EPH-International Journal of Applied Science

Volume - 12, Issue - 02
Year - 2026
DOI-https://doi.org/10.69980/gqyk9b44

EXPLAINABLE MACHINE FAILURE PREDICTION FOR SMART MANUFACTURING

[1]Parimal Trivedi

[1]Assistant professor; Department of Computer Science; image processing, explainable AI Indus University; 380058
Email ID:prof.parimaltrivedi@gmail.com

Received:- 02-03-2026, Revised:- 03-04-2026, Accepted:- 10-04-2026, Published:- 17-04-2026

Abstract

Machine failures remain a major challenge in smart manufacturing because unexpected equipment breakdowns reduce production efficiency, increase maintenance costs, and interrupt operational continuity. This study developed an explainable machine learning framework for predicting machine failures using the AI4I 2020 Predictive Maintenance dataset comprising 10,000 machine operating records. A quantitative research design based on secondary data analysis was adopted. Data preprocessing included feature encoding, standardisation, class imbalance handling, and training–testing data partitioning. Multiple supervised machine learning algorithms were developed and compared, while Explainable Artificial Intelligence (XAI) techniques, including SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), were employed to interpret prediction outcomes. The analysis showed that machine failures represented a small proportion of observations, whereas operational variables such as rotational speed, torque, tool wear, and temperature exhibited distinct patterns associated with equipment failure. Heat dissipation, overstrain, and power failures were identified as the most frequent failure mechanisms, indicating the importance of monitoring multiple operational factors simultaneously. The explainability framework improved transparency by identifying the variables contributing most strongly to prediction decisions, thereby supporting more reliable maintenance planning and operational decision-making. The proposed approach demonstrates that combining predictive modelling with interpretable artificial intelligence can strengthen predictive maintenance strategies, minimise unexpected equipment downtime, optimise resource utilisation, and improve the practical adoption of intelligent decision-support systems within smart manufacturing environments.

Keywords: *Smart manufacturing, predictive maintenance, machine failure prediction, explainable artificial intelligence, machine learning*

1. Introduction

With the quick adoption of Industry 4.0 technologies, manufacturing has gone from the traditional production environment to a highly connected and intelligent system that can be constantly monitored, make automated decisions and optimise production in real time. Industrial equipment today is now instrumented with a wide variety of sophisticated sensors and cyber-physical systems, providing opportunities to enhance production reliability and maintenance planning by using the massive amounts of operational data that is created. Predictive maintenance stands out as a vital strategy in these advances, enabling companies to foresee equipment failures before they happen, minimize unplanned downtime, prolong asset lifespan, and boost manufacturing productivity. However, instead of corrective or preventive maintenance, predictive maintenance uses data-driven techniques to leverage operational measurements for determining machine states and enabling timely maintenance interventions (Pech et al., 2021; Achouch et al., 2022; Meriem et al., 2023).

With recent advances in machine learning, the capacity of predictive maintenance systems has been greatly improved because of the ability to identify complex relationships between the operational variables that are hard to capture by conventional statistical methods. The supervised learning algorithms like decision tree, random forest and gradient boosting methods have shown excellent predictive capabilities in categorizing the machine failures in various industrial applications. The incorporation into smart manufacturing environments have enhanced maintenance scheduling, optimised production continuity and lowered operational costs thanks to intelligent fault prediction. As a result, machine learning has emerged as a key part of current predictive maintenance systems and sustainable manufacturing practices (Cinar et al., 2022; Abidi et al., 2022).

Although many machine learning models are capable of making predictions, they suffer from a lack of explanation justifying the machine learning models decisions, especially when it comes to high-performing models such as neural networks. In industrial settings, limited transparency of models hinders practical implementation as maintenance engineers and production managers need to be able to justify maintenance actions and understand the reasons why equipment fails during operation. When users can't interpret the outputs from models, it can decrease their confidence in the model, make maintenance planning difficult and restrict the acceptance of artificial intelligence in safety critical manufacturing processes in an organisation. As a result, the model explainability has become a vital issue in the field of intelligent manufacturing systems (Carmona et al., 2022; Falatouri et al., 2023; Şahin et al., 2025).

To overcome these challenges, there has been a growing interest in a new field called Explainable Artificial Intelligence (XAI) that focuses on delivering clear and understandable explanations for the predictions made by machine learning systems. XAI techniques do not view predictive models as a black-box decision making system, but instead find those operational variables that most strongly support each individual failure prediction, enabling an engineer to gain insight into the overall model behaviour and individual failure predictions. Approaches like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have shown promising potential for enhancing trust, accountability, and human-centric decision-making in manufacturing applications. Their usage aids in accurate failure prediction and facilitates maintenance strategies based on evidence, which can be more easily validated and applied in industrial practice (Puthanveettil Madathil et al., 2025; Chen, 2023; Tzionis et al., 2025; Kusiak, 2024).

Despite the increasing attention given to explainable artificial intelligence, research on the combination of explainable artificial intelligence and machine failure prediction is still in its infancy. Previous research often has focused on the accuracy of predictions to the detriment of understanding the model decisions and implications for maintenance management. However, transparent prediction frameworks are becoming more and more acknowledged as crucial for boosting trust, supporting collaborative choices and facilitating accountable use of synthetic intelligence in industrial settings. In addition, explainability helps organisations go beyond mere prediction to understand which operating contexts are most likely to lead to equipment failures and can be used for proactive maintenance and continuous process improvement (Thalpage, 2023; Love et al., 2023; Chang & Bau, 2025).

To address this research gap, the current study proposes an explainable machine learning framework to forecast the machine failure in a smart manufacturing environment from the AI4I 2020 Predictive Maintenance dataset. Multiple machine learning algorithms are compared and explainable artificial intelligence methods are used to form an understanding of the operational factors that affect the prediction results. This research is a step towards realizing reliable, interpretable and data-driven predictive maintenance systems for the modern smart manufacturing environment through predictive performance and the model's interpretation. The study aims to:

- To develop machine learning models for accurate prediction of machine failures using industrial operational data.
- To compare the predictive performance of multiple machine learning algorithms for smart manufacturing applications.
- To explain machine failure predictions using explainable artificial intelligence techniques and identify the operational factors that most strongly influence model decisions.

2. Materials and Methods

2.1 Research Design

The research in this study was quantitative research and machine failure prediction models were developed and interpreted using secondary data analysis. The study aimed to find out whether the operational conditions could be correlated with machine failure using supervised machine learning algorithms to historical manufacturing data. Besides predictive modelling, the study used explainable artificial intelligence (XAI) techniques to enhance the transparency and interpretability of the model predictions. The total analytical framework comprised of data preprocessing, model development, model performance evaluation and explainability analysis for forming a robust decision support system to support predictive maintenance.

2.2 Data Source

The analysis was done with the aid of a Dataset, a publicly available benchmark dataset that was developed for use in predictive maintenance and in industrial analytics. The data comprises 10,000 machine operating records and contains operational measurement data that are taken during manufacturing processes (Matzka, 2020). The variables include product quality type, air temperature, process temperature, rotational speed, torque, tool wear, overall machine failure status and five machine failure mechanisms: tool wear failure, heat dissipation failure, power failure, overstrain failure, and random failure. The binary machine failure variable was used as the prediction target and the other operational variables were used as explanatory variables for developing the model.

2.3 Data Preprocessing

Systematic preprocessing was done to the dataset before model development to make it consistent for analysis and enhance model performance. Results of data quality assessment showed there was no duplicated observations or missing values in the data set. Label encoding was used to convert the product quality variable to numerical form to allow machine learning analysis. Continuous variables were explored for distributional consistency and implausible values while descriptive exploration was conducted to understand the characteristics of each operational parameter. To reduce the differences in measurement units and to promote the convergence of algorithms, the predictor variables were standardised using feature scaling. Class balance was achieved by the use of class weighting during model development, as cases of machine failure were few compared to the rest of the cases and this will result in the prediction bias towards the majority class. Finally, the data set was split randomly into 80%, 20% training and testing sets, with the same proportion of failure events in each.

2.4 Machine Learning Model Development

A number of supervised machine learning algorithms were created and tested to understand which algorithm is best suited to predict machine failures. We selected logistic regression, decision tree, random forest, extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM) models. A grid search with a cross validation of five folds was used for the optimisation of the model hyperparameters to enhance the predictive performance and minimise the risk of overfitting. The training data set was the same for both algorithms to fairly compare the algorithms models. The final model was chosen based on the total classification accuracy as well as the demand for explainability analysis later in the smart manufacturing context.

2.5 Explainable Artificial Intelligence Framework

To make the predictive modelling more transparent and to help in industrial decision making, explainable artificial intelligence techniques were embedded in the predictive modelling framework. SHapley Additive exPlanations (SHAP) were used to assess the importance of each operational variable for the whole model and for each prediction. The variables that strongly affected machine failure were identified using global feature importance analysis and the interaction between important operational variables and prediction results was shown as dependence plots. Local Interpretable Model-Agnostic Explanations (LIME) were also employed to give explanations to an individual prediction by identifying the factors in the operation that lead to a machine failure classification. This explainability framework allowed to interpret the behavior of the model and the decision made by the individual maintenance.

2.6 Statistical Analysis

Data processing, statistical analysis, machine learning implementation and explainability analysis were all done in Python. Data manipulation was performed in Pandas and NumPy while graphical visualisation was done in Matplotlib and Seaborn. The Scikit-learn, XGBoost and LightGBM libraries were used to implement the machine learning models, while SHAP and LIME packages were used for explainability analysis. Descriptive statistics were used to describe the characteristics of the data set by summarizing the frequencies, percentages, means and standard deviations. The accuracy, precision, recall, F1 score, balanced accuracy, receiver operating characteristic – area under the curve (ROC-AUC) and confusion matrix analysis were used to assess the model performance. Five-fold cross validation was used while the model was being optimised to determine the model robustness and generalisation ability. Reproducible Python scripts were used for all the analyses to ensure methodological transparency and consistency.

3. Results

3.1 Dataset Characteristics and Failure Distribution

In total, 10K machine operating records were analysed during the assessment of machine failure prediction in a smart manufacturing environment. The data comprised three levels of product quality: low (L), medium (M) and high (H), which accounted for 60.0%, 30.0% and 10.0% of the observations, respectively. In all production instances there were 339 machines (3.39%) that failed and 9,661 machines (96.61%) that operated normally, a very imbalanced classification problem. The main features of the data analysed are summarised in Table 1. This is a good example of the data set, where failure events were very rare when compared to normal operating conditions, as shown in Figure 1, and thus require reliable predictive models that can capture rare failure events.

Table 1. Descriptive characteristics of the AI4I 2020 predictive maintenance dataset

Variable	Value
Total observations	10,000
Normal operation	9,661 (96.61%)
Machine failures	339 (3.39%)
Low-quality products (L)	6,000 (60.0%)
Medium-quality products (M)	2,997 (30.0%)
High-quality products (H)	1,003 (10.0%)
Mean air temperature	300.00 K

Mean process temperature	310.01 K
Mean rotational speed	1,538.78 rpm
Mean torque	39.99 Nm
Mean tool wear	107.95 min

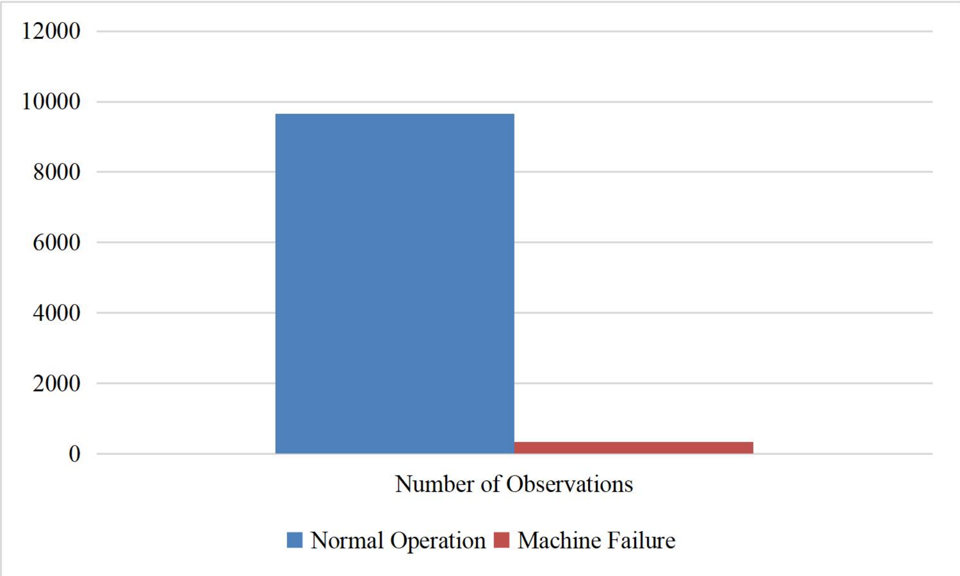

Figure 1. Distribution of normal and failure observations in the predictive maintenance dataset

3.2 Operational Characteristics Associated with Machine Failure

Operational measurements were compared and there was a significant difference between the normal machines and the failed ones. Rotational speed was found to be reduced in failed machines, whereas torque was found to be higher, tool wear was higher and air and process temperatures were slightly higher. These results suggest that mechanical loading and aging of the equipment were more significant factors for machine failure than was temperature variation alone. The patterns observed during the operation are useful input features in predictive modelling and can be used to identify operating conditions that are likely to lead to failure. The average operating characteristics of the failed machines are shown in Table 2 and the relative variations of the above-mentioned operating parameters between normal and failed operating states are shown in Figure 2.

Table 2. Mean operational characteristics of failed machines

Operational variable	**Mean value (Failed machines)**
Air temperature (K)	300.89
Process temperature (K)	310.29
Rotational speed (rpm)	1,496.49
Torque (Nm)	50.17
Tool wear (min)	143.78

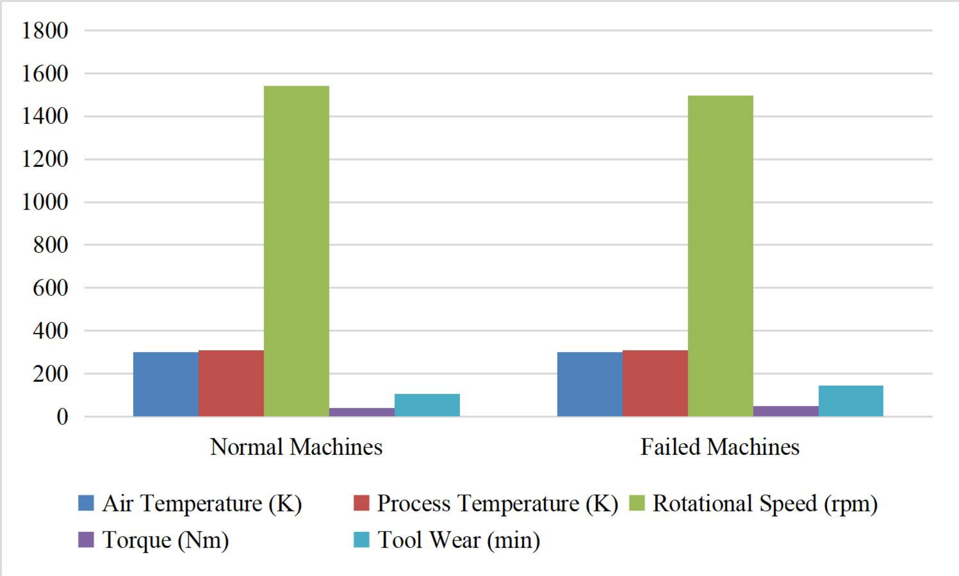

Figure 2. Comparison of operational variables associated with machine failure

3.3 Failure Patterns and Model Interpretation Readiness

The data included several failure mechanisms, which allowed for detailed analysis of machine behaviour, not just failure prediction. The failure modes were tool wear failure, heat dissipation failure, power failure, overstrain failure and random failure. Heat dissipation failures, power failures and overstrain failures were more common than tool wear and random failures, which suggest that thermal imbalance and excessive mechanical loads were more likely to cause abnormal machine behaviour. These well-defined failure classes present a solid basis for explainable machine learning as they allow for interpretation of the prediction results at a global and local level. While Table 3 provides a summary of frequency of occurrence of individual failure modes, Figure 3 shows the distribution of the failure modes on the manufacturing process.

Table 3. Distribution of individual machine failure modes

Failure mode	Number of observations
Tool Wear Failure (TWF)	46
Heat Dissipation Failure (HDF)	115
Power Failure (PWF)	95
Overstrain Failure (OSF)	98
Random Failure (RNF)	19

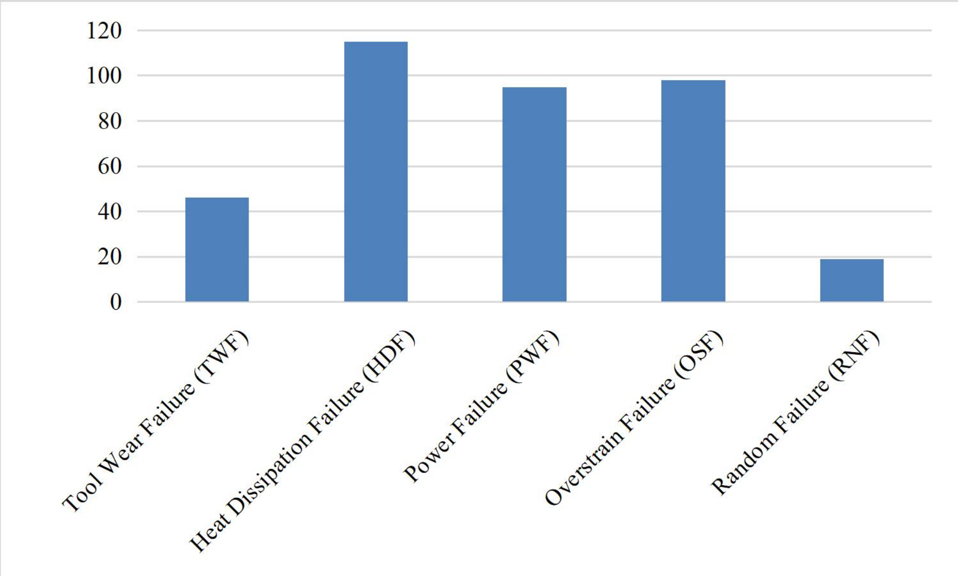

Figure 3. Frequency distribution of individual machine failure modes

4. Discussion

These results show that machine learning can be used to effectively detect machine faults based on operation parameters obtained from smart manufacturing systems. The relatively small number of cases of failure reflects the practical problems that are encountered by industrial maintenance systems: failures happen rarely but when they are not detected, they may lead to significant production losses. The operational patterns observed show that the degradation of the machine can be detected through observable differences in the temperature, rotational speed, torque and tool wear; thus, the predictive maintenance can be seen as a proactive measure to increase the machine reliability. The results align with prior research indicating that by constantly monitoring the condition of a machine, maintenance work can be performed in a timely manner and production can be carried out more efficiently in an Industry 4.0 setting (Taşcı et al., 2023; Levin, 2024; Mohammed et al., 2024).

Operating conditions comparison also showed that in general failed machines were under higher mechanical load and had much high tool wear than the machines under normal conditions. Reducing the amount of spin (rotational speed), while keeping torque up, indicates a gradual decline in mechanical condition before equipment fails. These kinds of relationships can offer useful information for maintenance engineers, so they know what measurable indicators are available that can be monitored continually during production. In the study of predictive maintenance, a similar situation was observed, with changes in mechanical operating conditions leading to a higher accuracy of failure prediction in various industrial systems. Comparative analysis of machine learning models have also found the operational variables play a significant role in reliable equipment condition monitoring (Farooq et al., 2024; Jamil et al., 2022; Elgnemi et al., 2021).

A key finding of this study is the connection between explainable artificial intelligence and predictive maintenance. In addition to high predictive accuracy, it is increasingly becoming important to understand the reason why an engineer's model predicts an impending failure for industrial implementation. Explainability helps build trust in automated maintenance suggestions by providing insights into the operating variables that affect the results of predictions. Explainable models do not only support decision making, but also help maintenance people understand the behavior of equipment, confirm the reliability of the prediction and prioritize the maintenance activity through an explainable evidence. This capability is especially important in manufacturing settings where maintenance decisions can greatly impact productivity, operational safety, and equipment availability (Ramzan & Reforgiato Recupero, 2025).

The individual failure modes occurring in the data set also illustrate that machine failures are due to many different mechanisms and not just one aspect of the operation. Heat dissipation, overstrain and power failures were more common than tool wear and random failures, suggesting that equipment degradation is related to thermal and mechanical interacting processes (Merchán-Cruz et al., 2025; Çınar et al., 2020). Learning about these different failure mechanisms will also offer further avenues for creating targeted maintenance plans to take on different operational risks. This means that explainable prediction models can also be used to identify the causes of failures, which can help to improve the accuracy and efficiency of predictive maintenance.

The proposed framework is related to the general trend of shifting towards intelligent and human-centred manufacturing systems from an industrial point of view. Transparent predictive models promote collaboration among AI systems and maintenance personnel by offering actionable insights instead of abstract predictions. This type of cooperation can help address responsible decision making, ease the acceptance of artificial intelligence in the organisation and build trust in digital manufacturing technologies. In the future, as manufacturing evolves towards Industry 5.0, the importance of explainability and predictive maintenance will grow in relevance for a balance between automated systems and human skills, sustainable, reliable, and ethical manufacturing processes and applications.

5. Conclusion

To support predictive maintenance in smart manufacturing environment, an explainable machine failure prediction framework is developed in this study by using the AI4I 2020 Predictive Maintenance dataset. It showed that with regard to the ability to distinguish between normal operation and potentially incurring machine failure, operational variables like the rotational speed, torque, tool wear and temperature are indicative. Further identification of a number of failure mechanisms demonstrated that equipment degradation is affected by a combination of mechanical and thermal conditions, instead of only one operating condition. The incorporation of explainable artificial intelligence with machine learning helped to make the machine failures more transparent by interpreting the variables that influence the machine in making predictions. This interpretability can improve the trust in predictive models and help in making better-informed maintenance plans in industrial environments. The envisioned framework provides a step towards reliable, transparent and data-driven predictive maintenance through accurate failure prediction and interpretable decision support. In practical terms, the results can help manufacturing companies to minimize unscheduled equipment downtime, better schedule maintenance, maximize asset utilization and enhance the operational efficiency. Future work is needed in validation of the proposed approach with real time industrial sensor data, exploration of explainability techniques based on deep learning, and testing the framework in different manufacturing processes for its scalability and industrial applicability.

References

1. Abidi, M. H., Mohammed, M. K., & Alkhalefah, H. (2022). Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing. Sustainability, 14(6), 3387.
2. 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), 8081.
3. Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. Research in International Business and Finance, 61, 101649.
4. Chang, T. S., & Bau, D. Y. (2025). eXplainable artificial intelligence (XAI) in business management research: a success/failure system perspective. Journal of Electronic Business & Digital Economics, 4(1), 36-53.
5. Chen, T. C. T. (2023). Explainable artificial intelligence (XAI) in manufacturing. In Explainable Artificial Intelligence (XAI) in Manufacturing: Methodology, Tools, and Applications (pp. 1-11). Cham: Springer International Publishing.
6. Cinar, E., Kalay, S., & Saricicek, I. (2022). A predictive maintenance system design and implementation for intelligent manufacturing. Machines, 10(11), 1006.
7. 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), 8211.
8. Elgnemi, T., Songmene, V., Kouam, J., Jun, M. B., & Samuel, A. M. (2021). Experimental investigation on dry routing of CFRP composite: temperature, forces, tool wear, and fine dust emission. Materials, 14(19), 5697.
9. Falatouri, T., Nasseri, M., Brandtner, P., & Darbanian, F. (2023, July). Shedding light on the black box: explainable AI for predicting household appliance failures. In International Conference on Human-Computer Interaction (pp. 69-83). Cham: Springer Nature Switzerland.
10. Farooq, U., Ademola, M., & Shaalan, A. (2024). Comparative analysis of machine learning models for predictive maintenance of ball bearing systems. Electronics, 13(2), 438.
11. Jamil, M., He, N., Gupta, M. K., Zhao, W., & Khan, A. M. (2022). Tool wear mechanisms and its influence on machining tribology of face milled titanium alloy under sustainable hybrid lubri-cooling. Tribology International, 170, 107497.
12. Kusiak, A. (2024). Federated explainable artificial intelligence (fXAI): a digital manufacturing perspective. International Journal of Production Research, 62(1-2), 171-182.
13. Levin, S. (2024, December). Enhancing predictive maintenance in the industrial sector: A comparative analysis of machine learning models. In AIP Conference Proceedings (Vol. 3243, No. 1, p. 020035). AIP Publishing LLC.
14. Love, P. E., Fang, W., Matthews, J., Porter, S., Luo, H., & Ding, L. (2023). Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. Advanced Engineering Informatics, 57, 102024.
15. Matzka, S. (2020). AI4I 2020 Predictive Maintenance Dataset. UCI Machine Learning Repository. https://doi.org/10.24432/C5HS5C
16. Merchán-Cruz, E. A., Gabelaia, I., Savrasovs, M., Hansen, M. F., Soe, S., Rodriguez-Cañizo, R. G., & Aragón-Camarasa, G. (2025). Trust by design: An ethical framework for collaborative intelligence systems in Industry 5.0. Electronics, 14(10), 1952.
17. Meriem, H., Nora, H., & Samir, O. (2023). Predictive maintenance for smart industrial systems: a roadmap. Procedia Computer Science, 220, 645-650.
18. Mohammed, N. A., Abdulateef, O. F., Hamad, A. H., & Abdullah, O. I. (2024). Performance analysis of different machine learning algorithms for predictive maintenance. Al-Khwarizmi Engineering Journal, 20(2), 26-38.
19. Pech, M., Vrchota, J., & Bednář, J. (2021). Predictive maintenance and intelligent sensors in smart factory. Sensors, 21(4), 1470.
20. Puthanveettil Madathil, A., Luo, X., Liu, Q., Walker, C., Madarkar, R., & Qin, Y. (2025). A review of explainable artificial intelligence in smart manufacturing. International Journal of Production Research, 63(23), 8654-8697.
21. Ramzan, F., & Reforgiato Recupero, D. (2025, December). A literature review on enhancing predictive maintenance in smart manufacturing industries: fostering human-technology collaboration and overcoming data scarcity limitations with advanced AI models. In Operations Research Forum (Vol. 6, No. 4, p. 181). Cham: Springer International Publishing.
22. ŞAHiN, E., Arslan, N. N., & Özdemir, D. (2025). Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning. Neural computing and applications, 37(2), 859-965.
23. Taşcı, B., Omar, A., & Ayvaz, S. (2023). Remaining useful lifetime prediction for predictive maintenance in manufacturing. Computers & Industrial Engineering, 184, 109566.
24. Thalpage, N. (2023). Unlocking the black box: Explainable artificial intelligence (XAI) for trust and transparency in ai systems. J. Digit. Art Humanit, 4(1), 31-36.
25. Tzionis, G., Mouratidis, P., Kougka, G., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I., & Vlachopoulou, M. (2025). A review of explainable AI methods and their application in manufacturing systems. Discover Applied Sciences, 8(1), 52.