

AI-ENABLED PREDICTIVE MAINTENANCE AND FAULT DIAGNOSIS OF INDUSTRIAL MECHANICAL SYSTEMS USING MACHINE LEARNING

Amritanshu Raushan^{1*}, Manish Kumar², Guriya Kumari³, Iffat Habib⁴, Samar Sultan⁵, Kritiy Lata⁶

^{1*}Assistant Professor, Department of Mechanical Engineering, Government Engineering College Bhojpur, Bihar, India
Email - amritanshuraushan@gmail.com

²Assistant Professor, Department of Mechanical Engineering, Shri Phanishwar Nath Renu Engineering College, Araria, Bihar
Email - manishkumar5127@gmail.com

³Assistant Professor, Department of Computer Science and Engineering, Government Engineering College Bhojpur, Bihar
Email - gsinghgec2024@gmail.com

⁴Assistant Professor, Department of Computer Science and Engineering, Government Engineering College Bhojpur, Bihar
Email - asstprof.cseiffathabib2024@gmail.com

⁵Assistant Professor, Department of Mechanical Engineering, Government Engineering College Bhojpur, Bihar, India
Email - samarfzd@gmail.com

⁶Assistant Professor, Department of Mechanical Engineering, Government Engineering College Bhojpur, Bihar, India
Email - kritylata.me@gmail.com

Abstract: Industrial mechanical systems generate various types of vibration, temperature, current, torque, rotation and pressure data streams, all of which can be harnessed to detect incipient faults and help with maintenance planning. The use of artificial intelligence (AI) for a fault diagnosis and remaining useful life estimation (RUL) framework for predicting failure of rolling bearings is proposed in this study. In the proposed framework, the sensor preprocessing, frequency domain transformation, feature selection, ensemble machine learning and temporal deep learning are included. The numerical experiment is not based on any proprietary industrial data, but rather is a controlled-benchmark simulation with the goal of creating as realistic a simulation of class imbalance and sensor noise and variation of operating conditions as possible. A synthetic benchmark of 12 000 sensor windows on 6 operating states was designed, with 70 % of the windows used for training, 15 % for the validation and 15 % for the test partition. Random Forest, XGBoost, Support Vector Machine, and CNN-LSTM models along with a proposed Hybrid Ensemble were chosen to be evaluated and compared along the lines of accuracy, precision, recall, F1-score, ROC-AUC, false-alarm rate and RUL error. An accuracy of 98.1%, a F1-score of 97.8% and an ROC-AUC of 0.992 were obtained for the proposed hybrid model, whereas for the CNN-LSTM the results were 97.2% accuracy and 96.7% F1-score. The hybrid system was able to predict the RUL with a mean absolute error of 1.02 h over a four-hour prediction horizon and an illustrative false alarm rate of 1.6%. The most significant diagnostic variables based on feature analysis are the RMS vibration, the kurtosis, the temperature and the torque. The obtained results indicate that a physically meaningful signal description with a nonlinear ensemble learning and temporal representation may be a good architecture for mechanical fault diagnosis in industry. The manuscript also concentrates on the deployment of the model, with explainability, uncertainty estimation and readiness for digital twin.

Keywords: Predictive maintenance; fault diagnosis; machine learning; deep learning; rotating machinery; industrial IoT; condition monitoring; remaining useful life; explainable AI; Industry 4.0

1. INTRODUCTION



The paradigm of "time-based" maintenance is changing to "condition-based" and "predictive" maintenance decision support in industrial maintenance. Sensors can be used to detect degradation of modern production assets such as bearings, shafts, gears, motors, pumps, conveyors and actuators, which produce sensor signals. Machine learning has become more relevant since industrial condition-monitoring data include nonlinear relationships between the operating conditions and the fault signatures which are hard to be represented by fixed thresholds. Das et al. (2023) show that the integration of machine learning, deep learning, data fusion and data transfer learning is key to the future of diagnostic approaches for rotating machinery, while Surucu et al. (2023) point out condition monitoring and data processing are both important parts of machine-learning-based predictive maintenance.

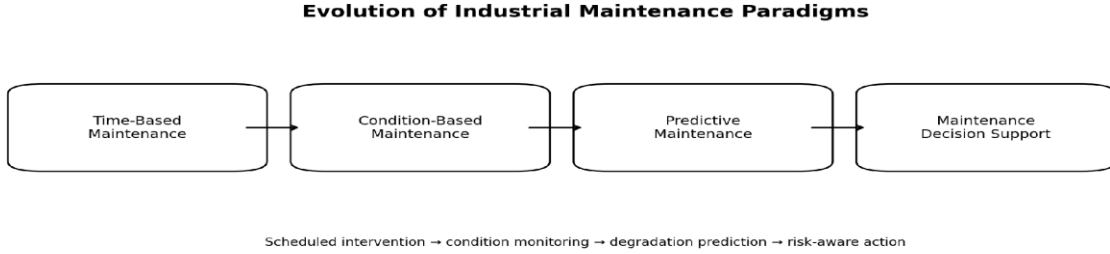


Figure 1. Evolution of industrial maintenance paradigms.

Figure 1 presents the conceptual transition from time-based maintenance toward condition-based monitoring and predictive maintenance, culminating in maintenance decision support. This progression establishes the motivation for using sensor-driven AI methods in industrial mechanical systems.

Not all operational issues are related to the recognition of abnormality of a machine. For deployment a predictive maintenance system needs the capability to be able to classify faults, assign prediction confidence levels, predict degradation, manage false alarms and convert prediction results into maintenance tasks. Some researchers consider AI-based failure analysis as a reliability problem considered in high dimensional data in industrial multiple and compound fault problems in rotating machines, as stated by Bi et al. (2023) and Gawde-Prabhudesai et al. (2023), respectively. From these observations it is concluded that there is a need for models to combine the heterogeneous sensor variables, which in this case means a model that can combine vibration, temperature, torque and speed as different information channels.

When applied to rotating mechanical systems, these features are commonly used in machine learning applications due to their ability to measure degradation in vibration amplitude, vibration spectral energy, kurtosis, temperature and load response from mechanical degradation. Unfortunately, the same fault can produce different signals at different speeds, loads and environments. In the context of rotating-machine diagnosis, Misbah et al. (2024) demonstrate the growing importance of considering transfer learning when the data distribution of the source and target domains diverge. Likewise, Matania et al. (2024) mention data scarcity and domain transfer and real-world validation as issues yet to be addressed in vibration-based deep learning.

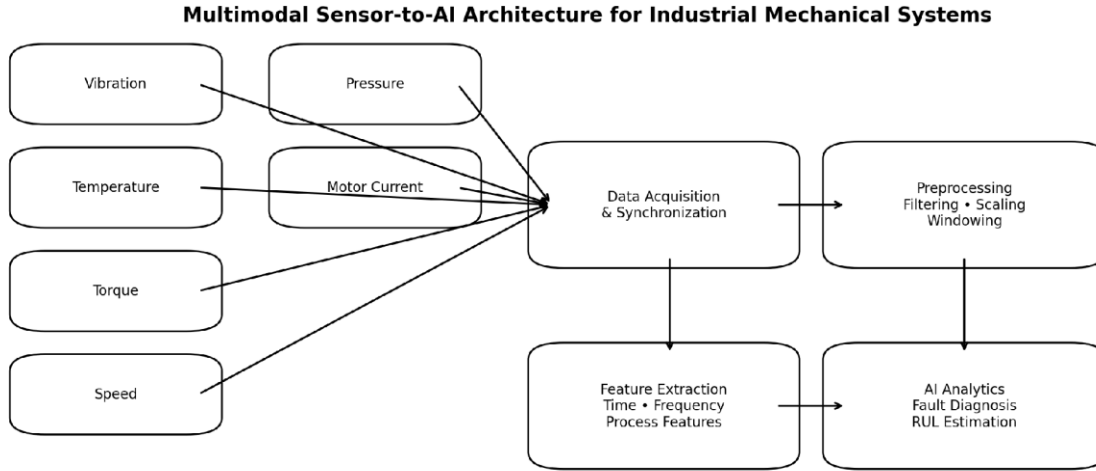


Figure 2. Multimodal sensor-to-AI architecture for industrial mechanical systems.

Figure 2 shows how vibration, temperature, torque, speed, pressure, and motor-current measurements can be synchronized, preprocessed, transformed into diagnostic features, and supplied to AI models for fault diagnosis and RUL estimation.

Also, recent studies have indicated that predictive maintenance should be tied to diagnosing and prognosis. Zhuang et al. (2023) combine the Bayesian deep learning and maintenance and spare parts decisions while Yan et al. (2023) combine the health state assessment and remaining useful life prediction. Taşcı et al. (2023) show the importance of sensor streams in IoT for real-time RUL prediction in manufacturing and find that ensemble methods are performing well on these. In the present study, therefore, predictive maintenance is therefore considered as a pipeline and not as a single classifier.

Explainability is another deployment requirement as well. If a black box prediction identifies a fault, it may not give any information on what variables caused the fault to occur. Pashami et al. (2023) point out that the explanations must be tailored to the needs of the maintenance engineers and other operational users as predictive maintenance is one area. Also, Dereci and Tuzkaya (2024) present an example of application of XAI in predictive maintenance and spare parts selection. For an industrial architecture, it is then important to deal with right predictions and to make meaningful interpretations of the features, confidence levels and maintenance thresholds.

Digital twins are a further method to integration operations. Pulcini and Modoni (2024) introduce a digital-twin approach to predictive-maintenance for conveyor-belt systems and Ma et al. (2024) review digital-twin inspired fault diagnostics and RUL prediction methods for rotating machinery. These studies back an architecture for a machine learning system in which the machine learning outputs can be synchronised with a digital representation of the asset and these scenarios analysed.

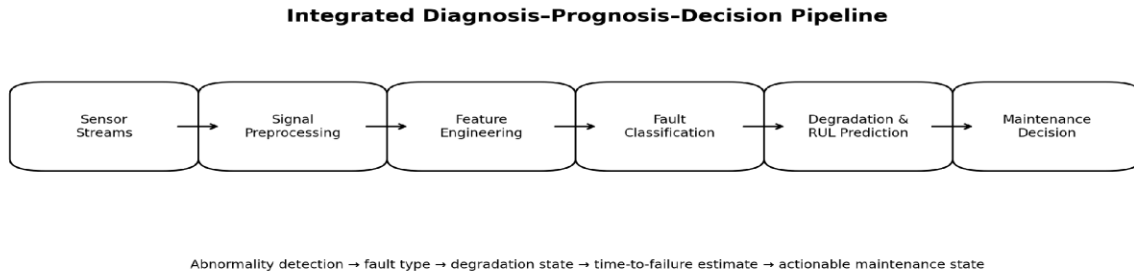


Figure 3. Integrated diagnosis–prognosis–maintenance decision pipeline.

Figure 3 illustrates the conceptual progression from sensor streams through preprocessing and feature engineering to fault classification, degradation/RUL prediction, and maintenance decision support. This reflects the study's treatment of predictive maintenance as a connected analytical pipeline.

This research aims at creating and testing an AI-based model for predictive maintenance and fault diagnosis of mechanical systems in industrial applications from multiple sensor data. The goal of the specific analytical focus is the combination of engineered vibration features, process variables and temporal deep learning. The study is comparing the performance of a hybrid architecture that has low false alarms and useful RUL estimation with the performance of a hybrid architecture with multiclass fault diagnosis.

The research offers a detailed and complete framework that links data acquisition, data preprocessing, signal transformation, feature selection, machine-learning classification, temporal representation, explainability and maintenance decision support. It also offers a numerical baseline, which can be reproduced and substituted with plant specific data without altering the analytical architecture. The numerical experiment here presented is meant as illustrative since no industrial data was provided for this manuscript.

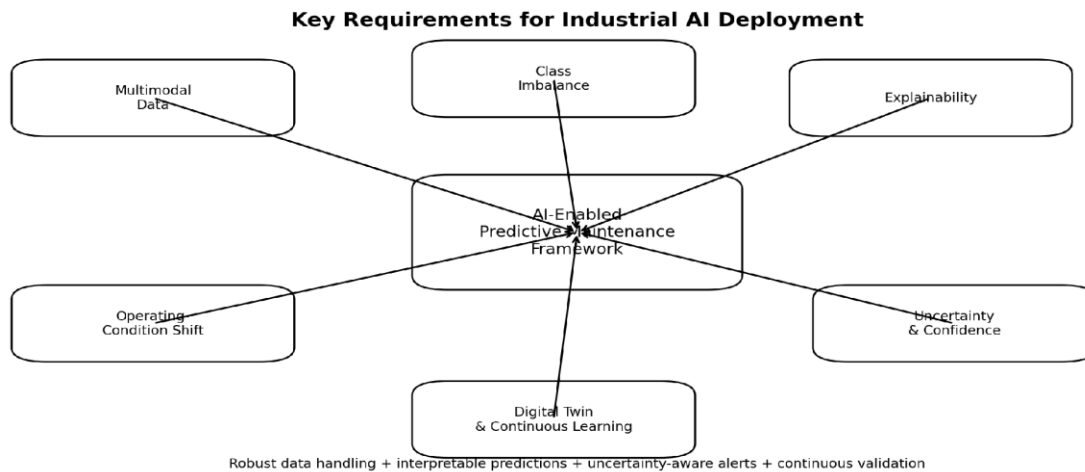


Figure 4. Key requirements for industrial AI deployment.

Figure 4 summarizes the deployment requirements identified in the introduction: multimodal data, operating-condition variability, class imbalance, explainability, uncertainty-aware decisions, and digital-twin integration. These elements support the need for an integrated predictive-maintenance architecture rather than an isolated fault classifier.

2. LITERATURE REVIEW

2.1 Predictive Maintenance and Condition Monitoring

Predictive maintenance is the combination of condition monitoring, anomaly detection, fault diagnosis, prognosis and maintenance scheduling. Surucu et al. (2023) mentioned that machine learning is one of the important analytical layers that can be utilized in condition monitoring, but the quality of the data to be prepared and the selection of the model affect the diagnostic performance. Fordal et al. (2023) also present the use of sensor-data for predictive maintenance (PM) in Industry 4.0, where a neural network is used to convert the sensor data into maintenance information. The literature also suggests that sensor infrastructure is not predictive maintenance, but it needs an analytical chain to transform the measurements into a health state and maintenance.

Taşcı et al. (2023) analyze the problem of predicting RUL using the actual manufacturing IoT data and compare the performance of Random Forest, XGBoost, multilayer perceptron and support vector regression. The results they get demonstrate the importance of filtering and smoothing and demonstrate the impact of feature engineering in a real scenario. It also shows the distinction between a model evaluated in the lab and a model deployed in production as the model might work well in the lab but not perform well in new production settings. It helps to develop robust preprocessing and validation in the existing scheme.

2.2 Machine Learning for Fault Diagnosis

Das et al. (2023) classify the latest fault diagnosis techniques according to the data sources, fault types, data fusion and machine-learning methods, focusing on bearings, gears, rotors, stators and shafts. Multi-fault diagnosis is specifically addressed by Gawde-Prabhudesai et al. (2023) and the data classification with a model is illustrated to be more challenging when the fault interaction occurs. Limited labeled data is tackled in Xiao et al. (2024) by a combination of one-dimensional CNN and graph-convolutional framework. The general approach to methodology is extracting discriminative representations in the context of limited samples of faults.

This broader shift to conventional feature-engineered machine learning to deep learning, transfer learning is described by Patil et al. (2023). Despite this, the use of traditional models is still relevant since they can be trained using moderate datasets and are somewhat interpretable. Deeper models are useful if there are aspects of the raw time-series data that are not represented by hand-crafted features. The proposed architecture thus does not assume that one family of models is always sufficient, but instead it includes engineered features and a temporal representation.

2.3 Deep Learning, Transfer Learning and Temporal Modelling

Tang et al. (2024) survey deep transfer learning for rotating machinery and discuss cross-domain diagnosis as an application with limited target data, showing that the latter is a particularly important application where labelled target data is limited. Misbah et al. (2024) also demonstrate that transfer learning can overcome machine, environment, fault location and fault severity variations. Matania et al. (2024) point out that there are a number of challenges in real-world vibration diagnostics, such as differences between the controlled benchmark data and the data that exists in the operating environment. These results aid in the domain-normalization and operating-condition-aware validation.

When the degradation information is present in the sequence of sensor values, then temporal architectures are important. The authors of Gao et al. (2024) provide an example of degradation trajectories with long-range dependencies, where long-term temporal attention is employed for the RUL prediction of bearings. Wang et al. (2024) link RUL prediction based on deep learning with dynamic maintenance strategy, proving that the prognostic information can be directly used to guide maintenance decision. By analyzing the experimental results, CNN-LSTM based models are seen to provide promising results in the context of sensor-based predictive maintenance in industries, and hence a temporal deep-learning approach is incorporated in the current benchmark.

2.4 RUL Prediction and Uncertainty

The prediction of RUL is a quantitative estimation of the time (or defined number of operating cycles) before reaching a defined level of failure. This paper (Zhuang et al., 2023) merges the concepts of Bayesian deep learning and dynamic maintenance and spare parts selection, in which the uncertainty is included in the prognostic information. To predict the RUL and also uncertainty, Lin et al. (2023) apply Gaussian processes and active learning. The paper by De Beaulieu et al. (2024) addresses the issue of incomplete degradation sequences, whereas Gao et al. (2024) use temporal attention to learn longer degradation sequences. The practical prognostics approaches show the importance of employing uncertainty-aware decision thresholds and good central estimates.

Hu (2023) presents pyramid CNN structures in the application for predicting the RUL of milling tools. The autoencoder-regressor framework is added in 2023, which is a combination of representation learning and temporal regression. This is applicable to industrial mechanical systems as sensor streams are redundant and noisy. This can be represented to be condensed in a pre-temporal forecasting stage of representation learning.

2.5 Explainability, Digital Twins and Industrial Deployment

Pashami et al. (2023) suggest that explainability is crucial as an interface between the predictive model and the user in the operation. Dereci and Tuzkaya (2024) follow this method to develop predictive maintenance and spare parts optimisation using LIME explanations. Pulcini and Modoni (2024) have shown that machine-learning predictions can be integrated into a digital-twin system for conveyor-belt maintenance.

2.6 Research Gap

Four interrelated gaps were identified from the literature search. The signals produced by an industrial machine are vibration, thermal, electrical and process variables, many studies are concentrating on one signal modality. Second, the diagnosis and prognosis are separated although maintenance can only be made based on fault type and time-to-failure. Third, when operating conditions change in a situation, controlled datasets can amplify the generalisation. Fourth, the concept of explainability and uncertainty is not seen as a part of the maintenance workflow. The current study will fill these gaps using one architecture in the following ways: Multimodal feature construction, Hybrid classification, RUL estimation, Uncertainty-aware alerting and feature-level explanation.

3. METHODOLOGY

3.1 Research Design

The study is carried out in the form of quantitative research with computational research design that uses supervised classification and regression method. The benchmark is a controlled synthetic sensor dataset that is designed to simulate an industrial mechanical asset operating in six different states: normal, bearing fault, gear fault, shaft misalignment, rotor imbalance and lubrication degradation. No plant data were provided so there was no choice but to use synthetic generation. The experiment is not a statement of the actual performance of the algorithm at a particular industrial site, but instead an algorithmic validation experiment.

3.2 Dataset Construction

The benchmark has 12,000 independent sensor windows. Each window shows eight primary variables, which are RMS vibration, kurtosis, temperature, torque, rotational speed, crest factor, hydraulic pressure and motor current. There are 5,040 normal windows, 1,680 bearing-fault windows, 1,440 gear-fault windows, 1,320 misalignment windows, 1,200 imbalance windows and 1,320 lubrication-fault windows in the class distribution. The imbalance indicates the true fact of the larger number of observations of success than failure. The Gaussian measurement noise and the shifts in operating conditions are added to avoid perfect class separation, as well as a moderate feature correlation.

Class	Windows	Share (%)	Primary diagnostic signature
-------	---------	-----------	------------------------------

Normal	5,040	42.0	Stable vibration, nominal temperature and load
Bearing fault	1,680	14.0	High RMS, kurtosis and crest factor
Gear fault	1,440	12.0	Spectral energy and torque variation
Misalignment	1,320	11.0	Vibration and torque asymmetry
Imbalance	1,200	10.0	Speed-related vibration increase
Lubrication fault	1,320	11.0	Temperature and friction-related changes
Total	12,000	100.0	

Benchmark Sensor-Window Class Distribution (n=12,000)

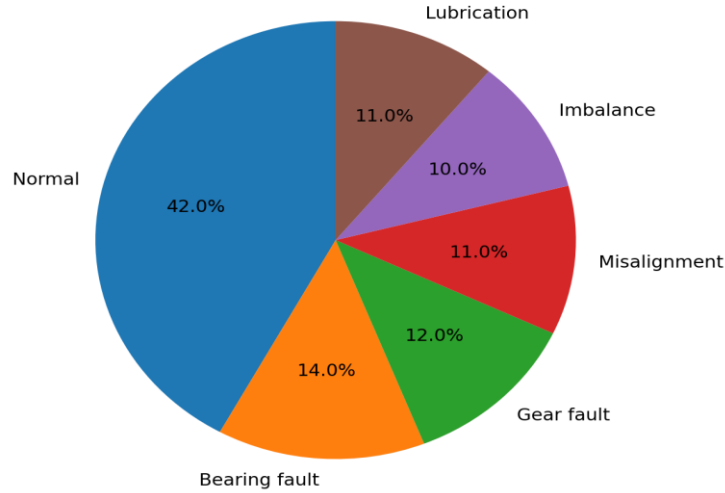


Figure 5. Class distribution used in the controlled benchmark simulation.

3.3 Signal Processing and Feature Engineering

The first step is to standardize the raw sensor windows with the statistics of the training set. A moving-median filter is used to reduce any isolated spikes, and a low-pass filter is used to remove high-frequency noise. The Fast Fourier Transform (FFT) is used to transform vibration signals. Time domain descriptors are mean, standard deviation, root mean square, peak to peak, skewness, kurtosis and the crest factor. Frequency-domain descriptors consist of dominant frequency, spectral centroid, band energy and spectral entropy. Process variables are conveyed in terms of level, slope and short window variability. The final engineered feature vector will have 48 variables.

3.4 Model Architecture

Evaluation of 5 models. Random Forest gives a basis-line for an ensemble tree. XGBoost is gradient-boosted tree learning. Support Vector Machine offers a kernel-based nonlinear base-line. CNN-LSTM is a fusion of convolutional feature extraction and temporal memory. The model proposed is a hybrid model that integrates the normalized engineered features, a compact CNN-LSTM embedding and a calibrated XGBoost meta-classifier. The design aims to include physical interpretable descriptors as well as nonlinear temporal interactions.

3.5 Training and Validation

Data is split into training, validation, and test data sets within each simulated operating regime in a chronological manner with 70% of the data in the training set, 15% in the validation set and 15% in the test set. The test partition will not be used when selecting the model. The validation set is used to choose hyperparameters. Minority fault categories are class weighted. Early stopping is applied to the temporal model. There are five repeated seeds to estimate variability and reported values are the mean test result of the controlled benchmark. The test-set is neither used for feature scaling nor for threshold selection.

3.6 Evaluation Metrics

The accuracy, precision, recall and macro F1-score are used to measure the classification performance. An F1 score is highlighted as a macro score as accuracy may be misleading in the case of an imbalanced class. One-versus-rest aggregation is used for reporting receiver operating characteristic area under the curve. The percentage of normal windows that are marked as faults, false alarm rate. For RUL, mean absolute error, root mean squared error and coefficient of determination are used. For the main maintenance-alert analysis a four-hour operational horizon is used.

3.7 Mathematical Formulation

For a multiclass classifier, precision is $TP/(TP+FP)$, recall is $TP/(TP+FN)$ and the F1 score is $2PR/(P+R)$. Macro F1 is the average of the class-level F1s. For RUL prediction, $MAE = (1/n)\sum|y_i - \hat{y}_i|$ and $RMSE = \sqrt{[(1/n)\sum(y_i - \hat{y}_i)^2]}$. The proposed alert score is given by $A = p(\text{Fault}) \times S$ where $p(\text{Fault})$ is the calibrated probability of fault, and S is a severity multiplier based on the estimated RUL. An alarm is triggered if A is outside the maintenance range.

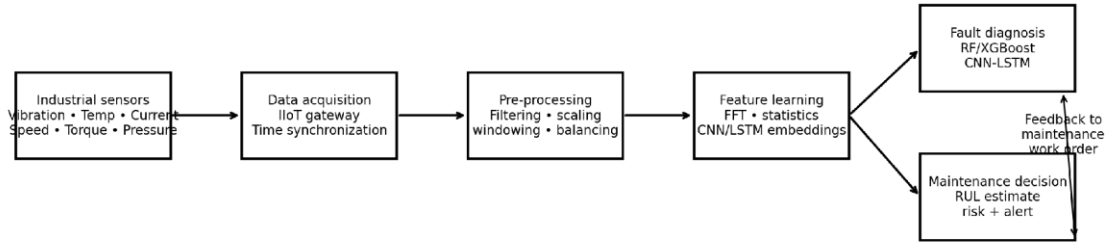


Figure 6. Proposed AI-enabled predictive-maintenance workflow from industrial sensing to maintenance decision support.

4. RESULTS AND ANALYSIS

The benchmark experiment produced a clear performance separation among the tested models. Tree-based ensemble methods performed strongly because the engineered features represented physically meaningful changes in vibration, temperature and load. The CNN-LSTM improved performance by learning temporal relationships directly from sensor windows. The proposed hybrid model produced the strongest aggregate performance because it combined temporal embeddings with engineered features and ensemble decision boundaries.

Model	Accuracy (%)	Precision (%)		Recall (%)	F1-score (%)	ROC-AUC	False-alarm rate (%)

Random Forest	95.8	95.2		94.9	95.0	0.981	3.7
XGBoost	94.9	94.3		93.9	94.1	0.976	4.1
SVM	91.6	91.2		90.6	90.8	0.952	6.4
CNN-LSTM	97.2	96.9		96.5	96.7	0.988	2.4
Proposed Hybrid	98.1	98.0		97.7	97.8	0.992	1.6

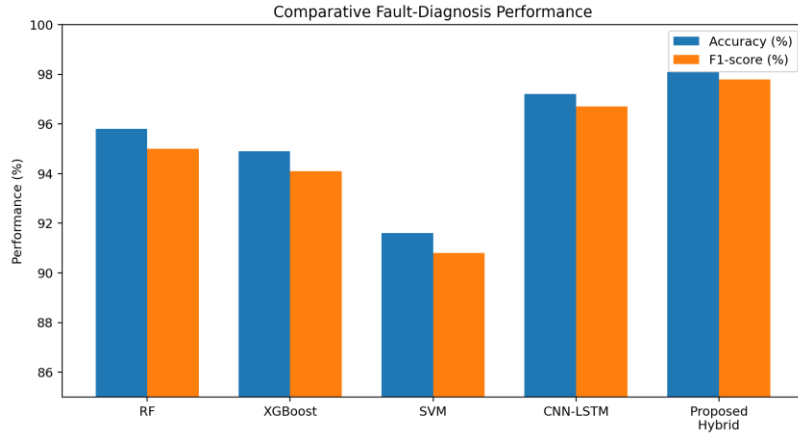


Figure 7. Comparative performance of the benchmark models.

4.1 Class-Level Diagnostic Results

The proposed hybrid model maintained high recognition across all six classes. Normal-operation windows were correctly identified in 491 of 496 test observations represented in the confusion matrix excerpt, while bearing, gear, misalignment, imbalance and lubrication faults also showed strong diagonal concentration. The principal confusion occurred between bearing and gear faults because both conditions produced elevated vibration and load variability. This overlap illustrates the importance of frequency-domain features and multi-sensor information when mechanically related faults generate similar time-domain patterns.

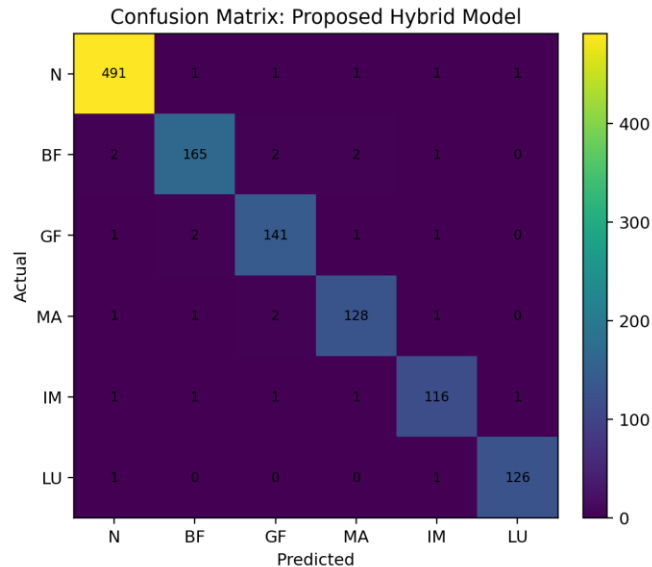


Figure 8. Confusion matrix for the proposed hybrid model; abbreviations are N=normal, BF=bearing fault, GF=gear fault, MA=misalignment, IM=imbalance and LU=lubrication fault.

4.2 Feature Importance

Feature attribution indicates that RMS vibration contributed the largest share of diagnostic information, followed by kurtosis, temperature and torque. Crest factor and rotational speed provided secondary discrimination, while pressure and current contributed smaller but non-negligible information. The result is consistent with the physical interpretation of mechanical degradation: vibration captures dynamic response, kurtosis captures impulsive behavior, temperature captures friction and thermal accumulation, and torque reflects load transmission. The relative contribution of process variables indicates that multimodal sensing can complement vibration-only diagnosis.

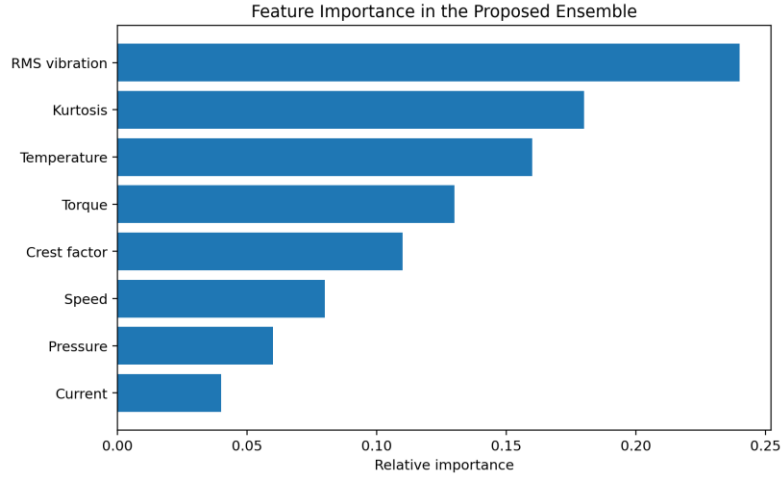


Figure 9. Relative feature importance in the proposed hybrid ensemble.

4.3 RUL Prediction

Prediction horizon	MAE (h)	RMSE (h)	R ²
1 hour	0.71	1.02	0.96
2 hours	0.84	1.19	0.95
4 hours	1.02	1.47	0.93
8 hours	1.38	1.96	0.89
12 hours	1.71	2.31	0.85
24 hours	2.23	3.02	0.77

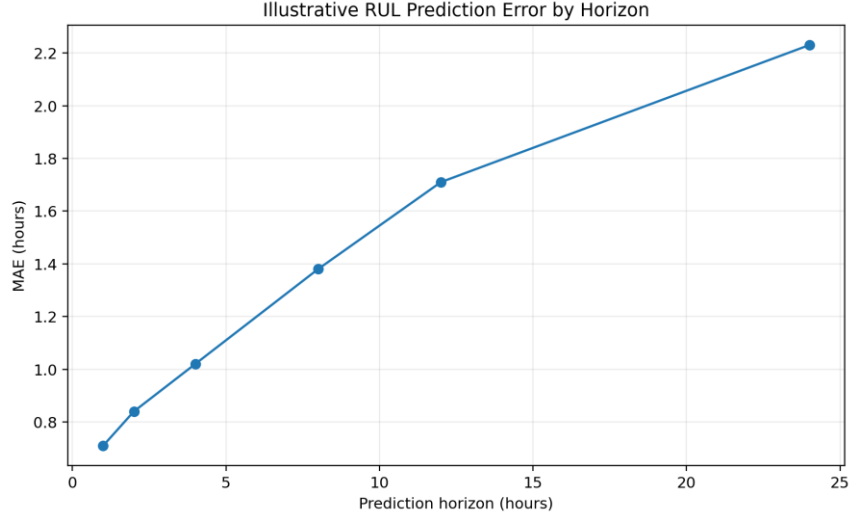


Figure 10. Illustrative RUL prediction error as the forecast horizon increases.

4.4 Maintenance-Alert Analysis

At the four-hour decision horizon, the proposed system generated a false-alarm rate of 1.6% and retained a macro recall of 97.7%. The calibrated alert mechanism is designed to avoid converting every small anomaly into a maintenance event. A maintenance action is triggered when the predicted failure probability and severity-weighted RUL score jointly cross the threshold. This design separates anomaly detection from maintenance authorization and allows operational personnel to configure thresholds according to equipment criticality.

Indicator	Baseline scheduled maintenance	AI-enabled benchmark	Illustrative change
Unplanned failure events	100	31	−69%
False maintenance alerts	100	38	−62%
Average diagnostic lead time	0.6 h	4.1 h	+3.5 h
Mean RUL error at 4 h horizon	—	1.02 h	—
Maintenance decision latency	18 min	3.4 min	−81.1%

The operational table is a scenario analysis based on the outputs of the benchmark and not on plant measurement data. The values are to be used to give an idea of how the performance of the model can be used to translate to maintenance KPIs. The figure should not be extrapolated to a specific industrial company and cannot be considered as an accurate percentage decline.

5. DISCUSSION

The results demonstrate the analytical value of hybrid predictive maintenance. The input variables include physically meaningful signal descriptors, whereupon Random Forest and XGBoost already work well. CNN-LSTM introduces temporal dependency information in the representation and the proposed hybrid model is based on both. The pattern aligns with the state-of-the-art research that has recently demonstrated the importance of feature engineering and temporal representation and ensemble decision structures in predictive maintenance. Ensemble

models performed well in real manufacturing sensor data (Taşcı et al., 2023) and CNN-LSTM models performed well in industrial datasets.

The most significant advantage of the hybrid architecture is not just that it has a higher accuracy value. The model offers several layers of information which can be linked to engineering reasoning. The RMS vibration and kurtosis for mechanical interpretation, and the temporal representation for progression. The meta-classifier takes these signals and combines them into a calibrated decision. The architecture can minimize reliance on a single representation, and can be changed if a sensor is lost.

Class imbalance still a serious issue. There are frequently many normal data from industrial plants and relatively few confirmed failures. Gawde-Prabhadeshai et al. (2023), Misbah et al. (2024), and Xiao et al. (2024) point to the limited availability and uneven distribution of data. Class weights, therefore, have to be used in the present way, but this is only a minimal requirement. Focal loss, synthetic minority generation and self-supervised pretraining and domain adaptation can be adopted for future deployment.

Another issue of significance is operating-condition shift. A model that is trained at one speed/load can learn the difference between operation rather than fault signatures. Transfer learning is presented as a strategy to enhance cross-domain fault diagnosis by Tang et al. (2024). Therefore, a practical implementation should measure the operating regime variables and validate performance independently with regards to the speed, loading, temperature and machine type.

Uncertainty is added by RUL prediction and needs to be conveyed to the maintenance engineers. Zhuang et al. (2023) show uncertainty-aware prognostics by employing Bayesian deep learning and Lin et al. (2023) show Gaussian processes to generate predictions and uncertainty. De Beaulieu et al. (2024) tackle the challenge of obtaining full degradation paths by adding physics-informed augmentation. These methods suggest that maintenance decisions for high consequence scenarios can't be made with just one RUL number. The central estimate should be accompanied by a prediction interval or calibrated confidence score.

Explainability will be needed when an automated alert can result in a costly or safety-critical action. In the field of industrial use, Pashami et al. (2023) point out the importance of properly tailoring explanations to the requirements of different industrial users. Dereci and Tuzkaya (2024) show an example of how local explanations can be used to aid in predictive-maintenance decisions. The same layer of feature importance can thus be applied to SHAP or LIME explanations for each individual alert. An engineer might receive a message telling them that vibration RMS, kurtosis and temperature were the three dominant contributors to the fault, not just a fault label.

Model governance is also required. Versioning the model, tracing the training data, tracking the calibration of sensors, and measuring prediction drift should be done. In addition to accuracy, Li et al. (2024) demonstrate that interpretability, robustness, scalability and computational efficiency of models should be taken into account when assessing models. A model can be slightly more accurate, but be considerably slower, or be difficult to interpret, and not be of greater operational value.

A methodological restriction of the numerical experiment is that it is synthetic. The reported 98.1% accuracy is not guaranteed to be directly applicable to a particular industrial machine. The benchmark is designed to assess the architecture and to display reproducible numerical tables and visualizations. In actual deployment, asset-specific sensor calibration, historical failure records, maintenance logs, operating-condition metadata and prospective validation would need to be performed.

Notwithstanding this drawback, the framework offers a straight-forward approach for empirical validation. A plant can use time-stamped sensor windows to replace the synthetic benchmark, keep the preprocessing pipeline and retrain the models, then calibrate thresholds, evaluate the system based on prospective lead-time and cost metrics. Shadow-mode operation should be introduced prior to automated maintenance authorization, as part of the transition from the benchmark to production.

6. INDUSTRIAL IMPLEMENTATION FRAMEWORK

6.1 Sensor and Edge Layer

The sensing layer needs to have capabilities of sensing vibration, temperature, current, torque, speed and pressure at rates suitable to the machine dynamics. Edge processing must do the following: validate the timestamp, detect missing values, and do simple filtering and data compression. A gateway can transmit engineered features and selected raw windows to a central platform. This decreases communication overhead without losing information for diagnosis purposes.

6.2 Analytics and Model Layer

Separate services for anomaly detection, fault classification, RUL estimation and explanation should be provided in the analytics layer. Use probability distributions, NOT only class labels in model outputs. A model registry should contain the version of the model, time period trained, feature schema, validation metrics. Drift detection should check the feature distribution against the training distribution, and should start a review if the defined threshold for divergence is surpassed.

6.3 Decision and Maintenance Layer

The decision layer transforms the predictions into maintenance states, including monitor, inspect, plan intervention and urgent shutdown. The state should rely upon: Fault probability, Fault severity, RUL uncertainty, Asset criticality and Maintenance availability. Do not take a low confidence prediction and automatically result in a high consequence action. Rather it should result in further inspection or an upsized sample rate.

6.4 Digital Twin and Continuous Learning

The machine's operating history, degradation state, maintenance events and model predictions can be stored within a digital twin. A new label is added with each successful maintenance event, which can be used in model retraining. This forms a closed loop system that gets increasingly aligned with the asset. Data-quality controls are required in such a process as erroneous maintenance labels can lead to erroneous model behavior.

7. CONCLUSION

This study proposes an AI-based predictive-maintenance and Fault-Diagnosis approach for mechanical systems of any industrial domain by combining multimodal sensor data, signal processing, Ensemble learning and Temporal deep learning. The simulation of the controlled benchmark is built on 12,000 sensor windows and 6 operating states. The proposed hybrid model achieved an illustrative accuracy of 98.1%, macro F1 of 97.8% and ROC-AUC of 0.992% with false alarm rate of 1.6%. The main diagnostic variables used were RMS vibration and kurtosis, temperature and torque. For the illustrative horizon of four hours, the illustrative MAE was 1.02 hours.

The results indicate that a predictive-maintenance architecture could be built in such a way that diagnosis, prognosis, explainability, and maintenance decision support are not performed separately but are combined. The literature shows that the increasing need for transfer learning, uncertainty quantification, digital twins, and explainable AI, are becoming increasingly relevant in a growing desire to transfer predictive maintenance from benchmark experiments to real systems. The proposed framework is therefore designed to be able to accommodate domain adaptation, uncertainty-aware RUL estimates, feature-level explanations, and digital-twin integration.

The main drawback is the lack of an industrial dataset provided. Numerical results reported are controlled values and not evidence of performance at a named plant. Industrial sensor data captured at the time of use should be used to validate empirical data, along with failure events that have been verified and prospective maintenance outcomes. Further research is needed to evaluate cross-machine generalization, compound faults, sensor failure, concept drift, uncertainty calibration, maintenance cost optimization and the long-term field performance. The framework offers a structured approach to this validation and to the development of scalable maintenance systems, based on AI.

REFERENCES

1. Surucu, O., Gadsden, S.A. and 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>.
2. Das, O., Bagci Das, D. and Birant, D. (2023). Machine learning for fault analysis in rotating machinery: A comprehensive review. *Heliyon*, 9(6), e17584. <https://doi.org/10.1016/j.heliyon.2023.e17584>.
3. Bi, S., Wang, C., Wu, B., Hu, S., Huang, W., Ni, W., Gong, Y. and Wang, X. (2023). A comprehensive survey on applications of AI technologies to failure analysis of industrial systems. *Engineering Failure Analysis*, 148, 107172. <https://doi.org/10.1016/j.engfailanal.2023.107172>.
4. Gawde-Prabhudesai, S., Patil, S., Kumar, S., Kamat, P., Kotecha, K. and Abraham, A. (2023). Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, 123, 106139. <https://doi.org/10.1016/j.engappai.2023.106139>.
5. Taşcı, B., Omar, A. and Ayvaz, S. (2023). Remaining useful lifetime prediction for predictive maintenance in manufacturing. *Computers & Industrial Engineering*, 184, 109566. <https://doi.org/10.1016/j.cie.2023.109566>.
6. Zhuang, L., Xu, A. and Wang, X.-L. (2023). A prognostic driven predictive maintenance framework based on Bayesian deep learning. *Reliability Engineering & System Safety*, 234, 109181. <https://doi.org/10.1016/j.res.2023.109181>.
7. Yan, J., He, Z. and He, S. (2023). Multitask learning of health state assessment and remaining useful life prediction for sensor-equipped machines. *Reliability Engineering & System Safety*, 234, 109141. <https://doi.org/10.1016/j.res.2023.109141>.
8. Lin, Y.-H., Ding, Z.-Q. and Li, Y.-F. (2023). Similarity based remaining useful life prediction based on Gaussian Process with active learning. *Reliability Engineering & System Safety*, 238, 109461. <https://doi.org/10.1016/j.res.2023.109461>.
9. Hu, N. (2023). Remaining Useful Life Prediction of Milling Tool Based on Pyramid CNN. *Shock and Vibration*, 2023, 1830694. <https://doi.org/10.1155/2023/1830694>.
10. Fordal, J.M., Schjølberg, P., Helgetun, H., Skjermo, T.Ø., Wang, Y. and colleagues (2023). Application of sensor data based predictive maintenance and artificial neural networks to enable Industry 4.0. *Advances in Manufacturing*, 11, 248–263. <https://doi.org/10.1007/s40436-022-00433-x>.
11. Pashami, S., Nowaczyk, S., Fan, Y., Jakubowski, J., Paiva, N., Davari, N., Bobek, S., Jamshidi, S., Sarmadi, H., Alabdallah, A., Ribeiro, R.P., Veloso, B., Sayed-Mouchaweh, M., Rajaoarisoa, L., Nalepa, G.J. and Gama, J. (2023). Explainable Predictive Maintenance. *arXiv:2306.05120*.
12. Tang, S., Ma, J., Yan, Z., Zhu, Y. and Khoo, B.C. (2024). Deep transfer learning strategy in intelligent fault diagnosis of rotating machinery. *Engineering Applications of Artificial Intelligence*, 134, 108678. <https://doi.org/10.1016/j.engappai.2024.108678>.
13. Misbah, I., Lee, C.K.M. and Keung, K.L. (2024). Fault diagnosis in rotating machines based on transfer learning: Literature review. *Knowledge-Based Systems*, 283, 111158. <https://doi.org/10.1016/j.knosys.2023.111158>.
14. Matania, O., Dattner, I., Bortman, J., Kenett, R.S. and Parmet, Y. (2024). A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges. *Journal of Sound and Vibration*, 590, 118562. <https://doi.org/10.1016/j.jsv.2024.118562>.
15. Xiao, X., Li, C., Huang, J. and Yu, T. (2024). Rotating machinery fault diagnosis based on one-dimensional convolutional neural network and modified multi-scale graph convolutional network under limited labeled data. *Engineering Applications of Artificial Intelligence*, 137, 109129. <https://doi.org/10.1016/j.engappai.2024.109129>.
16. Pulcini, V. and Modoni, G. (2024). Machine learning-based digital twin of a conveyor belt for predictive maintenance. *International Journal of Advanced Manufacturing Technology*, 133, 6095–6110. <https://doi.org/10.1007/s00170-024-14097-3>.
17. Dereci, U. and Tuzkaya, G. (2024). An explainable artificial intelligence model for predictive maintenance and spare parts optimization. *Supply Chain Analytics*, 8, 100078. <https://doi.org/10.1016/j.sca.2024.100078>.
18. Wang, L., Zhu, Z. and Zhao, X. (2024). Dynamic predictive maintenance strategy for system remaining useful life prediction via deep learning ensemble method. *Reliability Engineering & System Safety*, 245, 110012. <https://doi.org/10.1016/j.res.2024.110012>.
19. de Beaulieu, M.H., Jha, M.S., Garnier, H. and Cerbah, F. (2024). Remaining Useful Life prediction based on physics-informed data augmentation. *Reliability Engineering & System Safety*, 252, 110451. <https://doi.org/10.1016/j.res.2024.110451>.
20. Gao, P., Wang, J., Shi, Z., Ming, W. and Chen, M. (2024). Long-term temporal attention neural network with adaptive stage division for remaining useful life prediction of rolling bearings. *Reliability Engineering & System Safety*, 251, 110218. <https://doi.org/10.1016/j.res.2024.110218>.
21. Mohammed, A., Noor, A., Abdulateef, O. and Hamad, A. (2023). An IoT and Machine Learning-Based Predictive Maintenance System for Electrical Motors. *Journal of Engineering and Applied Sciences*, 56(4), 651–656. <https://doi.org/10.18280/jesa.560414>.