Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P6414-cd

Data-Driven Predictive Maintenance of Spare Parts for Smart Manufacturing

Kristian Have

Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, Denmark.
E-mail: kristian.have@hotmail.com

Parisa Niloofar

Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, Denmark. E-mail: parni@mmmi.sdu.dk

The rise of Artificial Intelligence (AI) and Industry 4.0 has led to a growing interest in predictive maintenance strategies, which offer benefits like reduced downtime, increased availability, and improved efficiency. This paper explores data-driven predictive maintenance of spare parts at a smart manufacturing company, based on AI methodologies to enhance efficiency and reduce downtime. The success of a smart manufacturing company is partly attributed to its advanced production facilities, particularly the precision injection moulds used for producing detailed and consistent parts. Injection moulding involves melting plastic and injecting it into a mould under high pressure. These moulds consist of many critical spare parts, such as gate bushes and inserts, which are prone to wear and tear due to intense pressures and temperatures. Failures in these small parts can halt production and affect efficiency. This study highlights the limitations of deep learning models due to insufficient data and the need for explainability and interpretability of models due to interaction with non-technical personnel. Also, results show that tree-based classification models, particularly Random Forest (RF) and XGBoost, perform best, with test accuracies of 69.59% for gate bushes and 69.23% for centre units. This investigation advances the manufacturing company's predictive maintenance capabilities, offering insights for future AI-driven maintenance optimization, leading to reduced costs, enhanced efficiency, and improved health and safety standards.

Keywords: Artificial intelligence, Data-driven modelling, Explainability, Predictive maintenance, Smart manufacturing, Spare parts.

1. Introduction

Maintenance can be defined as the upkeep of assets, such that they can be used to their full productive capacity. This involves the routine inspection, servicing, and repair of assets to prevent breakdown and prolong their lifespan. The key benefits to maintenance come in the form of preventing unexpected downtime, reducing long-term costs, and ensuring consistent quality of the produced goods.

Traditional approaches to maintenance involve either run-to-failure or preventive maintenance and have served the industry well for many years. Run-to-failure strategies such as corrective or reactive maintenance, which involves addressing issues as they arise, and preventive maintenance, which is based on scheduled servicing, have their own merits. However, they also come with inherent limitations. The potential for missed degradation, inaccurate lifespan estimates, and the resultant unexpected failures can lead to costly repairs, production disruptions, unplanned overtime costs, critical safety risks, and an overall decrease in operational stability Gulati and Smith (2009).

With the evolution of technology such as the emergence of artificial intelligence (AI), a new maintenance approach is emerging: predictive maintenance. This approach leverages historical data and Machine Learning (ML) to predict when a part is likely to fail, allowing for timely maintenance and replacement/repair. Additionally, Prognostics and Health Management (PHM) has become a crucial aspect of modern maintenance strategies. PHM integrates sensor data, anomaly detection, diagnostics, and prognostics to predict the Remaining Useful Life (RUL) of components and systems, thereby enabling proactive maintenance decisions. This project aims to explore the potential of AI, specifically explainable AI, in predictive maintenance, with a focus on the development and implementation of a predictive maintenance system on historical data of injection moulds in a smart manufacturing company.

2. Business Justification

At its essence, maintenance aims to minimize asset downtime and enhance production reliability. A well-designed maintenance strategy not only improves efficiency and production throughput, but also reduces worker overtime, enhances employee safety, improves product quality, and minimizes production delays (Mobley, 2002).

The current maintenance strategy in the smart manufacturing company (hereafter referred to as SMC) considered for this study consists of a mix of reactive and proactive approaches. In other words, the maintenance strategy combines run-to-failure strategies with scheduled maintenance based on either time or other relevant metrics. While these strategies are not inherently flawed, they may inadvertently lead to a decline in quality if an asset incurs damage or deteriorates prematurely. This can result not only in financial losses but also in a dip in customer satisfaction.

Furthermore, past studies referenced by Gulati and Smith (2009) indicate that a well-implemented predictive maintenance strategy can provide savings of 7-15% when compared to a standalone preventive approach. A study cited by Mobley (2002) demonstrated how the implementation of a computer-based maintenance system enabled a cable manufacturer to experience an increase of 50% in production capability without any corresponding increase in maintenance staff. This lead to a 60% increase in overall productivity.

As product quality remains a top priority for the SMC, any decrease in quality due to damaged or under-performing parts can result in significant financial losses and adversely affect customer satisfaction. The injection moulds, being custom-made and quite expensive, represent a substantial investment for the SMC. Damage to these moulds not only generates huge losses but also disrupts the production process, further adding to the financial and operational strain.

Implementing an AI solution for predictive maintenance can significantly mitigate these risks. By accurately forecasting when repairs or maintenance should be performed, the SMC can reduce maintenance costs and increase production efficiency. This approach ensures maintenance is performed exactly when needed, avoiding unnecessary downtime or premature interventions. Moreover, this predictive capability can potentially increase safety, reducing the chance of catastrophic failures that could endanger workers or halt production lines.

Additionally, this foresight leads to inventory reduction, as spare parts for the injection molds can be ordered on an as-needed basis rather than kept in large quantities. This not only saves on storage space but also optimizes cash flow and reduces inventory costs.

Therefore, considering the significant benefits that can be derived from an effective maintenance strategy, it is evident that implementing an AI solution for predictive maintenance would be highly advantageous.

3. Theoretical Basis and Related Work

This section will highlight important related work regarding the topic of predictive maintenance as well as provide a theoretical basis for understanding the ways in which such a problem can be solved. The findings in this section will be used to discover a methodology for solving the problem of predictive maintenance for the SMC.

3.1. *Theoretical Basis*

Over time, assets deteriorate due to a multitude of reasons such as reduced strength, increased stress, or design flaws. Determining the optimal timing for asset replacement or repair is done through what is known as a maintenance strategy. These strategies revolve around the balance of prolonging an asset's useful life while maximizing uptime. Choosing a run-to-failure approach may maximize an asset's useful life, i.e. the production capability, but it can result in additional costs stemming from potential damage and production downtime upon failure. Conversely, a preventive maintenance strategy, while enhancing uptime, often requires unnecessary repairs, leading to production losses and increased labor. With this in mind, it is the predictive maintenance strategy which strikes the best balance as it leads to maximized uptime and should only require repair just before the asset's useful life has been exerted (see Figure 1).

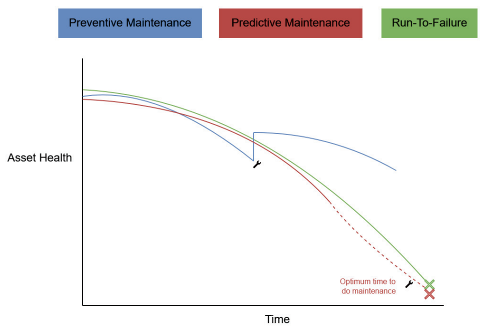

Fig. 1. Asset health over time for the preventive maintenance (marked in blue), predictive maintenance (marked in red), and run-to-failure (marked in green) strategies.

In predictive maintenance, two primary approaches are commonly employed; physical-based and data-driven approaches (Paolanti et al., 2018; Wen et al., 2022). Physical-based models utilize mathematical representations to present the degradation process. These models require a thorough understanding of degradation mechanisms, which often renders them impractical or ineffective in real-world applications due to system complexity or unclear degradation mechanisms. In contrast, data-driven models leverage ML techniques to detect patterns and anomalies within raw data, making them well-suited for predictive maintenance tasks, especially in the era of Industry 4.0, big data, and AI as a whole.

Data-driven models can be further categorized into two sub-types (Wen et al., 2022; Taşcı et al., 2023):

(1) Statistical-based models
(2) AI models

Statistical-based models typically monitor degradation trajectories in a probabilistic manner, while AI-based approaches utilize ML algorithms like RF, support vector machines, or deep learning to extract features and predict an asset's RUL. Recent advancements in AI have positioned data-driven approaches as the most promising method for predictive maintenance (Paolanti et al., 2018; Taşcı et al., 2023; Wen et al., 2022).

When employing AI models, the predictive maintenance problem can be approached in 3 ways:

(1) Binary classification
(2) Regression
(3) Multi-class classification

Binary classification predicts whether an asset will fail within a specific time frame or determines the asset's current state (e.g. healthy or unhealthy). Regression forecasts continuous values, such as calculating the RUL of an asset in days, kilometers, etc. Multi-class classification estimates multiple outcomes, such as predicting an asset's health status or likelihood of failure within different time intervals.

Selecting the appropriate problem classification depending on factors such as dataset type is crucial in the development of an effective predictive maintenance application.

3.2. *Performance Evaluation*

Performance evaluation is a critical aspect of assessing the effectiveness of predictive maintenance models. In the context of classification models, several common performance evaluation metrics are utilized to measure the model's accuracy and effectiveness. These metrics include accuracy, confusion matrices, precision, recall, and F1-score (Lee et al., 2019; Paolanti et al., 2018; Aslantas et al., 2022).

- **Accuracy:** Accuracy measures the proportion of correctly classified instances out of the total instances.
- **Confusion Matrices:** Confusion matrices provide a breakdown of correct and incorrect predictions made by a classification model.
- **Precision:** Precision quantifies the proportion of correctly predicted positive instances out of all instances predicted as positive.
- **Recall:** Recall calculates the proportion of

correctly predicted positive instances out of all actual positive instances.

- **F1-Score:** The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance.

For regression models, performance evaluation focuses on metrics that assess the accuracy and reliability of the model's predictions. Common performance evaluation metrics for regression models include Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2) (Wen et al., 2022; Ayvaz and Alpay, 2021).

- **MSE:** calculates the average of the squared differences between predicted and actual values.
- **RMSE:** measures the square root of the average of the squared differences between predicted and actual values.
- **MAE:** computes the average of the absolute differences between predicted and actual values.
- **MAPE:** represents the average percentage difference between predicted and actual values.
- **R^2:** quantifies the proportion of the variance in the dependent variable that is predictable by the independent variables.

These performance evaluation metrics (Gareth et al., 2013; Géron, 2022) play a crucial role in estimating the accuracy and reliability of predictive maintenance models.

3.3. *Related Work*

Predictive maintenance and PHM in general, has gathered significant attention in recent years using advanced ML techniques to forecast equipment failures. This section reviews several studies that have contributed to the field, showcasing a variety of approaches and methodologies.

Lee et al. (2019) focused on predictive maintenance for two critical components of machine tool systems: the cutting tool and spindle motor. Utilizing data from sensors installed on the equipment, they developed algorithms capable of predicting failure events categorized into normal, warning, and failure stages. Their approach employed support vector machines (SVM) and artificial neural networks (ANN), specifically recurrent neural networks (RNN) and convolutional neural networks (CNN), complemented by principal component analysis (PCA) for dimensionality reduction. The results were promising, with RNNs achieving an average accuracy of 93%, SVMs 87%, and CNNs 84%.

Paolanti et al. (2018) explored predictive maintenance within the context of Industry 4.0, focusing on induction motors monitored through internet of things (IoT) sensors. They also treated the problem as a multi-class classification task with four distinct classes, employing a RF model. This model demonstrated impressive performance, achieving an overall accuracy of 95% on a large dataset of 530,731 samples with 15 features collected in real-time.

Regarding prediction of RUL in PHM, an interesting approach was presented by Taşcı et al. (2023), who aimed to predict the RUL for equipment used in manufacturing consumer hygiene products. Without a predefined target variable for RUL, they calculated the time until the next stoppage thus creating their own target variable for RUL prediction. The study compared several ML models, including support vector regression (SVR), multilayer perceptron (MLP), RF, and extreme gradient boosting (XGBoost). The decision to exclude deep learning models from the comparison was strategic, driven by the relatively small size of the dataset and the desire to avoid the computational complexity associated with these models. One of the key aspects of this study was its comprehensive approach to dataset preparation. The researchers created multiple versions of the dataset to evaluate the impact of various preprocessing techniques on model performance. The standard dataset encompassed all collected data, serving as a baseline for comparison. In contrast, the stops-removed dataset excluded periods of production halt, based on the hypothesis that such intervals introduced noise rather than informative data. Another dataset employed an autoencoder to achieve dimensionality reduction. Additionally, the researchers experimented with clustering to segment the data into meaningful groups before model training, and they also explored the effects of removing highly correlated features to assess the impact on model accuracy.

Ayvaz and Alpay (2021) utilized real-time IoT sensor data for predictive maintenance in manufacturing production lines. Their dataset contained over 8.3 million rows and 101 features, analyzed through a regression task. Six ML algorithms were compared, with RF and XGBoost models outperforming others, achieving R-squared values of 0.982 and 0.979, respectively. PCA was also applied for dimensionality reduction.

Aslantas et al. (2022) investigated the prediction of RUL for plastic injection moulding machines using data from IoT sensors. They employed decision tree regression, XGBoost, and RF regression models across three datasets from different machines. The RF model exhibited the highest performance, with an average R-squared of 0.989 across the datasets, followed by XGBoost at 0.967.

These studies highlight the many applications and methodologies in predictive maintenance, underscoring the potential of ML techniques to significantly enhance maintenance strategies across various industries.

4. Methodology

4.1. *Data Description*

The dataset, comprises two primary components: notification data and master data. The notification data contains records of maintenance activities, including equipment numbers (injection mould IDs), spare parts maintained, shot counts at notification time, and maintenance dates. The master data provides static information about each injection mould, such as plastic type requirements, specifications, and production capabilities.

The dataset spans from April 2011 to March 2024, encompassing 64,020 maintenance notifications across 9,157 unique injection moulds and 2,553 unique spare parts. The master data contains 38 features, while the notification data comprises 15 features. Due to the dataset's complexity and the impracticality of predicting maintenance for all spare parts, the analysis focuses on gate bushes and centre units, identified through stakeholder consultation based on cost and delivery lead time considerations.

Data quality assessment revealed missing notification dates, which were excluded to maintain integrity. The notification frequency has shown a steady increase since the system's implementation in late 2012, with a notable decline in 2024 due to incomplete notifications.

4.2. *Problem Formulation*

The predictive maintenance challenge is approached through two distinct methods: regression and multi-class classification. The target variable for both approaches is defined as "shots until next maintenance" rather than calendar days, as shot count provides a more direct measure of mould usage and wear.

4.2.1. *Regression Approach*

The regression approach directly predicts the number of shots until the next maintenance event is required. This method provides the highest granularity for maintenance scheduling but faces challenges due to the significant variability in shot counts, which range from thousands to millions of shots between maintenance events.

4.2.2. *Classification Approach*

The classification approach segments the prediction into four time intervals based on production days: 0-49 days, 50-100 days, 101-300 days, and 301+ days. This categorization was informed by data quartiles and stakeholder requirements, particularly the need for 6-7 weeks advance notice for maintenance scheduling.

The final dataset comprises 981 instances for gate bushes and 605 for centre units. Gate bushes show a higher concentration of maintenance events in the shortest interval (0-49 days, n=370), while centre units display a more balanced distribution across intervals, with a slight skew toward longer maintenance periods (101-300 days, n=219). This distribution pattern suggests different maintenance characteristics between the two components, possibly reflecting their distinct roles and wear patterns in the injection moulding process.

4.3. *Model Selection*

Based on the literature review and dataset characteristics, two tree-based models were selected: RF and XGBoost. These models were chosen for their:

- Demonstrated effectiveness in handling complex, non-linear relationships
- Robustness to class imbalance
- Interpretability for non-technical stakeholders
- Suitability for the available dataset size

Deep learning approaches were excluded due to insufficient data volume and reduced interpretability for stakeholder communication.

5. Results and Discussion

The models were trained using 85% of the dataset with a temporal split for testing to prevent data leakage. A 5-fold cross-validation strategy was implemented to ensure result robustness. The dataset included 46 engineered features, derived from the original 15 features through various transformations and aggregations.

5.1. *Regression Performance*

The regression models attempted to predict the exact number of shots until next maintenance.

Table 1. Performance Metrics for Models with Feature Engineering

Metric	Gate Bushes		Centre Units	
	RF	XGB	RF	XGB
R^2 Score	0.152	0.087	0.182	0.012
RMSE (days)	74.14	75.96	82.06	88.17
MAE (days)	58.93	59.92	64.32	68.11

Note: Test metrics shown. RMSE and MAE in production days.

The regression results show limited predictive power, with R^2 scores below 0.2 for all models. This poor performance can be attributed to the high variability in the target variable, with shots until next maintenance ranging from 271 to 2,560,452 for gate bushes and 4,133 to 3,154,941 for centre units. This wide range and high standard deviation pose significant challenges for precise predictions.

5.2. *Classification Performance*

The classification approach categorized maintenance predictions into four intervals: 0-49 days, 50-100 days, 101-300 days, and 301+ days.

Table 2. Classification Accuracies for Models

Model	Mean CV Accuracy	Test Accuracy
Gate Bushes RF	0.507 ± 0.092	0.696
Centre Units RF	0.546 ± 0.068	0.692
Gate Bushes XGB	0.461 ± 0.105	0.615
Centre Units XGB	0.506 ± 0.081	0.637

Note: CV: Cross-validation results shown as mean ± standard deviation.

Table 3. Detailed Performance Metrics for RF Models

Time Interval	Gate Bushes		Centre Units	
	Precision	Recall	Precision	Recall
0-49 days	0.82	0.66	0.88	0.33
50-100 days	0.85	0.39	0.83	0.38
101-300 days	0.60	0.90	0.72	0.86
301+ days	0.58	0.88	0.56	1.00

Note: Results for RF models, which demonstrated superior performance.

The classification approach showed significantly better results than regression, with RF models achieving test accuracies of approximately 69% for both components. Particularly noteworthy is the high precision (0.82-0.88) for the critical 0-49 days category, indicating reliable predictions for immediate maintenance needs. This high precision is especially valuable given the cost implications of spare parts procurement.

The models demonstrate strong performance in identifying maintenance needs in the 101-300 days interval (recall $> 85\%$) but struggle with the 50-100 days interval (recall $\sim 38-39\%$).

This pattern suggests that the models are more reliable for short-term and long-term predictions than medium-term forecasts.

The detailed performance of the RF models can be visualized through confusion matrices (Figures 2 and 3).

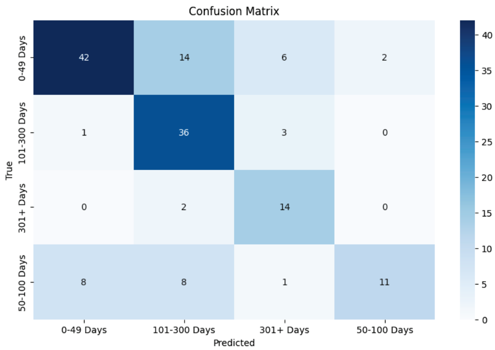

Fig. 2. Confusion Matrix for gate bushes RF model on test set

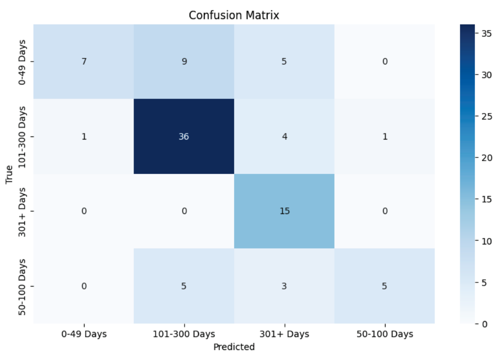

Fig. 3. Confusion Matrix for centre units RF model on test set

The confusion matrices reveal important patterns in model predictions. For gate bushes, the model correctly identifies 42 instances in the critical 0-49 days class, with most misclassifications falling into adjacent time intervals. This pattern of confusion between neighboring classes is expected and less problematic from a practical perspective. The centre units model shows similar patterns but with more pronounced difficulties in capturing all instances of immediate maintenance needs, often misclassifying them into longer time intervals.

Both models demonstrate stronger performance in the 101-300 days category, suggesting better capability in identifying medium-term maintenance needs. The varying performance across different time intervals likely reflects the underlying maintenance patterns of these components and the relative frequency of maintenance events in each time period.

Several factors contribute to the superior performance of the classification approach over regression:

- Simplified problem space through discretization of the highly variable target
- Reduced sensitivity to outliers and noise in the maintenance data
- Better alignment with practical maintenance planning needs

However, data quality remains a challenge, particularly due to:

- Unstructured communication between departments leading to duplicate entries
- Inclusion of cleaning activities alongside maintenance records
- Free-text input in the recording system causing information loss

Despite these limitations, the classification models, particularly RF, provide actionable insights for maintenance planning, with high precision in critical short-term predictions supporting cost-effective spare parts management.

6. Conclusion and Future Work

This study developed and evaluated AI models for predicting the maintenance of spare parts in SMC's injection moulds. While the regression approach showed limited success due to high data variability, the classification approach demonstrated promising results, achieving test accuracies of approximately 69% for both gate bushes and center units using RF models. Particularly noteworthy is the high precision (0.82-0.88) achieved for the critical 0-49 days category, indicating reliable predictions for immediate maintenance needs.

The complexity of the predictive maintenance challenge became evident throughout the study. Wide variability in spare part reliability, manual data entry leading to inconsistencies, lack of component state information, and diverse maintenance activities ranging from routine cleaning to complete replacement all contributed to the complexity of the problem. Despite these challenges, the results demonstrate the potential for implementing data-driven predictive maintenance at SMC.

Several immediate improvements could enhance the current approach. Data augmentation and sampling techniques could improve prediction accuracy for the critical 0-49 days class, while alternative binning strategies might optimize classification performance for different spare parts. For the regression approach, reducing prediction granularity to thousands of shots rather than individual shots could potentially improve model performance by reducing data noise and variability. Furthermore, standardizing maintenance logging procedures would significantly improve data quality and, consequently, model performance.

Available data type was limited to records of maintenance activities and static information about each injection mould. Hence, PHM could not be applied in its full capability, and looking ahead, there should be more focus on integrating additional data sources, particularly the IoT sensor data, which aligns with current industry best practices in predictive maintenance. Natural language processing techniques could be employed to extract valuable insights from free-text maintenance records, potentially revealing patterns and information not captured in the structured data. Advanced feature engineering methods for categorical data, such as embeddings or optimized one-hot encoding, could also enhance model performance.

While the current results may not match state-of-the-art performance levels, they provide a solid foundation for predictive maintenance capabilities. The classification approach, in particular, shows promise for practical implementation, potentially leading to improved maintenance scheduling and cost savings. To validate these benefits, comprehensive testing in production settings would be necessary to measure the system's impact on mean time between failures and overall maintenance efficiency.

References

Aslantas, G., M. Ozsarac, M. Rumelli, T. Alaygut, G. Bakirli, and D. Birant (2022). Prediction of remaining useful life for plastic injection molding machines using artificial intelligence methods. Last accessed on 09-04-2024.

Ayvaz, S. and K. Alpay (2021, July). Predictive maintenance system for production lines in manufacturing: A machine learning approach using iot data in real-time. Last accessed on 09-04-2024.

Gareth, J., W. Daniela, H. Trevor, and T. Robert (2013). *An Introduction to Statistical Learning*. New York, Ny Springer New York.

Gulati, R. and R. Smith (2009). *Maintenance and Reliability Best Practices*. New York, NY: Industrial Press.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.

Lee, W. J., H. Wu, H. Yun, H. Kim, M. B. Jun, and J. W. Sutherland (2019, January). Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. Last accessed on 09-04-2024.

Mobley, R. K. (2002). *An Introduction to Predictive Maintenance*. USA: Elsevier.

Paolanti, M., L. Romeo, A. Felicetti, A. Mancini, E. Frontoni, and J. Loncarski (2018, July). Machine learning approach for predictive maintenance in industry 4.0. Last accessed on 08-04-2024.

Taşcı, B., A. Omar, and S. Ayvaz (2023, October). Remaining useful lifetime prediction for predictive maintenance in manufacturing. Last accessed on 08-04-2024.

Wen, Y., M. F. Rahman, H. Xu, and T.-L. B. Tseng (2022, January). Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective. Last accessed on 08-04-2024.