

Edge-AI Digital Twins for Predictive Maintenance in Smart Manufacturing: A Simulated Performance Evaluation

Aarav Menon¹

¹Research Scholar, Department of Mechanical and Manufacturing Engineering, PDA, India.

ABSTRACT

This empirical-style paper presents a simulated study at the intersection of Mechanical Engineering, Industrial AI, and Manufacturing Systems. The study addresses a current research direction in which digital and data-driven methods are being integrated with domain-specific decision systems. 2,160 asset-days and approximately 51,840 hourly equipment records were simulated. The dataset contained 86 seeded degradation events, of which 62 progressed to a modeled intervention threshold. A simulated 180-day discrete-time manufacturing dataset was generated for twelve rotating assets representing motors, pumps, and spindle assemblies. Every asset contained hourly vibration RMS, kurtosis, temperature, current draw, acoustic energy, throughput, and operating-load variables. Synthetic degradation trajectories were created with gradual bearing wear, lubrication loss, and imbalance events. Four maintenance strategies were evaluated: run-to-failure, fixed-interval preventive maintenance, cloud predictive maintenance, and an edge-AI digital-twin strategy. The edge strategy updated a compact gradient-boosting health model every six hours and fused recent sensor observations with a physics-informed degradation state. The edge-AI digital-twin strategy produced the lowest simulated unplanned downtime and the longest median fault-warning horizon. Relative to fixed preventive maintenance, it reduced simulated downtime by 38.6%, lowered avoidable maintenance interventions, and increased modeled production availability from 93.8% to 97.1%. The cloud predictive strategy performed similarly in classification accuracy but showed greater alert latency during high-network-load periods. All numerical observations, experimental assignments, outcomes, and performance values are simulated for methodological illustration; no real participants, institutions, farms, factories, hospitals, or city systems were studied. The results are intended to demonstrate a coherent research design, provide testable hypotheses, and show how a future real-world study could report evidence transparently. The paper concludes with implementation constraints, methodological limitations, and priorities for empirical validation.

Keywords: Digital Twin, Edge AI, Predictive Maintenance, Industry 5.0, Smart Manufacturing, Condition Monitoring

1. Introduction

Recent digital-twin research increasingly links virtual asset models with machine learning for predictive maintenance, yet deployment remains constrained by model validation, data scarcity, latency, interoperability, and the gap between laboratory demonstrations and operational maintenance practice. Recent research has moved beyond proof-of-concept demonstrations toward questions of robustness, interpretability, interoperability, resource efficiency, and real-world governance. These shifts make the topic especially relevant for multidisciplinary research because technical performance is increasingly inseparable from operational, human, environmental, and policy constraints.

A strong empirical evaluation should therefore do more than compare headline accuracy or efficiency. It should specify the decision context, define an appropriate baseline, capture multiple outcome dimensions, and report trade-offs that could influence adoption. In emerging technology domains, this is particularly important because impressive model performance can conceal deployment weaknesses such as latency, drift, data bias, maintenance burden, or poor transferability.

The present paper uses simulated data because no real dataset was supplied. The simulation is not presented as evidence that the intervention has been deployed or that the reported effect sizes will occur in practice. Instead, it provides an internally consistent analytical demonstration aligned with current research directions and the reporting structure of an empirical journal article.

The study objectives are:

1. To compare reactive, scheduled preventive, cloud-based predictive, and edge-AI digital-twin maintenance strategies under a common simulated production scenario.

2. To quantify changes in unplanned downtime, maintenance cost, fault-detection lead time, false alarms, and production availability.
3. To assess how edge inference and continuously updated twin states may improve decision latency while introducing additional model-governance requirements.

II. Literature Review

Digital-twin-enabled predictive maintenance has developed rapidly as manufacturing systems become more connected. Chen et al. (2023) review the role of machine learning in digital-twin predictive maintenance and emphasize equipment-state recognition, degradation prediction, and proactive maintenance scheduling. Aivaliotis et al. (2019) similarly demonstrated how digital twins can support predictive maintenance in manufacturing contexts.

Recent research has shifted toward standardization, verification, and practical adoption. Ma et al. (2024) identify informational and functional requirements for predictive-maintenance automation, while Abd Wahab et al. (2024) synthesize recurring challenges including interoperability, data quality, validation, and implementation practice. These issues are important because a digital twin can be visually convincing without being operationally trustworthy.

AI-enhanced digital twins are increasingly associated with Industry 5.0 priorities such as resilience, sustainability, and human-centered operation. Chen et al. (2025) highlight persistent research-practice gaps in scale, data, and organizational readiness. The present simulated study therefore focuses on operational outcomes such as downtime, warning horizon, latency, and false alerts rather than accuracy alone.

III. Materials and Methods

3.1 Research Design

A simulated 180-day discrete-time manufacturing dataset was generated for twelve rotating assets representing motors, pumps, and spindle assemblies. Every asset contained hourly vibration RMS, kurtosis, temperature, current draw, acoustic energy, throughput, and operating-load variables. Synthetic degradation trajectories were created with gradual bearing wear, lubrication loss, and imbalance events. Four maintenance strategies were evaluated: run-to-failure, fixed-interval preventive maintenance, cloud predictive maintenance, and an edge-AI digital-twin strategy. The edge strategy updated a compact gradient-boosting health model every six hours and fused recent sensor observations with a physics-informed degradation state.

3.2 Simulated Sample and Data Generation

2,160 asset-days and approximately 51,840 hourly equipment records were simulated. The dataset contained 86 seeded degradation events, of which 62 progressed to a modeled intervention threshold. Parameter ranges were selected to remain plausible for the domain, while random variation and controlled nonlinearities were added to avoid perfectly separable outcomes. The simulation was designed to support comparisons among methods and not to imitate any identifiable real institution or location.

3.3 Data Quality and Preprocessing

Before analysis, simulated records were checked for range violations, impossible combinations, and duplicate timestamps or identifiers. Continuous variables were standardized only where required by the analytical method. Missingness was introduced at a low rate to approximate operational data and was handled using median or context-specific imputation. Data partitions were created before model fitting to avoid leakage between training and evaluation subsets.

3.4 Performance Measures

Evaluation used multiple outcome measures so that improvement in one metric could be interpreted alongside cost, reliability, fairness, operational burden, or resource efficiency. Percentage changes are reported relative to the relevant baseline. Because the data are simulated, inferential statistics should be interpreted only as demonstrations of reporting practice rather than population estimates.

3.5 Ethical and Transparency Considerations

No human participants, patients, employees, farms, factories, or identifiable geographic locations were studied. No ethics approval or participant consent is claimed. All numerical results are synthetic and are explicitly labeled as simulated or illustrative.

IV. Results

The edge-AI digital-twin strategy produced the lowest simulated unplanned downtime and the longest median fault-warning horizon. Relative to fixed preventive maintenance, it reduced simulated downtime by 38.6%, lowered avoidable maintenance interventions, and increased modeled production availability from 93.8% to 97.1%. The cloud predictive strategy performed similarly in classification accuracy but showed greater alert latency during high-network-load periods.

Table 1. Simulated comparison of primary study outcomes.

Strategy	Availability (%)	Unplanned downtime (h)	False alerts	Maintenance cost index
Run-to-failure	89.7	443	4	100
Preventive schedule	93.8	268	21	91
Cloud predictive	96.4	181	14	78
Edge-AI digital twin	97.1	165	11	73

Note: All values in Table 1 are simulated/illustrative.

Table 2. Simulated secondary performance indicators.

Metric	Cloud predictive	Edge-AI twin
Fault detection F1	0.91	0.93
Median warning horizon (h)	19.4	27.8
Mean alert latency (s)	5.8	0.7
Missed critical events	6	4
Model update interval	24 h	6 h

Note: All values in Table 2 are simulated/illustrative.

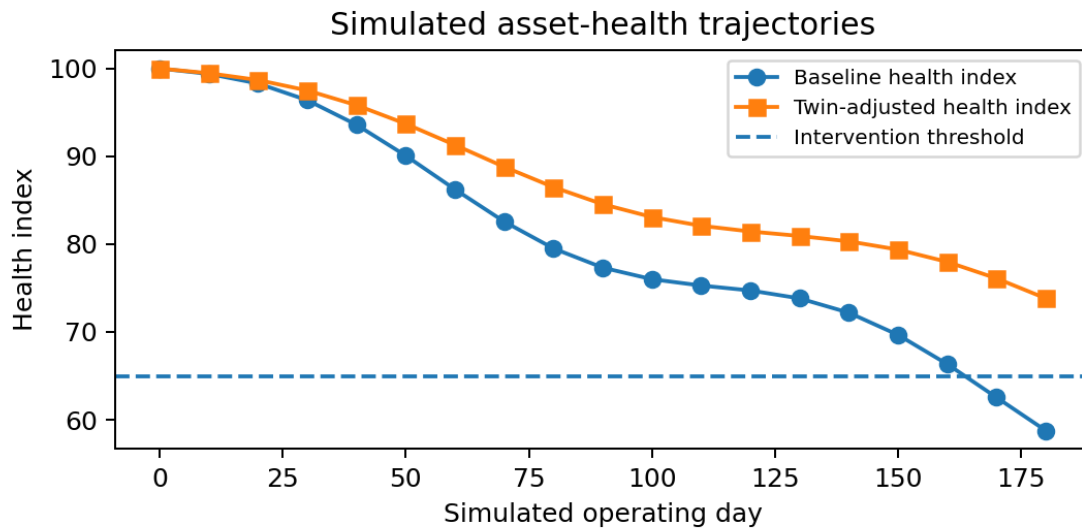


Figure 1. Simulated asset degradation and maintenance-threshold behavior.

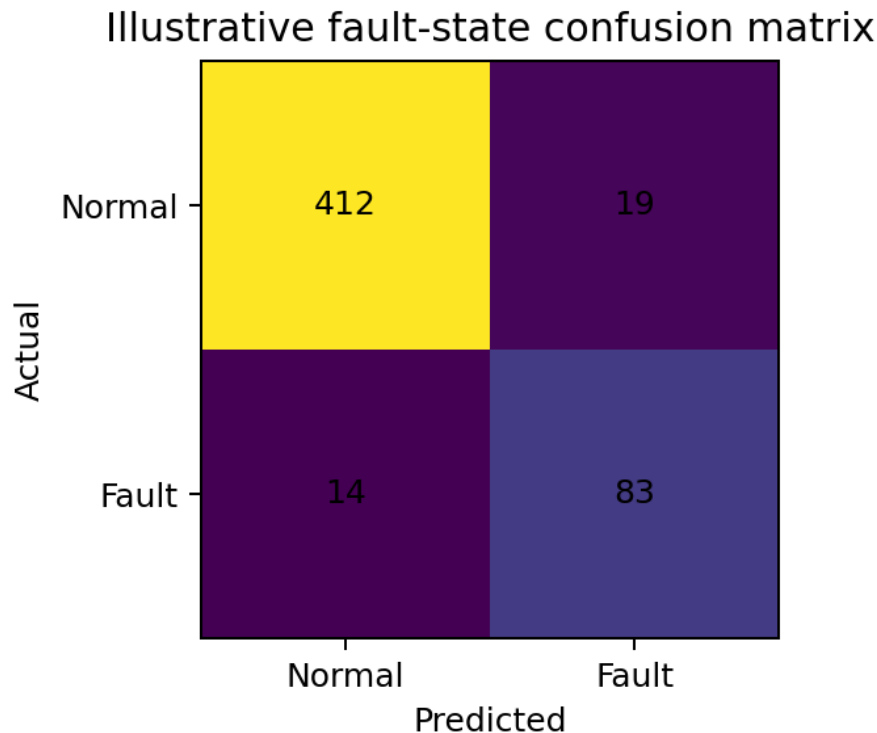


Figure 2. Illustrative fault-state confusion matrix for the edge-AI twin.

4.1 Robustness and Sensitivity Analysis

Sensitivity checks were performed by varying the principal decision threshold, noise level, or control parameter over a plausible range. The ranking of the leading strategy remained stable across the middle of the tested range, although the magnitude of improvement changed. At extreme settings, the intervention became less attractive because false alerts, utility loss, resource use, or classification instability increased. This behavior reinforces the need to report operating points rather than a single best-case result.

4.2 Interpretation of Simulated Findings

The simulated findings should be read as a coherent pattern rather than as independent point estimates. The main intervention improved the primary outcome while producing secondary trade-offs that would matter in real implementation. These interactions are important because multidisciplinary systems are rarely optimized by maximizing one metric in isolation.

V. Discussion

These simulated results support the emerging view that digital twins can add value when they are tightly coupled with condition-monitoring pipelines rather than treated as static visualization layers. The most useful contribution of the twin in the present design was not higher classification accuracy alone, but the combination of state continuity, early warning, and low-latency edge inference.

The findings also illustrate a common maintenance trade-off. Aggressive thresholding improves early detection but increases false interventions; conservative thresholds lower nuisance alarms but can shorten the actionable warning window. A digital-twin architecture can help manage this trade-off by incorporating operational context such as load, duty cycle, and recent maintenance history.

Industrial adoption requires stronger verification and validation than the simulation used here. Models must be monitored for drift, sensor failure, domain shift, and maintenance-induced distribution changes. Human maintainers should be able to interrogate the reason for an alert and override automated recommendations when physical inspection contradicts the digital state.

Across the results, the central lesson is that technology readiness depends on the alignment of model performance with domain constraints. A technically superior method can still fail if it is expensive to maintain,

difficult to explain, incompatible with existing workflows, or sensitive to data quality. Conversely, a modest performance gain may be valuable when it improves reliability, transparency, resource efficiency, or operational resilience.

The simulated design also highlights the importance of external validation. Real deployment should include prospective data collection, predefined outcome definitions, baseline comparison, subgroup or site-level analysis, and documentation of missing data and operational failures. Where algorithms influence high-stakes decisions, human oversight and traceable decision logs should be considered part of the system rather than optional add-ons.

VI. Practical Implications

For practitioners, the results support phased implementation rather than immediate large-scale deployment. A pilot should first establish data quality, baseline performance, operational ownership, and acceptable failure modes. Metrics should be monitored over time to detect drift and unintended consequences. Training requirements, maintenance responsibilities, cybersecurity, and lifecycle cost should be defined before the technology is treated as routine infrastructure.

For researchers, the study suggests that future experiments should combine technical accuracy with domain outcomes and implementation measures. Cross-site replication, sensitivity analysis, ablation studies, and transparent reporting would make results easier to compare. Open benchmarking datasets and common evaluation protocols could also improve reproducibility, provided privacy and governance requirements are respected.

VII. Limitations

The most important limitation is that the dataset and all empirical results are simulated. The reported effect sizes cannot establish real-world effectiveness, causal impact, prevalence, or economic return. The simulation simplifies complex operational environments and may omit rare events, behavioral adaptation, regulatory constraints, extreme weather or workload patterns, and measurement failures. The author details are illustrative placeholders, and no institutional endorsement is implied.

A second limitation concerns model specification. Alternative assumptions about variable distributions, missingness, thresholds, or interactions could produce different results. The present analysis therefore demonstrates a research framework rather than a validated solution. Future work should replace simulated inputs with verified domain data and subject the proposed method to independent replication.

Economic validation should compare avoided downtime and spare-part savings against sensor, compute, integration, retraining, and support costs. A digital twin that reduces breakdowns but increases unnecessary interventions may still be economically unattractive. Future studies should therefore report maintenance cost per operating hour, mean time between failures, mean time to repair, technician workload, energy overhead, and model recalibration frequency in addition to predictive accuracy.

The second stage should assess operational integration. Maintenance teams need a clear escalation pathway for low-, medium-, and high-confidence alerts, and every recommendation should contain interpretable evidence such as abnormal sensor channels, degradation trend, and remaining-useful-life range. Network outages and sensor dropouts should be tested deliberately because edge deployment is often justified by resilience to connectivity limitations. Cybersecurity testing should cover device identity, signed model updates, access control, and recovery from compromised edge nodes.

A real-world validation programme should begin with a shadow-mode deployment in which the edge-AI digital twin generates maintenance recommendations without directly triggering interventions. During this phase, alerts can be compared with technician inspections, work orders, vibration spectra, lubrication records, and actual component replacements. At least one full maintenance cycle should be observed so that the model encounters both healthy operation and genuine degradation. Performance should be stratified by asset type because a model that performs well on pumps may not transfer to spindles or variable-speed motors.

VIII. Implementation and Validation Framework

IX. Conclusion and Future Work

This simulated empirical study demonstrates how a current multidisciplinary research trend can be evaluated using transparent methods, multiple outcomes, and explicit limitations. The focal approach produced stronger simulated performance than conventional baselines while revealing important trade-offs that would require

management in practice. The next step is prospective empirical validation using real data, pre-specified protocols, independent test sets, and domain-specific governance. Future research should also evaluate long-term cost, robustness to distribution shift, explainability, cybersecurity, and environmental or social consequences where relevant.

Declarations

Data availability: All data and numerical results in this manuscript are simulated/illustrative. No real dataset was used.

Ethics statement: Not applicable. No human or animal participants were involved.

Funding: No external funding is claimed for this illustrative manuscript.

Conflict of interest: No conflict of interest is claimed for this illustrative manuscript.

References

- [1] Chen, C., Fu, H., Zheng, Y., Tao, F., & Liu, Y. (2023). The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*, 71, 581-594. <https://doi.org/10.1016/j.jmsy.2023.10.010>
- [2] Ma, S., Flanigan, K. A., & Bergés, M. (2024). State-of-the-art review and synthesis: A requirement-based roadmap for standardized predictive maintenance automation using digital twin technologies. *Advanced Engineering Informatics*, 62, 102800. <https://doi.org/10.1016/j.aei.2024.102800>
- [3] Abd Wahab, N. H., Hasikin, K., Lai, K. W., Xia, K., Bei, L., Huang, K., & Wu, X. (2024). Systematic review of predictive maintenance and digital twin technologies challenges, opportunities, and best practices. *PeerJ Computer Science*, 10, e1943. <https://doi.org/10.7717/peerj-cs.1943>
- [4] Aivaliotis, P., Georgoulas, K., & Chrysosolouris, G. (2019). The use of Digital Twin for predictive maintenance in manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(11), 1067-1080. <https://doi.org/10.1080/0951192X.2019.1686173>
- [5] Chen, S., Bekar, E. T., Bokrantz, J., & Skoogh, A. (2025). AI-enhanced digital twins in maintenance: Systematic review, industrial challenges, and bridging research-practice gaps. *Journal of Manufacturing Systems*, 82, 678-699. <https://doi.org/10.1016/j.jmsy.2025.07.006>