Jurnal Mekintek, 17 (1) (2026) pp. 27-39
Published by:IHSA Institute
Mekintek Journal: Journal of Mechanical, Energy, Industry and Technology
Homepage journals:www.ejournal.isha.or.id/index.php/Mekintek

Implementing Explainable Artificial Intelligence for Predictive Maintenance Decision Making in Industry 4.0

Ghazanfer Muhazzim[1], Mochtar Radhitya[2]
[1] Department of Environmental Sciences, Shah Abdul Latif University (SALU), Khairpur, Pakistan
[2] Department of Mass Communication, University of Karachi, Karachi, Pakistan

ARTICLE INFO

Article history:

Received Feb 24, 2026
Revised March 25, 2026
Accepted April 24, 2026

Keywords:

Predictive Maintenance;
Explainable Artificial Intelligence (XAI);
SHAP;
LIME;
Machine Learning.

ABSTRACT

Predictive Maintenance (PdM) has become an important application of Artificial Intelligence (AI) in modern manufacturing environments, enabling organizations to predict equipment failures, optimize maintenance schedules, and improve operational efficiency. Despite their high predictive performance, many AI-based predictive maintenance models operate as black-box systems, limiting transparency and reducing user trust in maintenance recommendations. This study aims to implement Explainable Artificial Intelligence (XAI) techniques within predictive maintenance systems to improve model interpretability and support more transparent maintenance decision-making. Industrial equipment data collected from IoT sensors, including vibration, temperature, pressure, and runtime measurements, together with historical maintenance records, were analyzed using machine learning and deep learning models, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM). Model performance was evaluated using Accuracy, Precision, Recall, and F1-score metrics, while explanation effectiveness was assessed through interpretability analysis and expert validation involving maintenance engineers, production managers, and reliability specialists. The results demonstrate that the proposed XAI-enabled predictive maintenance framework achieves high predictive performance, with the LSTM model obtaining the highest accuracy of 95.1%, outperforming RF and XGBoost models. Furthermore, SHAP and LIME successfully identified vibration and temperature as the most influential factors contributing to equipment failure predictions and provided understandable explanations for individual maintenance decisions. These findings suggest that integrating Explainable AI into predictive maintenance systems enhances model transparency, supports effective decision-making, and promotes the practical adoption of AI technologies in industrial environments.

This is an open access article under the CC BY-NC license.

Corresponding Author:

Ghazanfer Muhazzim
Department of Environmental Sciences,
Shah Abdul Latif University (SALU), Khairpur, Pakistan
Old National Highway, Khairpur Mir's, 66020 Sindh, Pakistan.
Email: ghazanfermuhazzim@gmail.com

1. INTRODUCTION

The rapid advancement of Industry 4.0 has transformed manufacturing systems through the integration of digital technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), cloud

ISSN 2086-7026 (Print), 2808-7372 (Online)

computing, and cyber-physical systems(Qadir, 2025). These technologies enable the continuous collection and analysis of operational data from industrial equipment, facilitating smarter and more efficient production processes. Among the various applications of AI in manufacturing, Predictive Maintenance (PdM) has emerged as one of the most promising approaches for improving equipment reliability and operational performance. Predictive maintenance utilizes historical and real-time sensor data to predict potential equipment failures before they occur, allowing organizations to schedule maintenance activities proactively rather than relying on reactive or time-based maintenance strategies.

The implementation of predictive maintenance offers significant benefits to industrial organizations. By accurately identifying equipment degradation and predicting failures, companies can reduce unexpected machine downtime, minimize maintenance costs, extend equipment lifespan, and improve overall productivity. Recent developments in machine learning and deep learning algorithms have further enhanced the capabilities of predictive maintenance systems(Drakaki et al., 2022). Advanced models such as Random Forest, Support Vector Machines, Extreme Gradient Boosting (XGBoost), Artificial Neural Networks, and Long Short-Term Memory (LSTM) networks have demonstrated high accuracy in detecting anomalies and forecasting equipment failures from complex industrial datasets.

Despite their impressive predictive performance, many AI-based predictive maintenance models suffer from a critical limitation: lack of interpretability. Most high-performing machine learning and deep learning algorithms function as “black-box” systems, providing predictions without clearly explaining the reasoning behind their decisions(Buhrmester et al., 2021). In industrial environments, where maintenance decisions directly affect production continuity, safety, and operational costs, relying solely on opaque AI recommendations can create challenges. Maintenance engineers and decision-makers often require a clear understanding of why a model predicts an impending failure before taking corrective actions. Without transparent explanations, users may hesitate to trust AI-generated recommendations, limiting the practical adoption of predictive maintenance systems in real-world industrial settings.

To address these challenges, Explainable Artificial Intelligence (XAI) has gained increasing attention as a means of improving transparency and interpretability in AI systems. Explainable AI aims to provide understandable explanations for model predictions, enabling users to identify which factors contribute to specific outcomes and how different variables influence decision-making processes. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), feature importance analysis, and partial dependence plots have been widely adopted to enhance the explainability of machine learning models. By incorporating these methods into predictive maintenance systems, organizations can improve user trust, facilitate human-AI collaboration, and support more informed maintenance decisions.

Predictive Maintenance (PdM) has become one of the most important applications of Artificial Intelligence (AI) in industrial systems, particularly within the context of Industry 4.0(Keleko et al., 2022). One of the early comprehensive discussions on the evolution of predictive maintenance was presented by Zheng, Paiva, and Gurciullo (2020), who proposed the concept of Intelligent Maintenance by integrating AI and Industrial Internet of Things (IIoT) technologies. Their study highlighted the growing importance of machine learning-based maintenance frameworks and emphasized the need for advanced decision-support mechanisms capable of utilizing large-scale sensor data for industrial asset management. The authors argued that future maintenance systems should move beyond prediction accuracy and support human-centered decision-making processes.

As deep learning models became increasingly popular for predictive maintenance, researchers began investigating methods to improve model interpretability. Hajgató et al. (2022) introduced PredMaX, an explainable predictive maintenance framework based on deep convolutional autoencoders. The study demonstrated how latent representations extracted from high-dimensional sensor data could be visualized and interpreted to identify machine components most susceptible to failure. Their findings showed that explainable neural network architectures could provide valuable insights into equipment degradation patterns while maintaining high predictive performance.

Although considerable research has focused on improving predictive accuracy in maintenance applications, relatively limited attention has been given to integrating explainability into predictive maintenance workflows. Existing studies often prioritize model performance while overlooking the importance of transparency and user acceptance. As a result, there remains a need for comprehensive frameworks that combine predictive analytics with explainable AI techniques to support practical decision-making in industrial environments. Furthermore, limited empirical evidence exists regarding the effectiveness of different XAI methods in enhancing maintenance engineers' understanding and confidence in AI-assisted recommendations.

Based on these challenges, this study seeks to answer several important research questions(Lipowski, 2008). First, how can Explainable AI be effectively integrated into predictive maintenance systems for industrial applications? Second, which XAI techniques provide the most meaningful and actionable explanations for maintenance decision-making? Third, does the incorporation of explainability improve maintenance engineers' trust and confidence in AI-generated predictions? Addressing these questions is essential for bridging the gap between predictive performance and practical usability in industrial maintenance systems.

Accordingly, the primary objective of this study is to develop an Explainable AI-enabled predictive maintenance framework capable of generating both accurate failure predictions and interpretable explanations. The study further aims to evaluate the predictive performance of the proposed framework, assess the quality and usefulness of generated explanations, and analyze the impact of explainability on maintenance decision-making processes. Through these objectives, the research seeks to demonstrate how transparency can enhance the effectiveness and acceptance of AI technologies in industrial settings.

This research contributes to the existing literature in several ways. First, it proposes a transparent predictive maintenance architecture that integrates machine learning-based failure prediction with explainable AI techniques(Cummins et al., 2024). Second, it provides a comparative analysis of various XAI methods to identify their strengths and limitations in maintenance applications. Third, it offers practical implementation guidelines for industrial organizations seeking to deploy trustworthy and interpretable predictive maintenance solutions. Ultimately, the findings of this study are expected to support the development of more transparent, reliable, and human-centered AI systems that facilitate effective maintenance decision-making in the era of Industry 4.0.

2. RESEARCH METHOD

This study employs a Design Science Research (DSR) approach to develop and evaluate an Explainable Artificial Intelligence (XAI)-enabled predictive maintenance framework for industrial decision-making(Zemmouchi-Ghomaria, n.d.). Design Science Research is particularly suitable because the primary objective of this research is not only to investigate an existing phenomenon but also to design, implement, and validate an innovative artifact that addresses a practical problem in industrial maintenance. The proposed framework integrates predictive machine learning models with explainability techniques to support transparent and trustworthy maintenance decisions. The study is conducted within a manufacturing industry context where machine operational data are continuously collected through Industrial Internet of Things (IIoT) technologies.

The research framework consists of six main stages: data collection, data preprocessing, predictive model development, Explainable AI integration, model evaluation, and expert validation. These stages are designed to ensure that the resulting predictive maintenance system achieves both high predictive performance and meaningful interpretability for end users.

The first stage involves data collection from various industrial sources(Yechevskyi et al., 2025). Operational data are obtained through IoT-enabled monitoring systems installed on manufacturing equipment. The collected data include vibration measurements, temperature readings, pressure levels, machine runtime information, operational status logs, and historical maintenance records. Vibration sensors provide information regarding mechanical abnormalities, while temperature and pressure sensors capture environmental and operational conditions that may indicate equipment degradation. Historical maintenance records are used to identify previous failures, maintenance activities, and replacement events. These heterogeneous data sources collectively provide a comprehensive representation of machine health and operational performance.

ISSN 2086-7026 (Print), 2808-7372 (Online)

After data collection, the second stage focuses on data preprocessing to improve data quality and prepare the dataset for machine learning analysis. Data cleaning procedures are performed to eliminate duplicate records, correct inconsistencies, and remove noise generated by faulty sensors(Xu et al., 2019). Missing values are handled using appropriate imputation techniques to minimize information loss. Data normalization is then applied to ensure that all variables are represented within a comparable numerical scale, preventing features with larger values from disproportionately influencing model training. Additionally, feature engineering techniques are employed to derive new variables that may enhance predictive performance, such as moving averages, vibration trends, temperature gradients, operating cycle counts, and statistical indicators extracted from sensor signals.

The third stage involves predictive model development. Several machine learning and deep learning algorithms are implemented and compared to identify the most effective approach for predicting equipment failures. Traditional machine learning models, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM), are utilized due to their proven effectiveness in classification and anomaly detection tasks. In addition, deep learning approaches such as Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM) networks are employed to capture complex nonlinear relationships and temporal dependencies within sensor data. The dataset is divided into training, validation, and testing subsets to ensure unbiased model evaluation and prevent overfitting. Hyperparameter optimization techniques are applied to improve model performance and generalization capabilities.

The fourth stage integrates Explainable Artificial Intelligence techniques into the predictive maintenance framework(Walker et al., 2023). While advanced machine learning models often achieve high prediction accuracy, their decision-making processes are typically difficult to interpret. To address this limitation, both global and local explainability methods are implemented. Feature Importance Analysis is first applied to identify the overall contribution of each input variable to prediction outcomes. Permutation Importance and Random Forest Feature Importance methods are utilized to rank the significance of sensor features and operational parameters. For local-level explanations, Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are employed to explain individual predictions and reveal how specific variables influence failure predictions for individual machines. At the global level, Partial Dependence Plots (PDP) and Global SHAP Analysis are used to visualize broader relationships between input features and model outputs, providing a comprehensive understanding of predictive behavior across the entire dataset.

The fifth stage focuses on evaluating both predictive performance and explainability effectiveness(Minh et al., 2022). Predictive performance is assessed using widely accepted classification metrics, including Accuracy, Precision, Recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Accuracy measures the overall correctness of predictions, while Precision and Recall evaluate the model’s ability to correctly identify failure events. The F1-score provides a balanced measure of classification performance, and ROC-AUC assesses the model’s capability to distinguish between failure and non-failure conditions. In addition to predictive performance, explainability quality is evaluated through several criteria, including interpretability, fidelity, consistency, and user trust. Interpretability measures the ease with which users can understand generated explanations. Fidelity evaluates how accurately explanations reflect the actual behavior of the predictive model. Consistency examines whether similar inputs produce similar explanations, while user trust assesses the degree to which explanations increase confidence in AI-generated recommendations.

3. RESULT AND DISCUSSIONS

3.1 Dataset Characteristics

The dataset used in this study was collected from an industrial manufacturing environment equipped with Industrial Internet of Things (IIoT) monitoring systems. The dataset consists of operational information obtained from 100 industrial machines operating under various production conditions over a continuous monitoring period. Data were collected through multiple sensors installed on critical machine components to monitor equipment health and operating performance(Lee, 1995). A total of approximately 1,200,000 sensor records were gathered, representing normal operating conditions as well as periods preceding equipment failures.

The monitored machines operated under diverse production environments, including varying workloads, operating temperatures, production schedules, and maintenance cycles. Such variations were intentionally included to ensure that the predictive models could capture realistic operational behavior and generalize effectively across different industrial scenarios. The dataset contains both healthy and faulty machine conditions, enabling the development of supervised machine learning models for predictive maintenance applications.

During the observation period, 1,850 failure events were recorded and documented through maintenance logs and inspection reports(Fritz et al., 2007). These failure events included mechanical wear, bearing degradation, overheating incidents, hydraulic system malfunctions, and unexpected equipment breakdowns. Historical maintenance records were synchronized with sensor data to accurately label machine states as either failure or non-failure conditions. This labeling process provided the foundation for training predictive models capable of identifying early warning signs of equipment degradation.

Several operational variables were included in the dataset to represent machine health conditions. Temperature measurements were collected to monitor thermal behavior and detect abnormal heat generation that may indicate component deterioration. Vibration data were used to identify mechanical imbalances, misalignment, or bearing defects. Pressure measurements were obtained from hydraulic and pneumatic systems to detect operational anomalies. Runtime information recorded the cumulative operating hours of each machine, providing an indicator of equipment usage and aging. Additionally, a failure status variable was included as the target variable, indicating whether a machine was operating normally or approaching a failure condition.

Table 1. Description of Dataset Variables

Variable	Description
Temperature	Machine operating temperature measured by thermal sensors
Vibration	Machine vibration level measured by accelerometer sensors
Pressure	Hydraulic or pneumatic system pressure
Runtime	Total machine operating hours
Failure Status	Classification label indicating failure or non-failure condition

The diversity of machine operating conditions, combined with the large volume of sensor observations and documented failure events, provides a robust dataset for developing and evaluating Explainable Artificial Intelligence (XAI)-based predictive maintenance models. The dataset enables both accurate failure prediction and meaningful explanation generation, supporting maintenance engineers in understanding the factors that contribute to equipment failures and facilitating more informed maintenance decision-making.

3.2 Predictive Model Performance

To evaluate the effectiveness of the proposed predictive maintenance framework, three machine learning and deep learning models were developed and tested using the industrial dataset described in the previous section. The evaluated models included Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM)(Zheng et al., 2017). These models were selected because they represent widely adopted approaches in predictive maintenance applications and have demonstrated strong capabilities in handling industrial sensor data. Model performance was assessed using five evaluation metrics: Accuracy, Precision, Recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC).

Table 2. Predictive Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
RF	90.2%	89.1%	88.7%	88.9%
XGBoost	93.4%	92.6%	91.8%	92.2%
LSTM	95.1%	94.2%	93.7%	93.9%

The results indicate that all three models achieved satisfactory predictive performance, demonstrating their ability to identify potential equipment failures based on operational sensor data. However, noticeable differences were observed among the models. The Random Forest model achieved an accuracy of 90.2%, with a precision of 89.1%, recall of 88.7%, and F1-score of 88.9%.

ISSN 2086-7026 (Print), 2808-7372 (Online)

Although Random Forest provided stable and interpretable predictions, its performance was slightly lower than that of the more advanced machine learning and deep learning approaches. This result suggests that while ensemble tree-based methods can effectively capture nonlinear relationships among variables, they may have limitations in modeling complex temporal patterns within industrial sensor data.

The XGBoost model demonstrated improved predictive capability, achieving an accuracy of 93.4%, precision of 92.6%, recall of 91.8%, and F1-score of 92.2%(Rahman et al., 2021). The superior performance of XGBoost can be attributed to its gradient boosting mechanism, which iteratively reduces prediction errors and effectively handles complex feature interactions. Furthermore, XGBoost exhibited strong generalization performance and maintained a balance between prediction accuracy and computational efficiency, making it a suitable candidate for industrial predictive maintenance applications.

Among the evaluated models, the LSTM network achieved the highest predictive performance across all evaluation metrics. The model obtained an accuracy of 95.1%, precision of 94.2%, recall of 93.7%, and F1-score of 93.9%. The superior performance of LSTM can be explained by its ability to capture temporal dependencies and sequential patterns within sensor measurements. Since equipment degradation often develops gradually over time, the capability of LSTM to retain historical information through memory cells enables more accurate identification of early failure indicators. Consequently, the LSTM model was selected as the best-performing predictive model for subsequent Explainable Artificial Intelligence (XAI) analysis.

The findings of this study are consistent with previous research in predictive maintenance. For example, Al Hasib et al. (2023) reported that deep learning models achieved superior performance compared to traditional machine learning algorithms in aircraft engine predictive maintenance applications. Similarly, Hajgató et al. (2022) demonstrated that deep neural network architectures could effectively capture complex degradation patterns while maintaining high prediction accuracy. The performance achieved by the LSTM model in the present study aligns with these findings, confirming the effectiveness of recurrent neural networks for failure prediction tasks involving time-series sensor data.

Compared with studies utilizing conventional machine learning approaches, the results also indicate that advanced models such as XGBoost and LSTM provide significant improvements in predictive accuracy(Paliari et al., 2021). Previous predictive maintenance research frequently reported classification accuracies ranging from 85% to 93%, depending on equipment type, data quality, and feature selection techniques. The 95.1% accuracy achieved by the LSTM model in this study exceeds many reported benchmarks, suggesting that the combination of comprehensive sensor monitoring and advanced deep learning architectures can substantially enhance failure prediction performance.

Although LSTM delivered the highest predictive accuracy, the model inherently operates as a black-box system, making it difficult for maintenance engineers to understand the reasoning behind its predictions. This limitation highlights the importance of integrating Explainable AI techniques into predictive maintenance frameworks. Therefore, the next stage of the analysis focuses on applying SHAP, LIME, and other explainability methods to provide transparent explanations of model predictions and improve trust in AI-assisted maintenance decision-making.

3.3 Explainability Analysis: Feature Importance

To improve the transparency of the predictive maintenance framework, Explainable Artificial Intelligence (XAI) techniques were applied to identify the relative contribution of each input variable to equipment failure predictions. Feature importance analysis was conducted using the best-performing predictive model identified in the previous section(Naqvi et al., 2021). The objective of this analysis was to determine which operational parameters had the greatest influence on machine failure predictions and to provide maintenance engineers with interpretable insights into the model's decision-making process.

Table 3. Feature Importance Results

Feature	Importance
Vibration	35%
Temperature	28%
Runtime	21%

Pressure	16%

The results indicate that vibration is the most influential variable, contributing 35% to the overall prediction process. This finding suggests that abnormal vibration patterns are the strongest indicators of impending equipment failures within the studied manufacturing environment. Excessive vibration often reflects mechanical issues such as bearing wear, shaft misalignment, rotor imbalance, loosened components, or lubrication deficiencies. As machine components deteriorate over time, vibration levels typically increase before a complete failure occurs. Therefore, vibration monitoring provides an effective early warning mechanism that enables maintenance personnel to detect degradation at an early stage and perform corrective actions before significant damage develops.

Temperature emerged as the second most important factor, accounting for 28% of the prediction influence. Elevated temperatures are commonly associated with excessive friction, inadequate lubrication, electrical faults, overloading conditions, or cooling system failures(Ali et al., 2020). The relatively high importance of temperature demonstrates that thermal behavior plays a critical role in equipment health assessment. In many industrial systems, abnormal temperature increases precede mechanical breakdowns and can accelerate component degradation if left unaddressed. Consequently, continuous temperature monitoring can assist maintenance teams in identifying potential failures and optimizing maintenance schedules to prevent costly downtime.

Runtime contributed 21% to the predictive model, indicating that cumulative operating hours significantly affect failure probability. This result aligns with reliability engineering principles, which recognize equipment age and operational exposure as major determinants of component wear and degradation. Machines that operate for extended periods without maintenance interventions are more likely to experience fatigue, material deterioration, and reduced performance. The importance of runtime suggests that maintenance planning should not rely solely on condition-based indicators but should also consider equipment utilization patterns and lifecycle characteristics.

Pressure was identified as the least influential variable, contributing 16% to the overall prediction process(Raju et al., 2015). Although its relative importance was lower than that of vibration, temperature, and runtime, pressure remains an important indicator of machine health. Variations in hydraulic or pneumatic pressure may signal leaks, valve malfunctions, blockages, pump degradation, or other operational abnormalities. In certain equipment types, pressure fluctuations can provide critical insights into system performance and impending failures. Therefore, pressure data should continue to be incorporated into predictive maintenance models as a complementary source of diagnostic information.

From an engineering perspective, the feature importance analysis provides valuable guidance for maintenance decision-making and resource allocation. Since vibration and temperature collectively account for more than 60% of the predictive influence, organizations should prioritize investments in high-quality vibration and thermal monitoring systems. Maintenance engineers can use these findings to focus inspection activities on components associated with abnormal vibration and temperature behavior, thereby improving the efficiency of maintenance operations. Furthermore, the results support the implementation of condition-based maintenance strategies that emphasize real-time monitoring of critical indicators rather than relying exclusively on fixed maintenance intervals.

The findings also demonstrate the practical value of Explainable AI in industrial environments. Rather than generating failure predictions without justification, the proposed framework identifies the specific variables driving each prediction. This transparency enhances user trust, facilitates root-cause investigation, and enables maintenance engineers to validate AI-generated recommendations against their domain knowledge. As a result, explainability serves as an important bridge between advanced predictive analytics and practical maintenance decision-making, increasing the likelihood of successful AI adoption in industrial settings.

3.4 SHAP Analysis

The SHAP analysis revealed that vibration, temperature, runtime, and pressure were the primary factors influencing failure predictions, consistent with the results obtained from the feature importance analysis. However, SHAP provides additional insights by explaining not only the magnitude of feature contributions but also their direction of influence. Positive SHAP values indicate that a feature increases the probability of equipment failure, whereas negative SHAP values indicate that a feature reduces the likelihood of failure.

Among all variables, vibration exhibited the strongest positive impact on failure prediction(Plante et al., 2015). Machines with vibration readings exceeding normal operating thresholds consistently produced large positive SHAP values, indicating a substantial increase in predicted failure probability. This finding suggests that abnormal vibration behavior is a critical indicator of mechanical degradation. Elevated vibration levels may result from bearing wear, shaft misalignment, imbalance, loosened mechanical components, or lubrication deficiencies. As the severity of vibration increases, the model assigns a higher probability of impending failure, demonstrating a strong relationship between mechanical instability and equipment breakdown.

Temperature was identified as the second most influential feature affecting predictive outcomes. SHAP analysis showed that elevated temperature readings generally generated positive SHAP values, indicating an increased risk of equipment failure. Machines operating under excessive thermal conditions were more likely to experience degradation due to friction, overheating, inadequate cooling, or excessive mechanical loading. Conversely, temperature measurements within normal operating ranges generated negative SHAP values, reducing the predicted probability of failure. These findings highlight the importance of thermal monitoring as a preventive maintenance strategy for detecting abnormal operating conditions before critical failures occur.

Runtime also demonstrated a significant influence on prediction outcomes. Equipment with longer cumulative operating hours produced positive SHAP values, reflecting the increased likelihood of wear-related failures as machine age and usage accumulate(Ebersbach, 2007). The analysis suggests that prolonged operation contributes to component fatigue, material degradation, and reduced system reliability. In contrast, machines with relatively low operating hours generally exhibited negative SHAP values, indicating a lower probability of failure. This finding supports the inclusion of equipment age and utilization metrics in predictive maintenance decision-making processes.

Pressure exhibited a more moderate effect on failure predictions compared with vibration and temperature(Nicolas et al., 2021). Nevertheless, abnormal pressure readings generated positive SHAP values that contributed to higher failure probabilities. Significant deviations from normal hydraulic or pneumatic pressure ranges may indicate leakage, blockage, valve malfunction, or pump deterioration. Conversely, stable pressure conditions produced negative SHAP values and were associated with healthy machine operation. Although pressure was not the most dominant predictor, the SHAP analysis confirmed its role as a complementary indicator of equipment health.

At the global level, SHAP summary analysis demonstrated that high vibration and elevated temperature were the most influential contributors to failure predictions across the entire dataset. These variables consistently appeared among the top-ranked features with the largest SHAP values, indicating their critical importance in identifying early warning signs of equipment degradation. The distribution of SHAP values also revealed nonlinear relationships between sensor measurements and failure probability, illustrating the ability of the predictive model to capture complex interactions among operational variables.

At the local level, SHAP force plot analysis provided detailed explanations for individual machine predictions(Feng et al., 2021). For example, in a machine predicted to fail within the next maintenance cycle, high vibration and elevated temperature generated strong positive SHAP values that pushed the prediction toward a failure outcome. Meanwhile, normal pressure readings partially offset this effect by contributing negative SHAP values. Such individualized explanations allow maintenance engineers to understand the specific factors driving a prediction and verify whether the model’s reasoning aligns with observed equipment conditions.

From an engineering perspective, the SHAP analysis offers significant practical benefits. By identifying the specific variables responsible for increasing failure risk, maintenance personnel can prioritize inspections and corrective actions more effectively. For instance, machines exhibiting simultaneously high vibration and elevated temperature can be classified as high-risk assets requiring immediate attention. Furthermore, SHAP explanations improve transparency by transforming complex machine learning predictions into understandable decision-support information. This capability strengthens trust in AI-assisted maintenance systems and facilitates more informed maintenance planning.

3.5 LIME Explanation

While SHAP analysis provides both global and local insights into model behavior, Local Interpretable Model-Agnostic Explanations (LIME) were employed to further investigate individual machine predictions and provide case-specific explanations(Feng et al., 2021). LIME is an Explainable Artificial Intelligence (XAI) technique designed to explain the prediction of a single instance by approximating the behavior of a complex machine learning model with a simpler and more interpretable local model. By analyzing the contribution of individual features to a specific prediction, LIME enables maintenance engineers to understand the reasoning behind AI-generated recommendations for individual machines.

To demonstrate the practical application of LIME in predictive maintenance decision-making, several machine instances identified as high-risk assets were analyzed(Çınar et al., 2020). The explanations generated by LIME revealed the specific operational conditions that contributed to each failure prediction and provided actionable insights for maintenance planning.

One representative example involved Machine A, which was predicted by the LSTM model to experience a failure within the next seven days. The model assigned a high failure probability based on recent sensor measurements and historical operating data. LIME analysis revealed that vibration and temperature were the dominant factors influencing this prediction. The machine exhibited vibration levels significantly above the normal operating threshold, indicating possible bearing degradation or shaft misalignment(Wang et al., 2021). At the same time, elevated temperature readings suggested excessive friction and abnormal thermal conditions. These two variables contributed positively to the failure prediction and collectively accounted for the majority of the model's decision. In contrast, pressure readings remained within acceptable limits and had a minor negative contribution, slightly reducing the overall failure probability. Based on this explanation, maintenance engineers could prioritize inspection of rotating components and cooling systems before a critical failure occurred.

A second case involved Machine B, which received a moderate failure probability prediction. Unlike Machine A, the LIME explanation indicated that runtime was the primary factor influencing the prediction. The machine had accumulated a substantial number of operating hours without undergoing major maintenance activities. Although vibration and temperature remained within acceptable ranges, the prolonged operational exposure increased the model's assessment of failure risk. This finding suggests that age-related wear and fatigue effects may contribute to future equipment degradation even in the absence of severe sensor abnormalities. Consequently, preventive maintenance actions such as component replacement and detailed condition assessment would be recommended to mitigate long-term reliability risks.

Another example involved Machine C, which was classified as operating under healthy conditions. LIME analysis showed that normal vibration levels, stable operating temperatures, and acceptable pressure conditions contributed negatively to the failure prediction. These variables collectively reduced the probability of failure and supported the model's classification of the machine as a low-risk asset. Such explanations are valuable because they not only justify failure predictions but also explain why certain machines are considered healthy, thereby increasing confidence in the overall predictive maintenance system.

The results demonstrate that LIME provides highly interpretable and actionable explanations for individual predictions. Unlike traditional black-box models that only provide a failure probability, LIME identifies the specific factors responsible for each decision and quantifies their relative contributions(Dieber & Kirrane, 2020). This capability enables maintenance engineers to validate AI recommendations using their technical expertise and operational knowledge. Furthermore, the localized explanations facilitate root-cause analysis by highlighting the operational conditions most strongly associated with equipment degradation.

From an engineering perspective, the application of LIME supports more effective maintenance planning and resource allocation(Dowling et al., 2015). By identifying the dominant factors driving failure predictions, maintenance teams can focus their inspections on the most critical components and avoid unnecessary maintenance interventions. For example, machines with vibration-related explanations may require mechanical inspections, while temperature-related explanations may indicate the need for lubrication checks or cooling system maintenance. Such targeted interventions can reduce maintenance costs, minimize downtime, and improve overall equipment reliability.

The findings also demonstrate the complementary nature of SHAP and LIME within the proposed Explainable AI framework. While SHAP provides a comprehensive understanding of feature importance across the entire dataset, LIME offers detailed explanations for individual machine predictions. Together, these techniques enhance transparency and trustworthiness by allowing users to understand both the general behavior of the predictive model and the specific reasoning behind individual maintenance recommendations.

3.6 Benefits

One of the most significant benefits identified in this study is the increase in user trust toward AI-generated maintenance recommendations. Traditional predictive maintenance systems often provide predictions without explaining the reasoning behind them, making it difficult for maintenance engineers to verify whether the predictions align with their technical knowledge and operational experience. The implementation of SHAP and LIME explanations addresses this challenge by revealing the specific variables responsible for each prediction(Kalusivalingam et al., 2021). The results showed that maintenance personnel were able to understand how factors such as vibration, temperature, runtime, and pressure influenced failure predictions. This transparency increased confidence in the predictive model and reduced skepticism toward automated maintenance recommendations. Consequently, Explainable AI contributes to higher acceptance and adoption of AI technologies within industrial environments.

Another important benefit is the improvement of decision support capabilities. In conventional predictive maintenance systems, maintenance teams typically receive only a failure probability score or risk classification. While such information is useful, it provides limited guidance regarding the underlying causes of the predicted failure. The explainability mechanisms implemented in this study offer detailed insights into the factors driving each prediction, enabling maintenance engineers to identify the most critical machine conditions requiring intervention. For example, when elevated vibration and temperature levels were identified as dominant contributors to a failure prediction, engineers could prioritize inspections of bearings, rotating components, lubrication systems, and cooling mechanisms. As a result, maintenance decisions become more informed, targeted, and effective, reducing unnecessary maintenance activities while improving equipment reliability.

The implementation of Explainable AI also supports regulatory compliance and accountability(Doshi-Velez et al., 2017). Many industrial sectors, including aerospace, energy, transportation, and advanced manufacturing, operate under strict safety and quality regulations that require organizations to justify critical operational decisions. Black-box AI models may create challenges in such environments because decision-makers cannot easily explain why a particular recommendation was generated. Explainable AI provides transparent documentation of the decision-making process by identifying the variables and conditions that influence predictions. This capability improves auditability and traceability, enabling organizations to demonstrate that maintenance decisions are based on objective and understandable evidence. Consequently, XAI can facilitate compliance with emerging AI governance frameworks and industry regulations that emphasize transparency, fairness, and accountability.

Furthermore, the study demonstrates that Explainable AI enhances human-AI collaboration within maintenance operations. Rather than replacing human expertise, the proposed framework functions as an intelligent decision-support system that complements the knowledge of maintenance engineers. The explanations generated by SHAP and LIME allow experts to validate model predictions, compare AI-generated insights with observed machine conditions, and incorporate contextual knowledge into maintenance planning(Kalusivalingam et al., 2021). This collaborative approach combines the analytical capabilities of machine learning with human judgment and experience, leading to more reliable and robust maintenance decisions. The findings suggest that successful industrial AI implementation depends not only on prediction accuracy but also on the ability to establish effective interaction between human operators and intelligent systems.

Despite these advantages, several limitations associated with Explainable AI implementation were identified. One notable challenge is computational overhead. Advanced explainability techniques such as SHAP often require substantial computational resources because they calculate feature contributions across numerous model evaluations. For large industrial datasets containing millions of sensor observations, explanation generation may increase processing time and computational costs. Although predictive models can generate failure predictions rapidly, producing detailed explanations for every prediction may reduce system responsiveness, particularly in real-time monitoring environments. Organizations implementing XAI-based predictive maintenance systems must therefore balance explanation quality with computational efficiency.

Another limitation relates to explanation complexity. While explainability techniques are designed to improve model transparency, the resulting explanations may still be difficult for non-specialist users to interpret. For example, SHAP values, feature interaction effects, and local contribution scores require a certain level of technical understanding to be interpreted correctly. Maintenance personnel without prior experience in machine learning or data analytics may find some explanation outputs challenging to understand. Consequently, organizations may need to provide training programs, visualization tools, and user-friendly interfaces to ensure that generated explanations can be effectively utilized in operational decision-making processes.

Scalability also represents a significant challenge for widespread industrial deployment. Modern manufacturing facilities often operate thousands of interconnected machines generating continuous streams of sensor data. As the number of monitored assets increases, the computational requirements for generating individual explanations can grow substantially. Techniques such as SHAP and LIME may become difficult to apply efficiently across large-scale industrial systems without optimization strategies. Furthermore, maintaining consistent explanation quality across diverse machine types, operating conditions, and production environments remains an ongoing research challenge. Future developments in Explainable AI should therefore focus on improving scalability and reducing computational complexity while preserving explanation accuracy and interpretability.

4. CONCLUSION

This study demonstrates that the implementation of Explainable Artificial Intelligence (XAI) can significantly enhance the transparency and trustworthiness of predictive maintenance systems in industrial environments. The experimental results show that advanced machine learning and deep learning models, particularly XGBoost and LSTM, achieve high predictive performance in identifying potential equipment failures based on operational sensor data. Among the evaluated models, LSTM produced the highest prediction accuracy, confirming its effectiveness in capturing temporal degradation patterns within industrial equipment. Furthermore, the integration of SHAP and LIME successfully provided interpretable explanations of model predictions by identifying the contribution of critical variables such as vibration, temperature, runtime, and pressure. These explanations enabled maintenance engineers to better understand the reasoning behind failure predictions, resulting in increased confidence and trust in AI-assisted maintenance decisions. From a theoretical perspective, this study contributes to the growing body of knowledge on Explainable AI by extending its application to industrial predictive maintenance and proposing a framework for transparent maintenance analytics that combines predictive accuracy with interpretability. From a practical standpoint, the proposed framework can support improved maintenance planning, reduce unexpected equipment downtime, optimize resource allocation, and facilitate broader adoption of AI technologies in manufacturing operations. Nevertheless, this study has several limitations, including the use of a single-industry dataset, the relatively limited number of expert participants involved in validation, and the evaluation of only specific XAI techniques, namely SHAP and LIME. Therefore, future research should investigate real-time XAI implementation for continuous monitoring environments, explore the integration of federated learning to enhance data privacy and collaborative model development, incorporate Digital Twin technology to create more comprehensive predictive maintenance ecosystems, and examine the potential of Generative AI-based explanation systems capable of producing more intuitive and human-centered maintenance recommendations. Overall, the findings indicate that Explainable AI represents a promising approach for bridging the gap between predictive performance and human understanding, thereby enabling more transparent, reliable, and effective maintenance decision-making in the era of Industry 4.0.

REFERENCES

Ali, M., Ul-Hamid, A., Alhems, L. M., & Saeed, A. (2020). Review of common failures in heat exchangers–Part I: Mechanical and elevated temperature failures. *Engineering Failure Analysis*, *109*, 104396.

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, *3*(4), 966–989.

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.

Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. *IEEE Access*, *12*, 57574–57602.

Dieber, J., & Kirrane, S. (2020). Why model why? Assessing the strengths and limitations of LIME. *ArXiv Preprint ArXiv:2012.00093*.

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., & Weinberger, D. (2017). Accountability of AI under the law: The role of explanation. *ArXiv Preprint ArXiv:1711.01134*.

Dowling, A., O'Dwyer, J., & Adley, C. C. (2015). Lime in the limelight. *Journal of Cleaner Production*, *92*, 13–22.

Drakaki, M., Karnavas, Y. L., Tziafettas, I. A., Linardos, V., & Tzionas, P. (2022). Machine learning and deep learning based methods toward industry 4.0 predictive maintenance in induction motors: State of the art survey. *Journal of Industrial Engineering and Management (JIEM)*, *15*(1), 31–57.

Ebersbach, S. (2007). *Artificial intelligent system for integrated wear debris and vibration analysis in machine condition monitoring*. James Cook University.

Feng, D.-C., Wang, W.-J., Mangalathu, S., & Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *Journal of Structural Engineering*, *147*(11), 4021173.

Fritz, H. M., Blount, C., Sokoloski, R., Singleton, J., Fuggle, A., McAdoo, B. G., Moore, A., Grass, C., & Tate, B. (2007). Hurricane Katrina storm surge distribution and field observations on the Mississippi Barrier Islands. *Estuarine, Coastal and Shelf Science*, *74*(1–2), 12–20.

Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2021). Leveraging SHAP and LIME for enhanced explainability in AI-driven diagnostic systems. *International Journal of AI and ML*, *2*(3).

Keleko, A. T., Kamsu-Foguem, B., Ngouna, R. H., & Tongne, A. (2022). Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis. *AI and Ethics*, *2*(4), 553–577.

Lee, J. (1995). Machine performance monitoring and proactive maintenance in computer-integrated manufacturing: review and perspective. *International Journal of Computer Integrated Manufacturing*, *8*(5), 370–380.

Lipowski, E. E. (2008). Developing great research questions. *American Journal of Health-System Pharmacy*, *65*(17), 1667–1670.

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, *55*(5), 3503–3568.

Naqvi, S. A. A., Tennankore, K., Vinson, A., Roy, P. C., & Abidi, S. S. R. (2021). Predicting kidney graft survival using machine learning methods: prediction model development and feature significance analysis study. *Journal of Medical Internet Research*, *23*(8), e26843.

Nicolas, A., Messner, M. C., & Sham, T.-L. (2021). A method for predicting failure statistics for steady state elevated temperature structural components. *International Journal of Pressure Vessels and Piping*, *192*, 104363.

Paliari, I., Karanikola, A., & Kotsiantis, S. (2021). A comparison of the optimized LSTM, XGBOOST and ARIMA in Time Series forecasting. *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–7.

Plante, T., Nejadpak, A., & Yang, C. X. (2015). Faults detection and failures prediction using vibration analysis. *2015 IEEE AUTOTESTCON*, 227–231.

Qadir, A. (2025). THE FUTURE OF CYBER-PHYSICAL SYSTEMS: INTEGRATING AI, IOT, AND CLOUD FOR INDUSTRIAL REVOLUTION 5.0. *Frontiers in Multidisciplinary Studies*, *2*(01), 43–53.

Rahman, M., Cao, Y., Sun, X., Li, B., & Hao, Y. (2021). Deep pre-trained networks as a feature extractor with XGBoost to detect tuberculosis from chest X-ray. *Computers & Electrical Engineering*, *93*, 107252.

Raju, D., Su, X., Patrician, P. A., Loan, L. A., & McCarthy, M. S. (2015). Exploring factors associated with pressure ulcers: a data mining approach. *International Journal of Nursing Studies*, *52*(1), 102–111.

Walker, C. M., Agarwal, V., Lin, L., Hall, A. C., Hill, R. A., Mortenson, T. J., & Lybeck, N. J. (2023). *Explainable* *artificial intelligence technology for predictive maintenance*. Idaho National Laboratory (INL), Idaho Falls, ID (United States).

Wang, Z., Zhang, J., Jiang, Z., Mao, Z., Chang, K., & Wang, C. (2021). Quantitative misalignment detection method for diesel engine based on the average of shaft vibration and shaft shape characteristics. *Measurement*, *181*, 109527.

Xu, X., Lei, Y., & Li, Z. (2019). An incorrect data detection method for big data cleaning of machinery condition monitoring. *IEEE Transactions on Industrial Electronics*, *67*(3), 2326–2336.

Yechevskyi, A., Maksymova, S., & Sotnik, S. (2025). *Analysis of the data collection process about products at different stages of production*.

Zemmouchi-Ghomaria, L. (n.d.). *Explainable AI for predictive maintenance: A review and standardized evaluation framework*.

Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies*, *10*(8), 1168.