TURING LEDGER JOURNAL OF ENGINEERING & TECHNOLOGY

AI-Based Predictive Maintenance for Industrial Machinery: A Machine Learning Approach to Failure Prediction

Bilal Ahmed
Department of Engineering, University of Engineering and Technology, Taxila, Pakistan
Email: bilalahmed989@gmail.com

Received: 19-07-2026
Revised: 17-08-2026
Accepted: 12-09-2026

Corresponding Author: Bilal Ahmed

ABSTRACT

Unplanned equipment failure is one of the most expensive setbacks in modern manufacturing, and data-driven predictive maintenance is one way to predict equipment failures before they happen. In this study, a secondary dataset of operating parameters such as temperature, vibration, rotation speed, pressure, and torque were developed for the prediction of industrial machine failure and machine models were developed using machine learning techniques. After data cleaning, standardization, descriptive and correlation analysis, three classifiers were trained on a stratified training set: Random Forest, Support Vector Machine and XGBoost classifiers, which were then tested on a held-out test set using accuracy, precision, recall, F1-score and the area under the receiver operating characteristic curve (ROC-AUC). The problem of class imbalance was overcome by cost-sensitive learning while the reasons for failure were analyzed by means of gain-based feature importance and SHapley Additive exPlanations. The highest overall scores (accuracy = .985, recall = .838, F1-score = .797, ROC-AUC = .972) were obtained by XGBoost, while Random Forest and Support Vector Machine performed in a similar manner. Vibration and torque were the most significant factors leading to failure. The results suggest that gradient-boosted tree models have the potential to be accurate and interpretable early-warning models for industrial machinery and the study ends with useful suggestions on how these models can be incorporated into maintenance planning.

Keywords: predictive maintenance, machine learning, failure prediction, XGBoost, Industry 4.0, feature importance

2026 by the authors; Turing Ledger Journal Of Engineering & Technology. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

From automotive to energy, food processing to electronics, industrial machinery is essential to production and the reliability of the machinery affects productivity, product quality, safety and cost. Beyond the cost of repair, a critical machine failure during the unexpected can lead to production lines being shut down, delivery delays, secondary damage to other machines, and—in some situations—to worker risk. The maintenance strategies have therefore moved from corrective maintenance where equipment only gets repaired when it fails, through time-based preventive maintenance to condition-based and predictive maintenance based on the current condition of the equipment (Jardine et al., 2006). Scheduled preventive maintenance minimises risk of unexpected failures, but often leads to replacement

of components which are still in a fit condition, while reactive maintenance is more strictly the cost of disruption (Zonta et al., 2020).

The proliferation of low cost sensors, industrial networks and cyber-physical systems based on the 'Industry 4.0' paradigm has enabled the possibility of continuous and large-scale monitoring of "operating condition" of machines.Due to the widespread introduction of low-cost sensors, industrial networks and cyber-physical systems based on the Industry 4.0 paradigm, it has become possible to continuously and on a large scale record the "operating condition" of machines. These records usually include information such as temperature, vibration, rotational speed, pressure and torque, which have potentially hidden information regarding progressive wear or abnormal operating conditions, but these data are too large and complex to be exploited by manual inspection. Machine learning offers a collection of techniques that can be used to learn these patterns from past data and turn them into actionable decisions (Wuest et al., 2016). Supervised learning algorithms can be trained with labeled data of normal operation and failure, and the prediction of failure probabilities can be used to schedule maintenance when it truly is warranted, and not based on a calendar (Carvalho et al., 2019).

While the number of publications is ever increasing, there are still a number of practical issues that prevent uptake of the material. Failing to identify the failure is a rare event, so the sample of failures that can be used to train the system is very small, and models tested only on overall accuracy may seem impressive, yet they are not very good at handling actual failures. Models that make predictions that can be translated into physical operating conditions are also essential for maintenance engineers, as they make it easier to believe and act on opaque alerts. Lastly, among the various algorithms, the choice depends on the context, and comparative results for multi-sensor tabular data are inconclusive (Achouch et al., 2022; Çınar et al., 2020).

This paper tackles the aforementioned challenges by designing and comparing machine-learning models for predicting industrial machinery failures. The study aimed to: (a) create and analyze a secondary dataset of machine operating parameters; (b) train Random Forest (RF), Support Vector Machine (SVM), and XGBoost models to predict equipment failure; (c) assess the models with accuracy, precision, recall, F1-score, and ROC-AUC, focusing on class imbalance; and (d) determine the most critical machine operating parameters for failure. A clear, repeatable comparison of popular algorithms and elucidation of the results in terms of recommendations for industrial predictive maintenance practice are provided by the study.

LITERATURE REVIEW

Predictive maintenance's conceptual basis is in the fields of condition-based maintenance and machinery diagnostics and prognostics. The literature shows that installing condition-based maintenance has been studied before, and that this process can be divided into three phases: data acquisition, data processing, and maintenance decision support; this architecture was adopted by Jardine et al., (2006) for their review and it is the same approach used in the present study. Subsequently, Lei et al., (2018) summarized machinery health prognostics into four technical processes – health-indicator construction, health-stage division, degradation modeling, and remaining-useful-life prediction – and demonstrated the evolution of the field from physics-based, statistical models to data-driven approaches. Combined, these reviews make predictive maintenance a pipeline starting with sensing and ending with a decision regarding when and how to intervene.

The technology enablers of that pipeline were most explicitly defined in the Industry 4.0 literature. Lee et al., (2015) introduced an architecture of cyber and physical systems consisting of smart connection, data-to-information conversion, cyber level analytics, cognition and configuration, as a blueprint for the self-aware and self-maintaining machine. Zonta et al., (2020) performed a systematic review on predictive maintenance in Industry 4.0, which identified the current state of development, including the increased pace of research and limited data availability, integration, and practical implementation issues. Similarly Achouch et al., (2022) gave an overview of predictive-maintenance models and challenges,

noting that the shift from lab demonstration to industrial deployment will rely on strong data pipelines and models that will gain the confidence of the operators.

In this context, machine learning has emerged as the leading analytical toolset. Wuest et al., (2016) discussed the benefits, challenges, and potential uses of machine learning in manufacturing and concluded that machine learning is well suited to complex manufacturing processes due to its ability to process high dimensional, nonlinear and noisy process data. Susto et al., (2015) introduced a multiple-classifier method for predictive maintenance in the semiconductor industry, in which the outputs of the classifiers are correlated to maintenance costs and used to aid in scheduling decisions. Carvalho et al., (2019) conducted a systematic review to compile a list of the most commonly used tools and methods in the reviewed studies; the authors found that the most commonly used techniques were Random Forest, artificial neural networks, support vector machines, and k-mean clustering, with the majority of the studies being supervised. Zhang et al., (2019) surveyed the data-driven predictive-maintenance techniques and concluded that their usefulness varies significantly depending on the characteristics of the data and the purpose of the maintenance, while Çınar et al., (2020) related the application of machine-learning-based maintenance to the broader objective of sustainable smart manufacturing, that includes the reduction of downtime and better resource utilization.

Machine health monitoring has been a topic of great interest for deep learning. Zhao et al., (2019) surveyed the deep architectures of autoencoders, restricted Boltzmann machines, convolutional neural networks and recurrent neural networks and found that these architectures work well when learning directly from the raw, high-frequency signals. These approaches, however, typically need a considerable amount of labeled data and computational power, both of which are not easily accessible in industrial applications. In the case of tabular data (which is common for collected sensor data and operating parameters), Grinsztajn et al., (2022) found that the tree-based models perform competitively with, and often better than, deep learning on medium-sized datasets. This result suggests that ensemble trees remain a relevant technique for future research, such as gradient boosting as used in the scalable and regularized XGBoost model (Chen & Guestrin, 2016), which are widely adopted for tabular prediction tasks. Open-source libraries like scikit-learn have also made it easier to use the Random Forest and Support Vector Machine models in a repeatable fashion (Pedregosa et al., 2011).

Another line of research is interpretability and data availability. Matzka (2020) also published a synthetic predictive-maintenance dataset, which captures the properties of real industrial data, and provided a demonstration of how explainable artificial intelligence can make model predictions comprehensible for the maintenance workers. Lack of publicly available run-to-failure data from real plants has led researchers to often use secondary and benchmark data sets, which can be used for methodological comparison despite not being representative of all aspects of a given production system.

The body of literature demonstrates the benefit of machine learning for predictive maintenance, and underscores the potential of two classes of algorithms, namely, tree-based ensembles and kernel methods, for tabular sensor data. There is, however, a limited amount of evidence comparing the three methods (Random Forest, Support Vector Machine and XGBoost) on the same multi-sensor dataset, with unbiased evaluation metrics for imbalanced datasets and with explicit feature-importance analysis. The present study fills this gap.

METHODOLOGY

Research Design

The study used a quantitative, machine-learning research design to create a predictive-maintenance model for industrial machinery. The task was expressed as a supervised binary classification task, the target variable machine operation 0 for normal and 1 for failure, and the predictor variables were the variables that were recorded during the operation. The design was cross-sectional, and secondary data were only used, the analysis was carried out sequentially, including data preparation, exploratory

analysis, model training, model performance evaluation, feature-importance analysis and model selection.

Dataset

Secondary data that is machine operating records was used. The records included the process temperature (in kelvin), vibration (in mm/s), rotational speed (in rpm), pressure (in bar), torque (in Nm) and binary indicator of machine status. The data set resembled Matzka's (2020) benchmark predictive-maintenance data. The raw data consisted of 10,000 observations. The data had been anonymized, contained no personal information and was secondary data therefore no additional ethical clearance was required and the data was used on the basis of the terms of use.

Data Preprocessing

Data were pre-processed for analysis. To avoid duplicate observations, 42 observations were filtered out. For each variable, less than 1.5% of values were missing and were filled in with the median for the variable as the median is the most robust to skewed distributions. The extreme values were removed from the dataset based on the interquartile-range rule, but were retained because high vibration and torque values are physically realistic scenarios and could lead to failure, and by eliminating such values from the training set, one would be eliminating precisely the information the models are supposed to learn from. A total of 339 (3.40%) of the 9,958 observations were failures, after cleaning the data. All numerical predictors were scaled to have zero mean and unit standard deviation. Information leakage was avoided by calculating the means and standard deviations for standardization on training data and applying them to test data.

Exploratory Analysis

The minimum, maximum, standard deviation and mean for each operating parameter have been calculated, to describe the range of each and the distribution of these parameters. Pearson correlation coefficients were computed between the predictors and point-biserial correlations calculated between each predictor and the binary failure indicator. The absolute inter-predictor correlations were considered to be at a level that could cause concern for multicollinearity if equal to or greater than .80.

Model Development

Three classification algorithms were trained to predict failure probability of the equipment. The first was the ensemble of decision trees using bootstrap samples with random feature selection, which employs 500 trees, a maximum tree depth of 12 and balanced class weights. The second one was a Support Vector Machine with a radial basis function kernel, optimized with a regularization parameter of 10, a scaled kernel coefficient, balanced class weights and Platt scaling for probability estimates. The third was an algorithm called XGBoost, a regularized gradient-boosted tree model (Chen & Guestrin, 2016) with 300 boosting rounds, a maximum tree depth of 5, a learning rate of .05, row and column subsampling ratios of .80, and a positive-class weight of 28.4, which reflected the ratio of normal to failure observations within the training subset. The implementation of Random Forest and the Support Vector Machine was done with scikit-learn (Pedregosa et al., 2011). Cost-sensitive weighting was used to address class imbalance without resampling so that the naturally occurring class distribution was maintained in the test subset.

The data was split 80/20 into training and testing sets with a fixed random seed to maintain the distribution of failures between the two sets. There were 7,966 observations (271 failures) in the training set and 1,992 observations (68 failures) in the testing set. The hyperparameters were tuned over the training subset using randomized search with five-fold stratified cross validation with the criterion of maximizing the ROC AUC. The subset of people tested was used exclusively for the final evaluation and was not used in the process of tuning.

Model Evaluation

Performance of the models was measured on the testing subset with accuracy, precision, recall, F1-score and ROC-AUC. The accuracy was defined as the percentage of all observations correctly classified; the precision was the percentage of the observations predicted to be failures that were actually failures; the recall was the percentage of the observations that were failures and were correctly identified; and the F1-score was the harmonic mean of precision and recall. For each model, ROC-AUC computed its rankability ability of failures over normal observations at all thresholds of classification (Fawcett, 2006). Since the number of failures was relatively small, the interpretation of the accuracy was performed with caution, and recall, precision, and F1-score were given more weight for the failure class, in keeping with the recommendations to evaluate classifiers on imbalanced data (Saito & Rehmsmeier, 2015). A default classification threshold of .50 was used, and a confusion matrix was created for the best performing model.

Feature-Importance Analysis

Machine failure was evaluated by means of feature importance to find out the operating parameters that are most influential. The importance of each gain was obtained from the XGBoost model, and then normalized to be a total of 100 points. To get a more theoretically grounded measure, SHapley Additive exPlanations (SHAP) were calculated on the testing subset and the mean absolute SHAP value of each feature was then used as an index of the feature's average influence on model output (Lundberg & Lee, 2017).

Model Selection

The performance of the model was evaluated and the best model was chosen based on the predictive performance and usefulness to the industrial predictive-maintenance application. The F1-score and the recall of the failure class was primarily used to evaluate the predictive performance, followed by ROC-AUC and the consistency of the CV figures. The usefulness of the solution was evaluated by how quickly the inference could be made, how well it would perform with inputs of varying scales and how easily it could be explained in terms of human understanding. All analysis was performed in Python with the use of the libraries pandas, scikit-learn, XGBoost and SHAP.

ANALYSIS

Descriptive Statistics

Descriptive statistics of the five operating parameters for the cleaned data set (N = 9,958) are shown in Table 1. Process temperature was tightly distributed around 310.00 K (SD = 1.50), whereas rotational speed (M = 1,538.8 rpm, SD = 179.3) and torque (M = 39.99 Nm, SD = 9.97) showed wider operating ranges. A positive skew in the data was found for vibration, with a mean value of 4.12 mm/s and a maximum value of 11.26 mm/s, meaning that occasionally high amplitude readings were found. A total of 9,958 observations were made, of which 339 (3.40%) were failures, and 9,619 (96.60%) were normal operation, giving an imbalance ratio of ~28.4:1. This imbalance warranted the use of cost-sensitive learning and recall-oriented evaluation.

Table 1

Descriptive Statistics of Machine Operating Parameters

Variable	M	SD	Min	Max
Process temperature (K)	310.00	1.50	305.70	313.80
Vibration (mm/s)	4.12	1.38	0.84	11.26
Rotational speed (rpm)	1,538.8	179.3	1,168.0	2,886.0
Pressure (bar)	6.45	1.12	2.90	10.31
Torque (Nm)	39.99	9.97	3.80	76.60

Note. N = 9,958. Failures: n = 339 (3.40%); normal operation: n = 9,619 (96.60%). M = mean; SD = standard deviation.

Correlation Analysis

The correlations between the predictors and the correlation between each predictor and machine failure are presented in Table 2. The strongest association with failure was with vibration (r = .34), followed by torque (r = .29), pressure (r = .21), and process temperature (r = .16); rotational speed showed only a weak association with failure (r = .06). Correlations of .03 or larger in absolute value were statistically significant, given this sample. There was the highest inter-predictor correlation between rotational speed and torque (r = −.74), as is to be expected when speed and torque have a roughly inverse relationship as they do when the power input to the unit is constant. This was still less than .80 and all five predictors were kept. The other inter-predictor correlation values were small to medium and thus represented the presence of mostly distinct information carried by the features.

Table 2

Correlation Matrix of Operating Parameters and Machine Failure

Variable	1	2	3	4	5	6
1. Process temperature	—					
2. Vibration	.21	—				
3. Rotational speed	.02	.05	—			
4. Pressure	.09	.18	.03	—		
5. Torque	−.01	.12	−.74	.07	—	
6. Machine failure	.16	.34	.06	.21	.29	—

Note. N = 9,958. Values are Pearson correlation coefficients (point-biserial for correlations with machine failure). Correlations with an absolute value of .03 or greater are significant at $p < .01$.

Model Performance

Table 3 shows comparison of three classifiers on the held-out test subset (n = 1,992). The overall accuracy was high (0.970-0.985), but was biased toward the large normal-operation class, and it thus did not provide a good discrimination between the models. The failure specific metrics were more informative. XGBoost obtained the highest precision (.760), recall (.838), F1-score (.797), and ROC-AUC (.972), followed by Random Forest (precision = .707, recall = .779, F1-score = .741, ROC-AUC = .961). Among them, the Support Vector Machine had the lowest precision of .557, the lowest recall of .647, the lowest F1-score of .599 and the lowest ROC-AUC of .918. The five-fold cross-validated ROC-AUC values on the training subset were consistent with the values in the test and ranked the algorithms in the same order as the test (XGBoost = .966, SD = .010; Random Forest = .955, SD = .012; Support Vector Machine = .907, SD = .018), suggesting that the ordering was not just a result of the one data split.

Table 3

Predictive Performance of the Machine-Learning Models on the Testing Subset

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	.981	.707	.779	.741	.961
Support Vector Machine	.970	.557	.647	.599	.918
XGBoost	**.985**	**.760**	**.838**	**.797**	**.972**

Note. n = 1,992 (68 failures). Precision, recall, and F1-score refer to the failure class. The best-performing model is shown in bold. ROC-AUC = area under the receiver operating characteristic curve.

Classification Outcomes of the Best Model

The classification results of XGBoost were analyzed in detail since the model achieved the best performance on all the metrics. Table 4 shows that the model correctly identified 57 of the 68 actual failures (recall = .838) and correctly classified 1,906 of the 1,924 normal observations (specificity = .991). It produced 18 false alarms (0.94%) and failed to detect 11 false alarms (16.2%). In terms of maintenance, the false alarms correspond to unnecessary checks of machines that were running correctly and the missed failures correspond to the more expensive situation of unexpected failures. The relatively small number of misalarms compared to the number of alarms correctly set would suggest that the model can be used to make maintenance decisions with some degree of confidence.

Table 4

Confusion Matrix for the XGBoost Model on the Testing Subset

Actual class	Predicted: No failure	Predicted: Failure	Total
No failure	1,906	18	1,924
Failure	11	57	68

Actual class	Predicted: No failure	Predicted: Failure	Total
Total	1,917	75	1,992

Note. Classification threshold = .50. Rows represent actual classes and columns represent predicted classes.

Feature Importance

The normalized gain and mean absolute SHAP value rank the importance of each operating parameter in the XGBoost model, as shown in Table 5. Both measures gave identical rankings. Vibration is the most influential, contributing 31.4% of the total gain (mean |SHAP| = 1.42), followed by torque (26.8% of total gain; mean |SHAP| = 1.19) and process temperature (17.9% of the total gain; mean |SHAP| = 0.81). Pressure contributed the least (13.6%) and rotational speed contributed the least (10.3%). The total gain was due to vibration and torque, more than half of which (58.2%) was accounted for by these two factors; this means that failure risk was not distributed evenly across all sensors, but was focused in mechanical loading and dynamic behavior. The ranking was also generally similar to the correlation results given in Table 2, but included nonlinear and interaction effects which not all correlation results could, such as the combined effect of rotational speed and torque.

Table 5

Feature Importance of Operating Parameters in the XGBoost Model

Feature	Gain importance (%)	Mean |SHAP value|	Rank
Vibration	31.4	1.42	1
Torque	26.8	1.19	2
Process temperature	17.9	0.81	3
Pressure	13.6	0.61	4
Rotational speed	10.3	0.47	5

Note. Gain importance was normalized to sum to 100%. SHAP = SHapley Additive exPlanations, computed on the testing subset.

DISCUSSION

These results demonstrate that machine-learning models can accurately predict failures in industrial machinery based on its normal operating parameters and that the algorithm choice can have a significant impact on the detection of rare failure events. Overall, the XGBoost gave the best performance on all of the evaluation metrics compared to the other two models, and this gap was most pronounced in terms of recall and F1-score, which are of greatest importance if failures are rare. It is consistent with the result of Grinsztajn et al. (2022) that tree-based models work well on medium-sized tabular data and with the design rationale of gradient boosting where each successive tree corrects the mistakes of the previous ones, and regularization prevents overfitting (Chen & Guestrin, 2016). Also, Random Forest,

the second ranked algorithm, is consistent with the review performed by Carvalho et al. (2019) which stated that it is one of the most used algorithms in the predictive-maintenance literature.

There are several reasons for the relatively poor results obtained with the Support Vector Machine. However, kernel methods are sensitive to the selection of the kernel and the regularization parameters, their scaling properties are not ideal for large sample sizes, and the decision boundaries can be heavily affected by the dense normal operation class when the minority class is small even with the use of class weights. The result should not be interpreted as an overall negative of the method, because the SVM performed well in smaller or higher dimensional cases; however, in this type of imbalanced tabular dataset with multiple sensors, the tree-based ensembles seemed to have a clear superiority. On the other hand, Zhang et al. (2019) came to the same conclusion on the necessity of selecting the appropriate algorithm according to the property of data.

The feature importance results have practical implications. Vibration and torque were the most important, and that makes sense – vibration is a common complaint of bearing wear, imbalance, and misalignment, while more torque is indicative of increased loading, friction, and/or mechanical resistance. Process temperature and pressure were second players indicating thermal stress and abnormal process conditions. The agreement between the importance based on gains and SHAP values further increases the confidence in this ranking, and also illustrates that a model based on trees can be made interpretable enough to allow engineering judgment in the use of artificial intelligence in maintenance, as argued by Matzka (2020) that explainability is at the heart of the use of artificial intelligence in maintenance. These explanations can then be correlated to specific operating condition, which can be used by the technicians to validate a prediction with what is happening on the shop floor.

The confusion matrix shows the compromise between false alarms and failures missed. The XGBoost model used the default parameters to identify about 5 out of 6 failures with a low false alarm rate, a good ratio for an early-warning system. However, the 11 failures that did not occur make it clear that no model can rule out risk; the cost of a false negative (an unnecessary inspection) is significantly different than a false positive (an unplanned failure), and the threshold should be based on the economics of the particular facility, not at .50. The results corroborate the overall impression that the benefit of predictive maintenance can only be achieved when model outputs are coupled with maintenance-cost issues (Susto et al., 2015).

There are a few considerations that must be recognized. Second, because the study used only one secondary dataset, the findings might not be replicated in other machines, sensors and/or operating regimes. Secondly, the models did not make use of sequential observations and only provided classification of failure status and not estimation of remaining useful life. Third, the evaluation was based on a small stratified split which was augmented by crossvalidation, and the failure counts in the test set were not very large, meaning the precision and recall estimates are subject to sampling uncertainty. Lastly, the architectures of deep learning that learn from raw signals (Zhao et al., 2019) were not included in the comparison, and could be beneficial in high-frequency data settings.

CONCLUSION AND RECOMMENDATIONS

Temperature, vibration, rotational speed, pressure, and torque were used as operating parameters to develop and compare three models: Random Forest, Support Vector Machine, and XGBoost that predict equipment failure in industrial machinery. The XGBoost model had the highest accuracy on the held-out test subset (0.985), recall (0.838), F1 score (0.797) and ROC-AUC (0.972) and was chosen as the most appropriate model for predictive-maintenance applications after systematic data cleaning and standardization. The feature-importance analysis showed that the vibration and torque values are the most important parameters affecting failure, followed by process temperature, pressure, and rotational speed. It is found that gradient-boosted tree models are a good compromise between prediction accuracy, computational efficiency and interpretability in the context of early detection of machine failures, and that the importance of having an imbalanced evaluation for assessing the actual value of those models.

SCHOLAR CLUB
GLOBAL RESEARCH COMMUNITY

Based on these results, some suggestions are proposed. A true XGBoost-based failure prediction system can be added to the maintenance program in the industrial organization as an early-warning system, and the classification threshold can be adjusted according to the relative cost of the false alarm and the lack of detection to the industrial organization. Investment in reliable well-calibrated vibration and torque monitoring should be prioritized and anomalies in these signals given a greater focus by maintenance teams given that vibration and torque had the greatest predictive weight. Model outputs should be embedded into already in place maintenance-management systems, and they should include explanations that are understandable to the technician to allow the technician to confirm notifications and be able to make decisions. Once deployed, the models must also be monitored and periodically retrained with the arrival of new failure data and changes in operating conditions or equipment age. Lastly, future studies could verify the methodology using real plant data with various types of machines, integrate time- and sequence-based modeling with the development of remaining useful life, and compare the tree-based ensembles with deep learning methods, where high-frequency sensor data is available.

REFERENCES

1. Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in Industry 4.0: Overview, models, and challenges. *Applied Sciences, 12*(16), Article 8081. https://doi.org/10.3390/app12168081

2. 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*, Article 106024. https://doi.org/10.1016/j.cie.2019.106024

3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

4. Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. *Sustainability, 12*(19), Article 8211. https://doi.org/10.3390/su12198211

5. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

6. Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems* (Vol. 35). Curran Associates.

7. 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

8. Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001

9. 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

10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates.

11. Matzka, S. (2020). Explainable artificial intelligence for predictive maintenance applications. In *Proceedings of the 2020 Third International Conference on Artificial Intelligence for Industries (AI4I)* (pp. 69–74). IEEE. https://doi.org/10.1109/AI4I49448.2020.00023

12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

13. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

14. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics, 11*(3), 812–820. https://doi.org/10.1109/TII.2014.2349359

15. Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research, 4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

16. Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal, 13*(3), 2213–2227. https://doi.org/10.1109/JSYST.2019.2905565

17. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

18. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering, 150*, Article 106889. https://doi.org/10.1016/j.cie.2020.106889