 5), and a visibly lower recall (0.20).

Random Forest achieved the highest ROC-AUC (0.849) among the three classifiers, indicating the strongest overall ability to discriminate failure from non-failure conditions despite the extreme class imbalance. This distinction matters specifically because the classifier's output is not used as a fixed-threshold decision rule in isolation, but as a continuous risk score feeding the threshold-based reinforcement-learning policies compared in Section 2.4 (thresholds of 0.7, 0.5, and 0.3); a model with strong ranking performance across the full probability range, as reflected in ROC-AUC, is therefore more valuable to this framework than one that is narrowly optimized for a single default cutoff. This is visible in Fig. 5(b): the Random Forest and Gradient Boosting curves both rise well above the diagonal chance line, particularly in the low false-positive-rate region most relevant to a screening application, whereas the Logistic Regression curve stays markedly closer to the diagonal despite its higher recall at the 0.5 threshold. On balance, Random Forest offered the best combination of robustness, ranking performance, and interpretability, and was retained as the failure-probability estimator feeding the reinforcement-learning layer.

Table 2. Model performance comparison (validation set)

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.969	0.012	0.45	0.023	0.684
Random Forest	0.984	0.013	0.25	0.024	0.849
Gradient Boosting	0.995	0.040	0.20	0.066	0.826

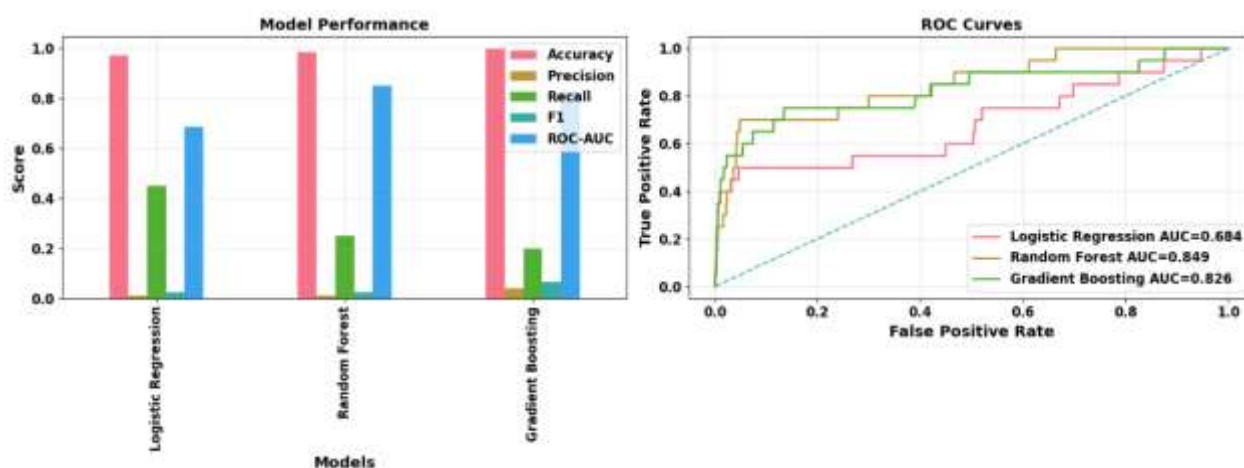


Figure 5. (a) Classification performance of the three predictive models; (b) corresponding ROC curves.

3.2 Feature Importance

Device-level historical failure count was overwhelmingly the strongest predictor of future failure (importance = 0.335), more than five times larger than any other feature in the top ten, consistent with the intuitive principle that a device's failure history is informative about its future reliability. The remaining nine features cluster into two related families. The first is built around load-cycle count: its current value (0.028), lagged values at one, two, and three preceding periods (0.027, 0.045, and 0.043, respectively), its rolling standard deviation (0.029), and its rolling maximum (0.038) together account for the majority of the remaining predictive signal, indicating that both the recent trajectory and the peak intensity of mechanical loading — not just its instantaneous level — carry information about impending failure. The second family reflects read-operation and retry behaviour: minimum read operations (0.063), the second-most-important feature overall, together with the rolling standard deviations of read operations (0.030) and spin-retry count (0.030), suggests that unusually low sustained throughput and increased variability in retry behaviour are early operational signatures of degrading hardware. The prominence of lagged and rolling-statistic versions of load-cycle count over its raw current value, in particular, underscores the value of the temporal feature engineering described in Section 2.2 (Fig. 6): a snapshot of a device's current state was consistently less informative than features that captured how that state evolved over the preceding cycles. Device identity alone showed little discriminatory value, indicating that failures were distributed broadly across the device population rather than concentrated in a small subset of units, which reinforces the case for sensor- and history-based, rather than device-based, risk modelling.

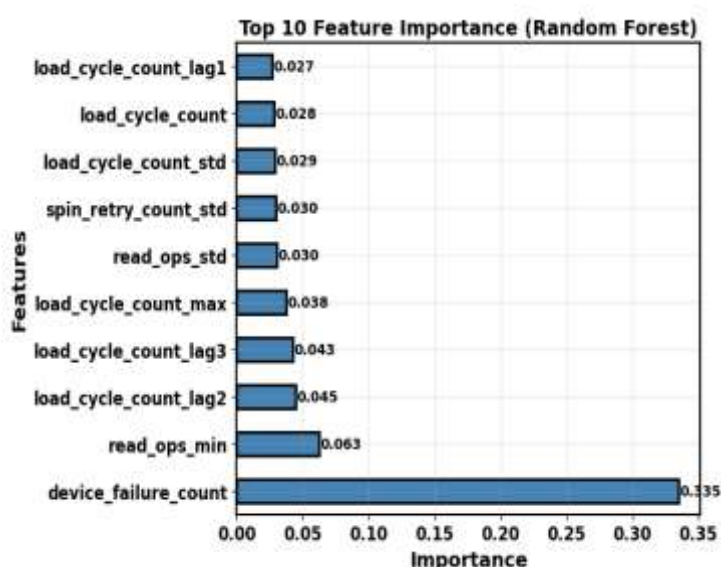


Figure 6. Top ten features contributing to failure prediction, Random Forest model.

3.5 Reinforcement-learning policy comparison

The reactive "always continue" baseline, in which equipment ran to failure before corrective action, incurred a total cost of US\$1,400 (Table 4), reflecting the cumulative failure cost (US\$1,000 per event, Section 2.4) accrued across the evaluated device trajectories with no proactive replacement at all. Both risk-based thresholds (≥ 0.7 and ≥ 0.5) substantially reduced cost, to US\$870 and US\$860 respectively - reductions of 37.9% and 38.6% relative to the reactive baseline - with the 0.5 threshold performing marginally better. This modest difference between the two thresholds indicates that, once the policy is proactive at all, its exact cutoff matters comparatively little within a moderate range: both intervened early enough to convert most incipient failures into far cheaper planned replacements (US\$500 per event) or maintenance actions (US\$100 per event), without triggering an excessive number of interventions on devices that were not, in fact, at meaningful risk.

In sharp contrast, the most aggressive threshold (≥ 0.3) increased total cost more than eleven-fold, to US\$16,680 (Table 4, Fig. 7). Because a threshold of 0.3 flags a device for replacement as soon as its predicted failure probability crosses a comparatively modest bar, it triggered replacement far more frequently than the higher thresholds - including on many devices whose elevated risk score did not, in fact, presage near-term failure - so that the cumulative cost of largely unnecessary US\$500 replacement actions came to dwarf the far smaller number of US\$1,000 failure events it was intended to prevent. This reveals a sharp

nonlinearity, rather than a smooth trade-off, in the relationship between intervention threshold and lifecycle cost: cost fell only modestly as the threshold relaxed from 0.7 to 0.5, then rose by more than an order of magnitude as it relaxed further to 0.3, implying that the cost-minimizing region of threshold space is comparatively narrow. These results indicate that policy performance is highly sensitive to the intervention threshold, and that a moderate, risk-based trigger point - rather than either an unconditionally reactive stance or an overly trigger-happy one - minimizes lifecycle cost.

Table 4. Cost comparison of replacement policies.

Policy	Total cost (US\$)
Always continue (reactive)	1,400
Threshold ≥ 0.7	870
Threshold ≥ 0.5	860
Threshold ≥ 0.3	16,680

3.6 Action distribution under the optimal policy

Under the cost-minimizing threshold-0.5 policy, 70.6% of interventions were replacements, 17.6% were preventive-maintenance actions, and 11.8% ended in unplanned failure (Fig. 7). Relative to a fully reactive baseline, in which deferred interventions culminate in failure, this corresponds to an approximate 88% reduction in unplanned-failure events, illustrating how timing intervention to predicted risk, rather than to a fixed calendar, can materially reduce both operational disruption and expenditure.

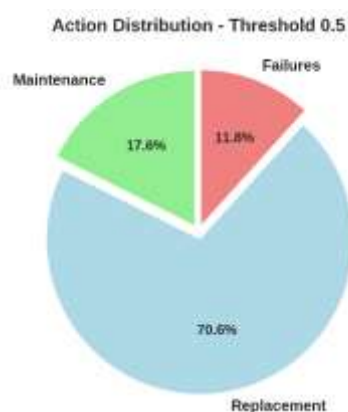


Figure 8. Distribution of actions under the optimal (threshold ≥ 0.5) policy.

3.7 Discussion

Together, these findings indicate that predictive failure modelling and reinforcement-learning-based policy optimization are complementary rather than substitutable capabilities: an accurate classifier identifies which devices are at risk, while the RL layer determines when intervention is economically optimal. The strong influence of historical failure count as a predictor, together with the essentially unblemished reliability of the low-risk category, suggests that a relatively small set of device-history features can support much of the risk stratification needed for lifecycle decision-making, which is encouraging for eventual deployment in resource-constrained HTM units. The clear cost advantage of moderate thresholds over both extremes also carries a practical caution: an overly aggressive, low-bar intervention policy - replacing equipment at the first sign of even modest predicted risk - can be far costlier than a well-calibrated moderate threshold, just as an unconditionally reactive stance is, reinforcing the case for a deliberately calibrated, evidence-based risk threshold over either extreme of the policy space.

Several limitations qualify these findings. First, the very high classification performance obtained for some configurations raises the possibility of information leakage through aggregated historical features, and should be re-examined with stricter, strictly prospective temporal validation before the framework is relied upon operationally, drawing on established sequential decision-making theory [26,27]. Second, the cost parameters used to define the reward function were fixed and simplified; real hospital environments involve more variable and multidimensional costs, including clinical risk, service disruption, and regulatory penalties, and future work should incorporate dynamic, context-sensitive cost functions consistent with SFDA and Ministry of Health requirements [28,29]. Third, the dataset used here, while structurally comparable to hospital equipment operation, is not itself a clinical dataset; validating the framework on real, de-identified HTM records - ideally spanning multiple device classes and accounting for Hajj- and Umrah-period demand surges - is a necessary next step toward deployment in Saudi hospitals consistent with Vision 2030's digital-health objectives [30]. Finally, because Random Forest offers only aggregate feature importance rather than case-level explanation, integrating explainable-AI techniques such as SHAP or LIME would likely improve clinical and HTM staff confidence in, and willingness to act on, individual replacement recommendations.

4. CONCLUSION

This study developed and evaluated a two-stage decision-support framework that couples supervised failure prediction with a reward-driven reinforcement-learning policy for healthcare-equipment replacement. Random Forest provided the strongest failure-discrimination performance among the classifiers tested (ROC-AUC = 0.849), with device-level failure history as its dominant predictor, and a moderate, risk-based intervention threshold (probability ≥ 0.5) minimized total lifecycle cost, reducing it by 38.6% relative to reactive, run-to-failure practice while cutting unplanned failures by roughly 88%. These results support the central premise of this work: healthcare equipment-replacement decisions can be shifted from static, calendar- or threshold-driven practice toward adaptive, evidence-based policies that jointly account for cost, reliability, and risk. For Saudi Arabian healthcare institutions operating under Vision 2030 and SFDA regulatory oversight, such a framework offers a practical pathway toward more resilient, cost-effective, and safety-conscious equipment lifecycle management, provided that future work validates it on real hospital data, incorporates dynamic and regulation-aware cost structures, and builds in the explainability needed to earn the confidence of HTM decision-makers.

AUTHOR CONTRIBUTIONS

Waleed Mohammed Alzuman: Conceptualization, Methodology, Data curation, Formal analysis, Writing—Review & Editing.

Mohamed Yacin Sikkandar: Validation, Writing—Review & Editing, Supervision.

Sabarunisha Begum S: Literature search, Data curation, Formal analysis, Investigation, Writing—Review & Editing.

Jackson Daniel: Validation, Writing—Review & Editing, Supervision.

Ahmed Farag Salem Babetat: Data curation, Formal analysis, Investigation, Writing—Review & Editing.

Faleh Ibrahim Alrashdi: Data curation, Formal analysis

Majed Atallah Almutairi: Literature search, Data curation, Formal analysis

Fahad Sattan Alrashdi: Data curation, Formal analysis

Khalid Munawir Alharbi: Literature search, Data curation, Formal analysis

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

Ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

The dataset used in this study is publicly available through the Kaggle repository. The Predictive Maintenance Dataset can be accessed at: <https://www.kaggle.com/datasets/hiimanshuagarwal/predictive-maintenance-dataset/data>.

GENERATIVE AI USE STATEMENT

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely for language editing, grammar correction, and improvement of clarity and readability. The authors reviewed and edited all AI-assisted content and take full responsibility for the accuracy, originality, and integrity of the final manuscript.

REFERENCES

- [1] Jayaraman, P., Desman, J., Sabounchi, M., Nadkarni, G.N., Sakhuja, A., 2024. A primer on reinforcement learning in medicine for clinicians. *npj Digital Medicine* 7, 337.
- [2] Zhang, H., Liu, J., Chen, Y., 2024. Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making. *ResearchGate Preprint*.
- [3] Frommeyer, T.C., Gilbert, M.M., Fursmidt, R.M., Park, Y., Khouzam, J.P., Brittain, G.V., Frommeyer, D.P., Bett, E.S., Bihl, T.J., 2025. Reinforcement learning and its clinical applications within healthcare: a systematic review of precision medicine and dynamic treatment regimes. *Healthcare* 13(14), 1752.
- [4] Al-Hamadani, M.N.A., Fadhel, M.A., Harangi, B., et al., 2024. Reinforcement learning algorithms and applications in healthcare and robotics: a comprehensive and systematic review. *Sensors* 24(8), 2461.
- [5] Pliego Marugán, A., Pinar-Pérez, J.M., García Márquez, F.P., 2024. A reinforcement learning agent for maintenance of deteriorating systems with increasingly imperfect repairs. *Reliability Engineering & System Safety* 252, 110466.
- [6] Ali, H., 2022. Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making. *International Journal of Computer Applications Technology and Research* 11(3), 88–104.
- [7] Enhaili, N., 2024. *Improving Healthcare Policies Using Reinforcement Learning on Patterns of Service Utilization*. Master's thesis, Simon Fraser University.
- [8] Adsule, A., Kulkarni, M., Tewari, A., 2020. Reinforcement learning for optimal policy learning in condition-based maintenance. *IET Collaborative Intelligent Manufacturing* 2(4), 182–188.
- [9] Mahmoodzadeh, Z., Wu, K.-Y., Lopez Droguett, E., Mosleh, A., 2020. Condition-based maintenance with reinforcement learning for dry gas pipeline subject to internal corrosion. *Sensors* 20(19), 5708.
- [10] Knowles, M., Baglee, D., Wermter, S., 2010. Reinforcement learning for scheduling of maintenance. In: *Research and Development in Intelligent Systems XXVII*, Springer, London, pp. 409–422.
- [11] Geurtsen, M., Adan, I.J.B.F., Atan, Z., 2023. Deep reinforcement learning for optimal planning of assembly line maintenance. *Journal of Manufacturing Systems* 69, 170–188.
- [12] Chen, Y., Liu, C., 2025. Sequential multi-objective multi-agent reinforcement learning approach for system predictive maintenance of turbofan engine. *Advanced Engineering Informatics* 67, 103553.
- [13] Abensur, E.O., Santos, B.P., Bandeira, A.A., 2023. Optimization models as applied to equipment replacement problems: review and trends. *Gestão & Produção* 30, e4022.
- [14] K B, Venkatesan L, Benjamin LS, K V, Satchi NS. Reinforcement Learning in Personalized Medicine: A Comprehensive Review of Treatment Optimization Strategies. *Cureus*. 2025 Apr 21;17(4): e82756.
- [15] Wu Q, Han J, Yan Y, Kuo YH, Shen ZM. Reinforcement learning for healthcare operations management: methodological framework, recent developments, and future research directions. *Health Care Manag Sci*. 2025 Jun;28(2):298-333.
- [16] Zhao, Y., Yang, J., Wang, W., Yang, H., Niyato, D., 2024. TranDRL: a transformer-driven deep reinforcement learning enabled prescriptive maintenance framework. *IEEE Internet of Things Journal* 11(21), 35432–35444.
- [17] Ong, K.S.H., Niyato, D., Yuen, C., 2020. Predictive maintenance for edge-based sensor networks: a deep reinforcement learning approach. In: *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, IEEE, pp. 1–6.

- [18] Ogunfowora, O., Najjaran, H., 2023. Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization. *Journal of Manufacturing Systems* 70, 244–263.
- [19] Liao, H.-Y., Cade, W., Behdad, S., 2021. Markov chain optimization of repair and replacement decisions of medical equipment. *Resources, Conservation and Recycling* 171, 105609.
- [20] Hu, Y., Miao, X., Zhang, J., Liu, J., Pan, E., 2021. Reinforcement learning-driven maintenance strategy: a novel solution for long-term aircraft maintenance decision optimization. *Computers & Industrial Engineering* 153, 107056.
- [21] Agarwal, H., 2024. Predictive maintenance dataset. *Kaggle*. Available at: <https://www.kaggle.com/datasets/himanshuagarwal/predictive-maintenance-dataset/data> (accessed on 16 January 2026).
- [22] Nair, K.S., Mughal, Y.H., Albejaidi, F., Alharbi, A.H., 2024. Healthcare financing in Saudi Arabia: a comprehensive review. *Healthcare* 12(24), 2544.
- [23] Yunfei Zhao, Carol Smidts, Reinforcement learning for adaptive maintenance policy optimization under imperfect knowledge of the system degradation model and partial observability of system states, *Reliability Engineering & System Safety*, 224, 108541, 2022.
- [24] M. Feng and Y. Li, Predictive Maintenance Decision Making Based on Reinforcement Learning in Multistage Production Systems, *IEEE Access*, vol. 10, 18910-18921, 2022.
- [25] Wang, Z., Tang, J., Wei, P. et al. Deep reinforcement learning-driven multi-objective optimization and its applications on lighting infrastructure operation and maintenance strategy. *Sci Rep*, 16, 8989, 2026.
- [26] Puterman, M.L., 2014. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Hoboken, NJ.
- [27] Sutton, R.S., Barto, A.G., 2018. *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, Cambridge, MA.
- [28] Saudi Food and Drug Authority, 2023. *Medical Devices Interim Regulation and Guidance*. Saudi Food and Drug Authority, Issued by SFDA Board Decision No. (3-29-1443) dated 19/02/1443 AH (accessed on 25 March 2026).
- [29] Ministry of Health, Saudi Arabia, 2022. *Health Sector Transformation Strategy*. Ministry of Health, Saudi Arabia (accessed on 26 March 2026).
- [30] Saudi Vision 2030, *Annual Report*, Kingdom of Saudi Arabia (accessed on 26 March 2026).