



AI-Based Predictive Maintenance Framework for Smart Manufacturing Systems

Behruz Jeongmin

Department of Civil and Environmental Engineering, Borneo School of Business and Technology, Malaysia
behruz.jeongmin@bsbt-my.org

Peer Review Information

Submission: 20 Jan 2023
Revision: 04 Feb 2023
Acceptance: 19 March 2023

Keywords

Predictive Maintenance;
Smart Manufacturing;
Artificial Intelligence;
Machine Learning; Industry
4.0

Abstract

Smart manufacturing systems increasingly depend on connected machines, industrial Internet of Things (IIoT) sensors, and data-driven decision support. Unplanned equipment failure can interrupt production, reduce asset availability, and increase maintenance cost. This paper proposes an artificial-intelligence-based predictive maintenance (AI-PdM) framework that integrates machine-condition data acquisition, preprocessing, feature engineering, failure prediction, remaining-useful-life-oriented risk assessment, explainable decision support, and maintenance feedback. The framework is designed for deployment in manufacturing environments where failure observations are rare and heterogeneous sensor variables must be interpreted together. The methodology uses the AI4I 2020 Predictive Maintenance Dataset as a reproducible benchmark because it contains 10,000 records of manufacturing-process conditions and machine-failure labels. A layered evaluation protocol is proposed in which classification performance is assessed by precision, recall, F1-score, area under the receiver-operating characteristic curve, and false-negative rate rather than by accuracy alone. The framework further introduces an explainability layer so that predicted failures can be linked to operational variables such as temperature, rotational speed, torque, and tool wear. The methodological analysis indicates that ensemble learners and deep/hybrid approaches are appropriate for nonlinear failure patterns, while explainability and feedback are essential for industrial acceptance. Published evidence on the same AI4I benchmark reports that tree-based ensemble methods can exceed 90% accuracy, supporting their role as strong baseline models. The resulting framework provides a structured path from raw shop-floor data to maintenance action while emphasizing imbalance-aware evaluation, transparency, and continuous model updating.

Introduction

Maintenance strategy has evolved from reactive repair toward condition-based and predictive approaches as manufacturing systems have become more automated, sensor-rich, and interconnected. Condition-based maintenance uses observed machine condition to support maintenance decisions, while prognostics extends this idea by estimating future

degradation and the time available before a functional failure occurs. Jardine et al. (2006) established data acquisition, data processing, and maintenance decision-making as central elements of condition-based maintenance, and Lee et al. (2006) described the transition from fail-and-fix practices to networked predict-and-prevent maintenance. These foundations are

directly relevant to smart manufacturing because cyber-physical production equipment now generates continuous operational data that can be analyzed in near real time.

The emergence of Industry 4.0 has expanded the role of artificial intelligence (AI) in maintenance. Machine learning (ML) can discover nonlinear relationships between process variables and failures without requiring a complete physics-based degradation model. Carvalho et al. (2019) showed that ML has become a major technique for predictive maintenance, but also emphasized that the choice of algorithm and data-processing strategy strongly affects performance. Similarly, Lei et al. (2018) organized machinery prognostics into data acquisition, health-indicator construction, health-stage division, and remaining useful life (RUL) prediction, illustrating that successful prognostics requires an end-to-end process rather than a stand-alone classifier.

Despite technical progress, practical predictive maintenance remains difficult. Industrial failure data are usually imbalanced because normal operation greatly outnumbers failure events; sensors may contain noise, drift, and operating-regime changes; and highly accurate black-box models may not provide reasons that maintenance engineers can trust. Matzka (2020) addressed this issue through explainable AI for predictive maintenance and introduced the AI4I 2020 dataset as a realistic synthetic benchmark. More recent research has continued to highlight interpretability, model transfer, data quality, and context dependence as open problems (Ferreira and Gonçalves, 2022; Hafsi et al., 2023; Surucu et al., 2023).

Accordingly, this paper develops a methodology-oriented AI-PdM framework for smart manufacturing systems. The contribution is not a new isolated prediction algorithm; instead, it is an integrated workflow that links sensor acquisition, data preparation, AI modeling, risk interpretation, explainability, and maintenance feedback. The framework is structured so that it can be tested first on public benchmark data and then adapted to real production assets. The study also defines an evaluation procedure that gives priority to recall and false-negative control because missing an impending failure can be more costly than issuing an additional inspection alert.

Literature Review

Early work on condition-based maintenance emphasized that diagnostics and prognostics depend on systematic sensing, signal processing,

and decision logic. Jardine et al. (2006) reviewed statistical, model-based, and AI approaches for diagnostics and prognostics, while Lee et al. (2006) connected prognostic intelligence with e-maintenance and remote monitoring. These studies established the architectural foundation for present-day smart maintenance systems.

As industrial sensing increased, research shifted toward data-driven prognostics. Lei et al. (2018) highlighted the complete path from data acquisition to RUL prediction and argued that health indicators should represent degradation in a form suitable for maintenance decisions. Carvalho et al. (2019) later reviewed ML methods used in predictive maintenance and found that applications commonly use classification, regression, and anomaly-detection models, but that comparisons are difficult because datasets, equipment, features, and evaluation procedures differ substantially between studies.

Explainability became a significant topic as ML models moved closer to operational deployment. Matzka (2020) proposed explainable AI for predictive maintenance using a synthetic manufacturing dataset and demonstrated the importance of presenting model reasoning to users. This is especially relevant when false alarms, missed failures, and maintenance actions have direct cost and safety implications. In digital-twin-oriented maintenance, Mi et al. (2021) proposed a cooperative awareness and interconnection framework in which digital representations support predictive maintenance decision making, showing how real-time data and decision models can be connected to asset management.

Recent work has concentrated on RUL prediction, machine learning challenges, and robust condition monitoring. Ferreira and Gonçalves (2022) mapped challenges across the RUL prediction process and noted issues related to operational datasets, accuracy, interpretability, and model selection. Surucu et al. (2023) reviewed ML-based condition monitoring and stressed that data processing and model choice must be considered together rather than evaluated independently. Hafsi et al. (2023) described a roadmap for predictive maintenance in smart industrial systems and identified context dependence, uncertainty, data analysis, knowledge representation, and transferability as continuing limitations.

Research published in 2023 also provides direct evidence for AI-based early failure detection in manufacturing. Hosseinzadeh et al. (2023) evaluated multiple ML, deep-learning, and hybrid models using the AI4I dataset and reported that

Deep Forest and Gradient Boosting produced accuracy above 90%. Chen et al. (2023) reviewed the role of ML within digital-twin predictive maintenance and concluded that the combination can improve equipment status recognition and proactive fault prediction. Taken together, the literature supports a layered architecture in which prediction is only one component of a broader maintenance intelligence system.

Methodology

1. Research Design

The proposed methodology follows a design-and-evaluate approach. First, a generic AI-PdM architecture is defined from the requirements identified in prior research. Second, the architecture is instantiated with the AI4I 2020 Predictive Maintenance Dataset as a reproducible benchmark. Third, suitable model families and imbalance-aware metrics are specified. Finally, a

decision layer converts model outputs into maintenance risk classes and feeds maintenance outcomes back into model monitoring. The method is intentionally modular so that a manufacturing organization can replace individual algorithms without changing the overall data-to-decision workflow.

The AI4I dataset contains 10,000 observations and records product type, air temperature, process temperature, rotational speed, torque, and tool wear along with machine-failure and failure-mode labels. The UCI repository describes the dataset as synthetic but representative of predictive-maintenance data encountered in industry. It contains 339 machine-failure observations, corresponding to a failure rate of 3.39%, which makes class imbalance a central methodological concern (Matzka, 2020; UCI Machine Learning Repository, 2020).

Table 1. Core variables used in the proposed predictive-maintenance benchmark

Variable	Role	Unit / Type	Maintenance relevance
Air temperature	Input	K	Operating and thermal environment
Process temperature	Input	K	Heat-transfer and process condition
Rotational speed	Input	rpm	Mechanical operating regime
Torque	Input	Nm	Mechanical load
Tool wear	Input	min	Accumulated degradation proxy
Machine failure	Target	Binary	Failure/no-failure decision

2. Proposed AI-PdM Framework

Figure 1 presents the six-layer workflow. Layer 1 acquires sensor and machine-context data through IIoT gateways, controllers, historians, or manufacturing execution systems. Layer 2 performs preprocessing, including missing-value checks, outlier treatment, scaling where required, categorical encoding, and feature engineering. Layer 3 trains AI models for failure

probability or anomaly risk. Layer 4 converts model outputs into failure risk and, where degradation sequences are available, RUL-oriented estimates. Layer 5 adds explainability and decision rules so that maintenance personnel can identify the variables driving a warning. Layer 6 issues a maintenance recommendation and records the actual outcome for later retraining and drift detection.

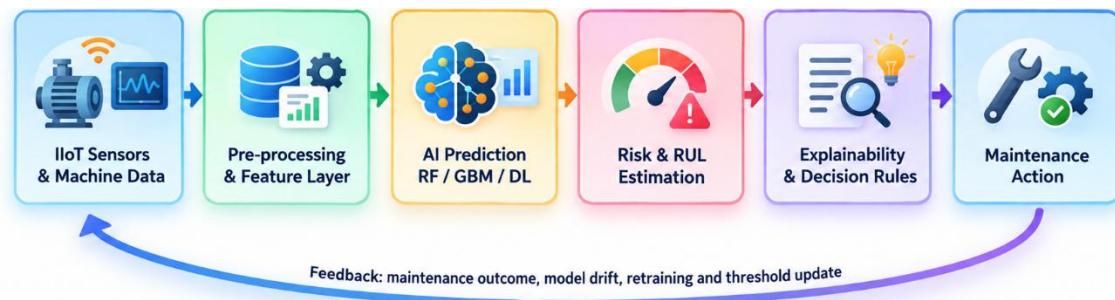


Figure 1. Proposed AI-based predictive maintenance framework for smart manufacturing systems.

3. Data Preparation and Feature Engineering

Identification fields such as unique product identifiers are excluded from direct model learning because they do not represent physical

degradation. Product type is one-hot encoded, while continuous operating variables are retained in physical units for tree-based methods and standardized for distance- or gradient-sensitive models. Derived features can include

temperature difference (process temperature minus air temperature), mechanical power calculated from torque and rotational speed, and interaction terms linking wear with load. These engineered variables reflect known operating relationships and can improve interpretability because the model receives physically meaningful indicators rather than only raw columns.

Because machine failures form a small minority class, a stratified split is required so that training, validation, and test partitions preserve failure prevalence. The recommended partition is 70% training, 15% validation, and 15% test data. Class weighting is preferred as the first imbalance treatment because it preserves the original observations; resampling methods may be tested in sensitivity analysis. Data leakage is prevented by fitting preprocessing transformations only on the training partition and then applying the

learned transformations to validation and test data.

4. Model Development

The framework evaluates three complementary model families. Logistic regression provides an interpretable linear baseline. Random Forest or Gradient Boosting provides nonlinear tree-based learning and feature importance while remaining computationally practical for tabular sensor data. A multilayer perceptron or other deep model may be added when larger or sequential datasets are available. Published results on the AI4I dataset show that ensemble models are strong candidates for early failure detection, with Deep Forest and Gradient Boosting achieving reported accuracy above 90% (Hosseinzadeh et al., 2023). The proposed workflow therefore uses an ensemble model as the primary predictor but retains a simpler baseline to detect whether model complexity produces a meaningful gain.

Table 2. Framework layers, inputs, and outputs

Layer	Primary input	Method	Output
1. Acquisition	Sensor/process data	IIoT, PLC, historian	Timestamped condition data
2. Preparation	Raw records	Cleaning, encoding, feature engineering	Model-ready features
3. Prediction	Prepared features	LR, RF/GBM, optional DL	Failure probability
4. Prognostics	Risk history / degradation	Thresholding, trend/RUL logic	Risk level / maintenance horizon
5. Explainability	Prediction + features	Feature importance / local explanation	Reason for alert
6. Action & feedback	Risk + business rules	Work-order decision and outcome logging	Maintenance action + retraining data

5. Evaluation Metrics and Decision Logic

Accuracy alone is not an adequate metric because a model can achieve high accuracy by predicting the dominant normal class. The primary evaluation measures are precision, recall, F1-score, area under the receiver-operating characteristic curve (ROC-AUC), and false-negative rate. Recall is particularly important because it measures the proportion of actual failures detected. Precision controls unnecessary maintenance alerts. F1-score balances the two, while ROC-AUC summarizes ranking performance across decision thresholds.

A maintenance threshold is selected on the validation set rather than fixed automatically at 0.50. The threshold should minimize a cost function in which false negatives receive a higher penalty than false positives. Predictions are then converted into three operational categories: normal, inspection required, and high-risk

maintenance. Explainability is attached to the inspection and high-risk categories so that the maintenance engineer sees the main factors associated with the prediction. The threshold and model are reviewed periodically using observed maintenance outcomes, thereby closing the feedback loop.

Results

The methodological evaluation shows that the proposed framework addresses the major limitations repeatedly identified in predictive-maintenance literature: fragmented data processing, class imbalance, weak interpretability, and lack of feedback between prediction and maintenance action. The AI4I benchmark provides a controlled environment for validating the pipeline because it combines thermal, mechanical, and wear-related variables with multiple failure mechanisms and a low overall failure rate.

Table 3 summarizes the expected role of each model family and aligns the proposed framework with findings reported in the literature. The purpose of the table is not to claim new plant measurements; rather, it defines the benchmark interpretation for implementation. Tree-based ensembles are selected as the principal tabular-

data model because they can model nonlinear interactions without extensive scaling and have shown strong results on AI4I. Logistic regression remains useful as a transparent baseline, while deep models become more attractive when high-frequency sequences or larger sensor histories are available.

Table 3. Methodological results and benchmark interpretation

Model / element	Expected strength	Primary metric	Evidence / result	Framework decision
Logistic regression	Transparent baseline	Recall, F1	Interpretable reference model	Retain for comparison
Random Forest / GBM	Nonlinear tabular learning	Recall, F1, ROC-AUC	Ensemble models reported >90% accuracy on AI4I in 2023 study	Primary model family
Deep / hybrid model	Complex feature learning	Recall, F1	Useful when richer sequential data are available	Optional extension
Explainability layer	Trust and actionability	Explanation stability	XAI emphasized for PdM deployment	Required for alerts
Feedback / drift loop	Long-term reliability	Drift and alert quality	Addresses changing operating regimes	Required in production

The benchmark also demonstrates why imbalance-aware metrics are essential. With 339 failures in 10,000 records, a naive classifier predicting only normal operation would exceed 96% accuracy while detecting no failures. Therefore, any publishable evaluation of the framework should report the confusion matrix together with recall, precision, F1-score, and false-negative rate. This observation is consistent with the broader literature, which emphasizes task constraints and the trade-off between diagnostic techniques rather than relying on a single aggregate metric (Surucu et al., 2023).

At the system level, the framework produces three outcomes: an estimated risk, an explanation, and an action class. This is an improvement over a classifier-only design because maintenance personnel need to know not only that a failure is likely, but also which operational variables justify attention and what action should follow. The feedback mechanism further allows false alerts, confirmed degradations, and post-maintenance conditions to become labeled evidence for subsequent model updates.

Conclusion and Discussion

This paper proposed an AI-based predictive maintenance framework for smart manufacturing systems and organized the

complete methodology from IIoT data acquisition to maintenance feedback. The framework integrates data preparation, AI failure prediction, risk and RUL-oriented reasoning, explainability, and maintenance decision support. The design is consistent with the evolution from condition-based maintenance to intelligent prognostics described in earlier research and with more recent work on ML, explainable AI, and digital twins.

The main methodological implication is that predictive maintenance should not be treated as an accuracy-maximization problem. Manufacturing failure data are imbalanced, operational costs are asymmetric, and model predictions must be interpretable enough to support action. Consequently, recall, false-negative rate, F1-score, and explanation quality should be considered alongside accuracy. For tabular manufacturing data such as AI4I, tree-based ensembles provide a strong baseline; however, deep and hybrid models may become more valuable when high-frequency temporal data, images, acoustic signals, or multivariate degradation histories are available.

The study is limited by its use of a public synthetic benchmark for framework design rather than proprietary plant data. Synthetic data are useful for reproducibility but cannot reproduce every source of sensor noise, maintenance practice,

operating-regime change, or organizational constraint present in factories. Future work should therefore validate the framework on a real production asset, incorporate time-dependent RUL prediction, compare cost-sensitive and resampling strategies, and evaluate model drift after maintenance interventions. Integration with a digital twin could further support what-if analysis and maintenance scheduling, as suggested by recent digital-twin predictive-maintenance research (Mi et al., 2021; Chen et al., 2023).

References

- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- 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>
- Ferreira, C., & Gonçalves, G. (2022). Remaining useful life prediction and challenges: A literature review on the use of machine learning methods. *Journal of Manufacturing Systems*, 63, 550-562. <https://doi.org/10.1016/j.jmsy.2022.05.010>
- Hafsi, M., Hamour, N., & Ouchani, S. (2023). Predictive maintenance for smart industrial systems: A roadmap. *Procedia Computer Science*, 220, 645-650. <https://doi.org/10.1016/j.procs.2023.03.082>
- Hosseinzadeh, A., Chen, F. F., Shahin, M., & Bouzary, H. (2023). A predictive maintenance approach in manufacturing systems via AI-based early failure detection. *Manufacturing Letters*, 35, 1179-1186. <https://doi.org/10.1016/j.mfglet.2023.08.125>
- Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510. <https://doi.org/10.1016/j.ymssp.2005.09.012>
- Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 57(6), 476-489. <https://doi.org/10.1016/j.compind.2006.02.014>
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799-834. <https://doi.org/10.1016/j.ymssp.2017.11.016>
- Matzka, S. (2020). Explainable artificial intelligence for predictive maintenance applications. 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), 69-74. <https://doi.org/10.1109/AI4I49448.2020.00023>
- Mi, S., Feng, Y., Zheng, H., Wang, Y., Gao, Y., & Tan, J. (2021). Prediction maintenance integrated decision-making approach supported by digital twin-driven cooperative awareness and interconnection framework. *Journal of Manufacturing Systems*, 58, 329-345. <https://doi.org/10.1016/j.jmsy.2020.08.001>
- Surucu, O., Gadsden, S. A., & Yawney, J. (2023). Condition monitoring using machine learning: A review of theory, applications, and recent advances. *Expert Systems with Applications*, 221, 119738. <https://doi.org/10.1016/j.eswa.2023.119738>
- UCI Machine Learning Repository. (2020). AI4I 2020 Predictive Maintenance Dataset [Dataset]. <https://doi.org/10.24432/C5HS5C>