Integrating Machine Learning-Based Remaining Useful Life Predictions with Cost-Optimal Block Replacement for Industrial Maintenance

Young-Suk Choo[1], Seung-Jun Shin [2]

[1]*Graduate School of Technology and Innovation Management, Hanyang University, Seoul, Republic of Korea*
smkyoungh@naver.com

[2]*School of Interdisciplinary Industrial Studies, Hanyang University, Seoul, Republic of Korea*
sjshin@hanyang.ac.kr

ABSTRACT

This study presents a preventive maintenance methodology to predict the remaining useful life (RUL) of mechanical systems and determine cost-effective replacement schedules. The approach integrates machine learning for RUL prediction, Weibull distribution for reliability analysis, and a block replacement model with minimal repair to optimize preventive maintenance. Many existing studies rarely incorporate RUL prediction results into determining optimal maintenance actions due to the high uncertainty in RUL prediction. To address this, the proposed methodology emphasizes not stopping at the prediction stage but integrating RUL predictions into actionable maintenance strategies to reduce uncertainty and improve applicability in industrial contexts. A case study using the open CMAPSS dataset demonstrates the feasibility of the approach. The value of this study lies in proposing a methodology that not only utilizes prediction-based proactive outcomes instead of predefined replacement intervals but also integrates them with subsequent maintenance strategies, providing practical and cost-effective solutions for industrial applications.

1. INTRODUCTION

Maintenance is important in mechanical machine systems because failures during operation can cause severe damage and breakdown. Accordingly, proper maintenance actions should be undertaken periodically and intermittently to avoid repetitive degradation that causes damage and safety problems in machine systems (Rebaiaia et al., 2017). This significance of maintenance has increased the development of optimal maintenance strategies to improve system reliability, prevent system failure, and reduce maintenance costs owing to system deterioration (Wang, 2002).

Maintenance strategies can be classified as corrective maintenance (CM), preventive maintenance (PM), condition-based maintenance (CBM), and predictive maintenance (PdM), based on the timing and actions of maintenance (Biggio and Kastanis, 2020). The CM fixes and recovers a system after failure. As PM is popular in the industry, it maintains a system at scheduled intervals before a failure occurs, and the intervals are determined by the system's expected age and replacement cycle. The CBM monitors and diagnoses the health of a system and performs maintenance actions when the system is detected to be unhealthy. PdM has recently gained the attention of industries because it can be implemented with advances in information and communication technology, such as artificial intelligence, data analytics, and internet of things, although it still remains a challenge. PdM is similar to CBM because of its dependency on the system's health; however, it differs in that proactive maintenance is pursued by predicting the future failure occurrence or remaining useful life (RUL) of a system.

In PdM, RUL prediction is a key technology for accurately forecasting the occurrence of system failures. RUL prediction directly affects the planning and execution of proactive maintenance actions (Aydemir and Acar, 2020). Good RUL prediction results lead to an increase in the overall equipment effectiveness and maintenance cost efficiency across the system lifecycle. Figure 1 illustrates the concept of RUL. The RUL represents the system's available time from the current time to the expected failure occurrence. RUL prediction relies primarily on the analysis of process data collected from the sensors attached to a system. Machine-monitoring data can provide meaningful information to represent a system's health and state, and

https://doi.org/10.36001/IJPHM.2025.v16i1.4242

they can be used broadly for anomaly detection, diagnostics and prognostics, and RUL prediction on a system.

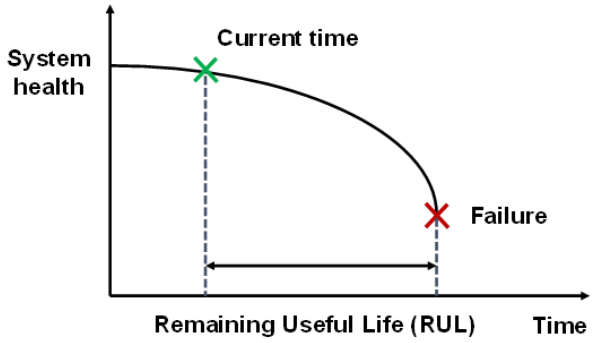


Figure 1. Concept of remaining useful life (RUL)

However, RUL prediction still remains a challenge because of the difficultly in RUL prediction at the current time due to the uncertainty of the next occurrence of a failure. Uncertainty induces the occurrence of expected or unexpected failures for various reasons such as degradation, deterioration, abrasion, friction, breakage, operation conditions, environment, or even unknown disturbances. Nevertheless, the RUL should be accurately predicted using process data. Well-predicted RULs aid in establishing optimal maintenance timing in PdM, as well as in determining optimal overhauls, repairs, and replacements to reduce maintenance costs.

The importance of RUL prediction in PdM has led to studies relevant to the development and validation of feasible methods and algorithms for performance measurement (Lei et al., 2018). They mostly developed RUL prediction models using machine learning techniques and compared the model performances to derive the best model for a target system. These studies have contributed valuable methods and algorithms to increase the prediction accuracy of RUL in a system. However, relevant studies rarely apply RUL prediction results to determine the optimal maintenance actions because they treat RUL prediction as a stand-alone study. Our approach directly addresses this gap by integrating RUL predictions with a cost-optimal block replacement model. This integration ensures that predictive insights are not merely theoretical but actionable, enabling industries to transition from reactive or static maintenance strategies to dynamic, data-driven decisions. Furthermore, by optimizing replacement schedules based on RUL predictions, this study offers a practical methodology for reducing operational costs and downtime. In addition, these studies did not consider diverse operation conditions on multiple systems because they targeted RUL prediction for a single system. To overcome these limitations, it is essential to develop a PdM-based method that can predict RULs for multiple systems and apply the RUL prediction results to determine the maintenance actions required to minimize maintenance costs. RUL prediction becomes more practical when applied to a holistic system comprising several systems and is integrated with the following maintenance strategy.

In contrast, PM is a maintenance strategy currently popular in industries. In PM, system or part replacements can occur at scheduled intervals or when necessary. This replacement mostly relies on an age replacement model or a block replacement model. The age replacement model replaces the system and/or its parts with predetermined replacement ages. Meanwhile, a system must be replaced immediately with a new one if failure occurs before the system's replacement age. This model is useful when the failure rate increases over time; otherwise, preventive replacement incurs lower expenses than replacement after failure (failure replacement). Meanwhile, the block replacement model replaces a system and/or its parts with predetermined schedules or intervals, regardless of their expected ages. Here, "block" represents a set of multiple components to be grouped as a system due to their homogeneity or similarity. This model has the advantage of convenient maintenance because it replaces a block as a whole at a given point in time, without considering failure replacement. However, this model may have a disadvantage because some components can be replaced with new ones, although their RULs remain sufficient. Accordingly, the block replacement model necessitates the determination of optimal replacement schedules to minimize the replacement of components that possess sufficient RULs and eventually reduces the wastage of cost incurred for maintenance. In this regard, RUL prediction can be a solution for optimizing replacement schedules because it enables the time-driven prediction of system failures. The block replacement model can be synergistically integrated with RUL prediction because the former is common in PM, whereas the latter is known to be a key technology in PdMs. This was the motivation for the present study. The proposed method is particularly relevant for industries where machinery reliability is critical and operational costs are high. Examples include the aviation industry, where optimizing the maintenance schedules of turbofan engines can prevent costly flight delays and enhance passenger safety; the manufacturing sector, where reducing unexpected equipment failures can increase production efficiency; and the energy industry, where predictive maintenance can extend the lifespan of turbines and reduce downtime in power generation systems. These contexts highlight the potential impact of integrating predictive analytics with preventive maintenance strategies.

Machine learning can underlie RUL prediction as a feasible data-driven approach. Machine learning has demonstrated good performance in PdM applications and has been broadly applied across manufacturing sectors (Thyago et al., 2019). The applications of machine learning in RUL prediction have increased owing to its good precision, applicability, and cost-effectiveness (Ferreira and

Gonçalves, 2022). Machine learning provides the capability for early prediction of failures by learning process data. In particular, machine learning has demonstrated excellent performance in failure prediction when sufficient data are trained.

To address the aforementioned challenges, this study explores the following research question: How can the integration of machine learning-based RUL predictions with cost-optimal block replacement models improve the effectiveness and efficiency of preventive maintenance strategies? This study proposes a preventive maintenance method to generate prediction models of the RUL and derive cost-optimal replacement times based on the RUL prediction model and the block replacement model with minimal repair. In the proposed method, (1) machine learning models are generated to predict the RULs of mechanical systems by training historical process data, (2) the Weibull distribution is fitted to identify the failure rate and reliability of the system by estimating its scale and shape parameters using the predicted RULs, and (3) the block replacement model with minimal repair is applied to determine the optimal preventive replacement times and costs based on the predicted RULs and the Weibull distribution. The proposed method aims to interconnect the PM and PdM through the integration of RUL prediction and block replacement models. In addition, the proposed method is targeted for a "block" system that combines multiple systems as a group due to their similarity. A case study was conducted to demonstrate the feasibility of the proposed method using an open and referential commercial modular aero-propulsion system simulation (CMAPSS) dataset for turbofan engine systems.

The remainder of this paper is organized as follows. Section 2 explains related work, Section 3 proposes the method, Section 4 describes the case study, and Section 5 discusses strategies and impacts of the proposed method. Section 6 concludes the paper.

2. RELATED WORK

2.1. RUL prediction using machine learning

Machine learning has been widely used to develop RUL prediction models based on historical processes and sensor data (Kraus and Feuerriegel, 2019). Models have been generated using typical data analytics procedures, including data extraction, feature extraction and classification, model building and training, and RUL prediction and evaluation (Ferreira and Gonçalves, 2022). Machine learning techniques include linear regression (LR), support vector regression (SVR), ridge regression, random forest (RF), extreme gradient boosting (XGBoost), and artificial neural networks (ANN). LR is a general technique used to model numerical relationships between independent and dependent variables. SVR is a regression version of a support vector machine, wherein a decision boundary is projected for classification and the margin width is adjusted to contain as much data as possible. Ridge regression comprises a residual sum of squares and a penalty term that can be differentiated and optimized as the sum of the squares of the parameters. RF and XGBoost are ensemble learning techniques that create several classifiers and combine their prediction results to obtain the final result. Ensemble learning aids in increasing accuracy and avoiding overfitting, which occurs in machine learning when a model is too attuned to the training data. An ANN constructs a network structure by mimicking neurons in layers, similar to the neural network in the human brain. The ANN outputs a predicted value by iteratively adjusting and converging the weights and bias of each neuron in a network with minimal error.

Shi et al. (2021) used XGBoost, RF, and SVM to predict the RULs of bearings. Tong et al. (2021) used XGBoost, AdaBoost, and gradient boost to predict the RULs of lithium-ion batteries and compared the performances of these models. Aydemir and Acar (2020) applied RF and long short-term memory (LSTM) to predict the RULs of turbofan engines using the CMAPSS dataset. Xue et al. (2020) predicted the RUL of a lithium-ion battery using an adaptive unscented Kalman filter and SVR. Park et al. (2020) used ridge regression to predict the RULs of bearings. Cailian and Chun (2020) predicted the RULs of lithium-ion batteries by applying a genetic algorithm to RF. Liu et al. (2021) used XGBoost, an RNN, and deep neural networks to predict the RULs of turbofan engines in the CMAPSS dataset. Wu et al. (2021) combined classification and regression using a CNN, SVR, and LSTM to predict the RUL of a CMAPSS turbofan engine. Chen et al. (2022) predicted the RULs of turbofan engines and optimized maintenance cost by balancing corrective and preventive maintenance costs. Lv et al. (2023) integrated a LSTM-based failure prediction model with a Deep Q-Network (DQN)-based maintenance decision-making system. Najdi et al. (2024) applied an adaptive residual attention LSTM for RUL prediction of rolling bearings. Zonta et al. (2022) predicted the RULs of industrial systems and optimized production schedules using multiple deep learning algorithms, such as CNN, LSTM, and RNN, in Microsoft Azure. Narayanan et al. (2024) applied RF and SVM to predict RULs for maintenance efficiency and effectiveness in manufacturing plants, energy production facilities, and transportation fleets. They validated the applicability of machine learning-based predictive maintenance using real data in various industrial domains.

Many previous studies utilized reference datasets such as CMAPSS turbofan engines, lithium-ion batteries, and bearings because of difficulties in data acquisition. RUL prediction essentially requires long-term data; thus, such referential data can be useful because they provide process data over the lifecycle of the system.

Furthermore, many of the previous studies have contributed to deriving the methods required to create accurate RUL prediction models by learning process data in target systems. However, these methods do not suggest the application of predicted RULs to determine the system's replacement actions during maintenance. This is because they typically apply more than two machine learning techniques for model generation and determine the best model through model verification and validation on a single system. They could not estimate the probability of failure for replacement time optimization in multiple systems. Replacement time should not be optimized at a single but at a holistic system because maintenance comes into action for a factory or shop floor where multiple machines exist. The rest of the studies do not fully address the integration of predictive insight into preventive maintenance, which pursues cost optimization. In contrast, our study proposes a quantitative method that integrates machine learning-driven RUL prediction with a block replacement model to decide optimal replacement times for cost optimization. Hence, the proposed method is expected to bridge the gap between industrial practice and technological advancement, ensuring a systematic and data-driven maintenance strategy.

2.2. Replacement models in preventive maintenance

2.2.1. Age replacement model

The Figure 2 illustrates the age replacement model. The age replacement model aims to replace the system immediately after the replacement time t_0. If system failure occurs before t_0, the system is replaced at that point. This model is useful when a system's failure rate increases over time, or when preventive replacement incurs lower expenses than failure replacement.

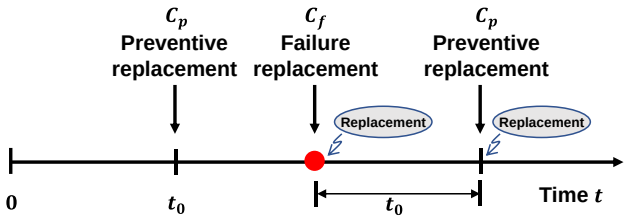


Figure 2. Concept of age replacement model

The system-replacement time can be either i) the time when a failure occurs before t_0 or ii) the time when replacement occurs as planned after t_0, given a failure time T, failure distribution function $F(t)$, and failure density function $f(t) = d/dtF(t)$. Equation (1) expresses the expected time $E(t)$. Equation (2) expresses the expected total cost $E(c)$, where C_p is the total cost of preventive replacement and C_f is the total cost of failure replacement. Generally, C_f is higher than C_p owing to the additional cost incurred by the failure of post-treatment and system recovery.

$$E(t) = \int_0^{t_0} tf(t)dt + t_0 \cdot \Pr(T \geq t_0) = \int_0^{t_0} (1 - F(t))dt \quad (1)$$

$$E(c) = C_p + (C_f - C_p)F(t_0) \quad (2)$$

In the age replacement model, an economic analysis should be conducted to determine whether system replacement is cost-efficient at a particular point in time. Equation (2) is inappropriate for economic analysis because $E(c)$ represents the total cost. For this purpose, an expected cost rate per unit time is preferable because costs can be measured and compared based on an identical unit. The expected cost rate per unit time corresponds to the unit cost obtained by dividing $E(c)$ by $E(t)$, i.e., an expected time consumed for complete recovery (Ross, 1980). Equation (3) expresses the expected cost rate per unit time, $C_A(t_0)$. Differentiating Equation (3) allows the determination of the optimal replacement time $t_0{}^*$. Equation (4) expresses the differentiation equation with respect to t_0, where $dC_A(t_0)/dt_0 = 0$ (Jin and Yamamoto, 2017). Here, $R(t)$ is a reliability function, where $R(t) = 1 - F(t)$.

$$C_A(t_0) = \frac{C_p + (C_f - C_p)F(t_0)}{\int_0^{t_0}(1 - F(t))dt} = \frac{C_p(1 - F(t_0)) + C_f F(t_0)}{\int_0^{t_0}(1 - F(t))dt} \quad (3)$$

$$h(t_0{}^*)\int_0^{t_0{}^*} R(t)dt - F(t_0{}^*) = \frac{C_p}{C_f - C_p} \quad (4)$$

In Equation (4), as the failure rate indicates $h(t_0)$ with regard to t_0, $h(t_0)$ equals to $f(t_0)/(1 - F(t_0))$. However, obtaining $t_0{}^*$ using Equation (4) is mathematically challenging. Alternatively, $t_0{}^*$ can be obtained using the Newton-Raphson method or reliability engineering software. $C_A(t_0{}^*)$ satisfies Equation (5) when $h(t_0{}^*)$ exhibits an increasing failure rate (IFR) (Zhao *et al.*, 2017).

$$C_A(t_0{}^*) = h(t_0{}^*)(C_f - C_p) \quad (5)$$

2.2.2. Block replacement model

Figure 3 depicts the block replacement model. Preventive replacements occur repetitively at constant intervals, denoted as t_p. This model is useful for maintenance planning and execution because maintenance actions are

performed regularly during a scheduled period or interval. In addition, this model can apply the concept of a block, as it targets multiple systems that can be grouped based on their similarity. The block replacement model involves the following three cases depending on the replacement timing and actions.

- Case I: A failed system is replaced immediately with a new system when failure occurs before t_p.
- Case II: A failed system is left unrepaired and replaced at the next t_p.
- Case III: A failed system is run with minimum repair and replaced at the next t_p. Minimum repair refers to the reuse of second-hand or remanufactured parts.

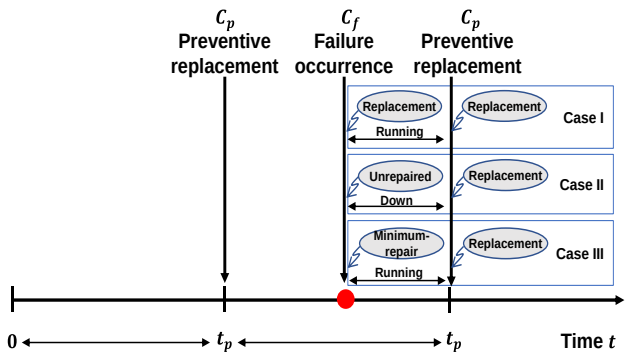


Figure 3. Concept of the block replacement model

In Case I, the replacement cost can be wasted because the failed system is replaced again when t_p is reached. Equation (6) expresses the expected cost rate per unit time, $C_{B1}(t_p)$, where $M(t_p)$ is a renewal function (Nakagawa, 1979). The differentiation of Equation (6) allows the determination of the optimal expected cost rate per unit time $C_{B1}(t_p{}^*)$. Equation (7) expresses the differentiation equation with respect to t_p, where $dC_{B1}(t_p)/dt_p = 0$. Here, $m(t_p)$ is a renewal density function, given $m(t_p) = d/dtM(t)$. Equation (8) expresses $C_{B1}(t_p{}^*)$ (Nakagawa, 1979).

$$C_{B1}(t_p) = \frac{C_p + C_f M(t_p)}{t_p} \tag{6}$$

$$t_p m(t_p) - \int_0^{t_p} m(t_p) = \frac{C_p}{C_f} \tag{7}$$

$$C_{B1}(t_p{}^*) = C_f m(t_p{}^*) \tag{8}$$

In Case II, the replacement cost can be beneficial; however, system availability may decrease because a failed system runs down until the next t_p. Equation (9) expresses the expected cost rate per unit time, $C_{B2}(t_p)$, where C_d is the downtime cost caused by the elapsed time from failure to the next t_p (Dong *et al.*, 2020). Similarly, Equation (10) expresses the differentiation equation of Equation (9), and Equation (11) expresses the optimal expected cost rate per unit time $C_{B2}(t_p{}^*)$ with respect to t_p, where $dC_{B2}(t_p)/dt_p = 0$ (Dong *et al.*, 2020).

$$C_{B2}(t_p) = \frac{C_p\left(1 - F(t_p)\right) + C_f F(t_p) + C_d \int_0^{t_p}(t_p - t)f(t)}{t_p} = \frac{C_p\left(1 - F(t_p)\right) + C_f F(t_p) + C_d \int_0^{t_p} F(t_p)dt}{t_p} \tag{9}$$

$$(C_p - C_f)\int_0^{t_p{}^*}[f(t_p{}^*) - f(t)]dt + C_d \int_0^{t_p{}^*}[F(t_p{}^*) - F(t)]dt = C_p \tag{10}$$

$$C_{B2}(t_p{}^*) = (C_p - C_f)f(t_p{}^*) + C_d F(t_p{}^*) \tag{11}$$

In Case III, the replacement cost can be reduced because a failed system runs without downtime when multiple systems operate simultaneously as they belong to a block. Case III was applied to the proposed method. This is because maintenance costs can be optimized for multiple systems grouped into blocks. To simplify the problem, the following assumptions are made: 1) a failed system is immediately repaired when a failure occurs, 2) the repair time is negligible, and 3) the failure rate of the system is stationary even after repair.

Equation (12) expresses the expected cost rate per unit time $C_{B3}(t_p)$, where $h(t)$ is a failure rate function, and C_k is the minimum repair cost (Nakagawa, 1979). Similarly, Equation (13) expresses the differentiation of Equation (12), and Equation (14) expresses the optimal expected cost rate per unit time $C_{B3}(t_p{}^*)$ (Rebaiaia *et al.*, 2017). In Equation (13), $C_{B3}(t_p{}^*)$ satisfies Equation (14) when $h(t)$ increases monotonically (Rebaiaia *et al.*, 2017).

$$C_{B3}(t_p) = \frac{C_p + C_k \int_0^{t_p} h(t)dt}{t_p} \tag{12}$$

$$t_p h(t_p) - \int_0^{t_p} h(t_p)dt = \frac{C_p}{C_k} \tag{13}$$

$$C_{B3}(t_p{}^*) = C_k h(t_p{}^*) \tag{14}$$

3. PROPOSED METHOD

The proposed method aims to optimize the preventive replacement time using RUL prediction models and a block replacement model to minimize the maintenance cost of a block system. Figure 4 illustrates the procedure of the

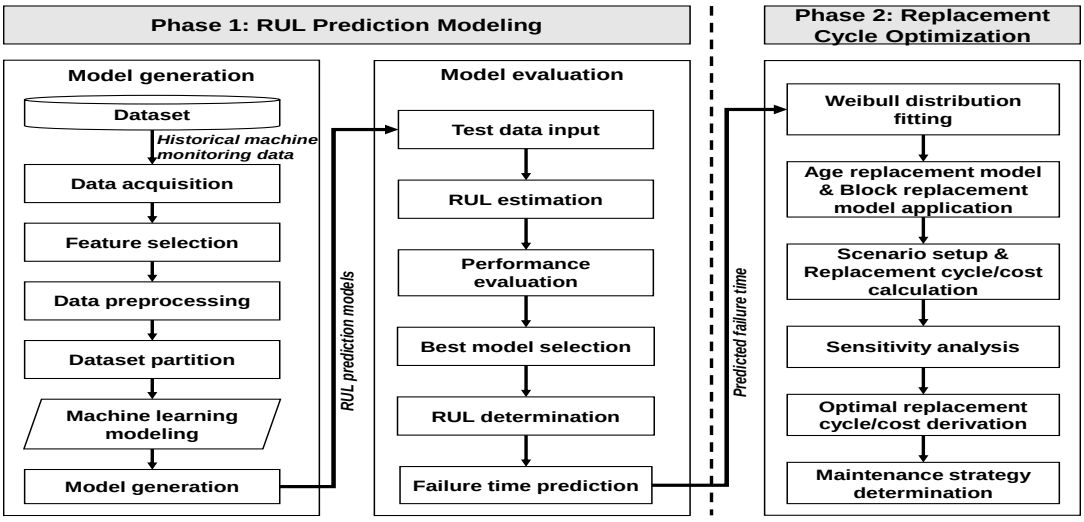


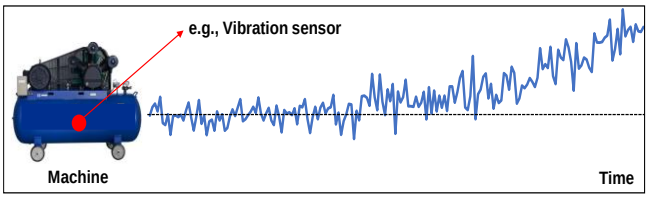
Figure 4. Procedure of the proposed method

proposed method. The method comprises two phases: RUL prediction modeling and replacement cycle optimization.

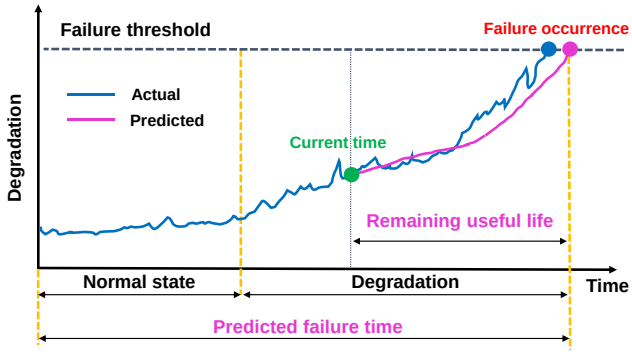
In the first phase, model generation is required to generate models that can predict the RULs of individual machines. Several types of models have been generated using machine learning techniques by training using historical process data. Model evaluation must be conducted to identify the best model among the models and then to use it for outputting a predicted system failure time. In the second phase, replacement cycle optimization is required to derive the optimal preventive replacement time and cost based on the predicted failure time and the two replacement models. The predicted system failure time is applied to specify scale parameter α and shape parameter β to be used for fitting a Weibull distribution, which enables the estimation of the failure rate and reliability of each system. These two parameters allow the calculation of preventive replacement times and their corresponding costs in certain scenarios. The block replacement model derives the optimal preventive replacement times to minimize maintenance costs. Finally, the best maintenance strategy is determined based on the optimal preventive replacement time and cost, considering the system operations and environments, including the system repairability, multiplicity, and performance.

3.1. RUL prediction modeling

The primary concept of RUL prediction is to infer an RUL from the process data that reflect system degradation. Some process data provide important information for RUL prediction because they represent the health state of a system by showing different, unusual, or gradually changing patterns from a normal pattern. Such dissimilar patterns can be detected and recognized using machine learning techniques by determining the causality between the process



(a) Historical process data



(b) System degradation

Figure 5. Concept of RUL prediction modeling

data and degradation features. Figure 5 illustrates the RUL prediction concept. As shown in Figure 5 (a), the sensor measures and acquires vibration data to monitor the health state of the system. Vibration data can act as a health indicator because they may exhibit dissimilarities with the normal pattern in a long-term trend along a system that continues to operate. In Figure 5 (b), the health indicator begins to increase gradually from the normal state, revealing

the appearance of continuous degradation in the system. Machine learning facilitates the detection and recognition of increasing patterns by training the historical vibration data recorded during the run-to-failure of the system. Failure occurrence can be set as the failure threshold for the next use. Machine learning then enables the prediction of the RUL at the current time because it can draw an increasing trend for the future without future data and determine the time point of threshold exceedance on degradation.

3.1.1. Model generation

The input of the model generation is historical process data of mechanical systems. The historical process data signify a dataset that records the health state and failure occurrence of a system at unit time during the system lifecycle. Process data are typically acquired from physical, chemical, or electrical sensors attached to a target system. Process data can also contain information on failure occurrences. Failure occurrences can be detected when sensor data are discontinued or exceptional at a certain time; otherwise, they can be acquired from an external data source or human input into a system. Historical process data require two premises for RUL prediction: the data should be fully recorded from a system's run-to-failure without missing data, and they should properly reflect the system's degradation with an unusual or dissimilar pattern that is different from the normal pattern.

Feature selection involves the identification of sensor attributes that directly affect model generation. Some features need to be excluded for efficient modeling if they are redundant, stationary, or irrelevant. Duplicate features can be found by correlation analysis with other features and need to be excluded from a dataset. Principal component analysis is a useful technique for reducing data dimensionality and estimating the distribution of multidimensional data. Data preprocessing includes data cleaning and data normalization. Data cleaning removes missing data, outliers, or noisy data that frequently occur in sensory data. Data normalization adjusts the original feature values to notionally common values within a specific range, such as the min-max and 0-to-1 scales. Feature selection and data preprocessing are important in machine learning because they affect the data quality. The performance of machine-learned models largely relies on the data quality (Woo *et al.*, 2018).

Dataset partition divides the entire dataset into training and testing datasets. A training dataset can be further segregated into training and validation datasets, if necessary. The training dataset is used to learn the data for generating the models, whereas the validation dataset is used to measure the learning error to check the correctness of the learning models. The testing dataset is used to measure the prediction error to verify the prediction accuracy when the models are applied to a target system. A ratio of 7:3 or 8:2 is generally used for the training or testing datasets. A ratio of 5:3:2 can be used when the dataset is separated into three sub-datasets.

Model generation is related to the generation of the RUL prediction models. In this study, the models were generated using six machine learning techniques: LR, SVR, RF, XGBoost, ridge regression, and ANN. These models were selected based on their established performance in predictive tasks, including RUL estimation, and their suitability for handling industrial datasets. Improved prediction accuracy has significant implications for operational cost reduction. When prediction models can estimate failure times more precisely, industries can identify optimal replacement schedules earlier during machine operation. This leads to fewer unexpected failures, less downtime, and reduced costs associated with corrective maintenance. For instance, XGBoost's ability to capture non-linear relationships makes it highly effective for modeling complex systems like turbofan engines, where sensor data exhibit intricate patterns. Similarly, ANN is well-suited for detecting subtle trends in high-dimensional data, which are critical for predicting system degradation accurately. By leveraging these strengths, this study demonstrates how precise RUL predictions can directly inform and improve maintenance strategies. They were selected because of their superior performance in numerical prediction problems in machine learning. Figure 6 (a) shows the structure of the XGBoost-based model. In XGBoost, a prediction model is generated by deploying multiple decision trees as separate predictors and combining their predictions to draw substantive conclusions. A decision tree creates a predictor to derive a predicted output, and the next tree creates another predictor successively to reduce the residual of the prior predictor, while the decision trees are dependently and consecutively combined. Each predictor is weighted to impose a high weight on incorrect answers and a low weight on correct answers. For example, the first decision tree creates $f_1(x_1)$ by training a historical process dataset, the following decision trees continuously create their $f_k(x_1)$, and the RUL is lastly predicted by summing the predicted values from all trees, as derived by $\sum_{k=1}^{N} f_k(x_i)$.

Figure 6 (b) shows the structure of the ANN-based model. An ANN predicts the value by constructing a network comprising neurons in multiple layers. The layers consist of an input layer to receive multiple input data, hidden layers to connect the input and output layers, and an output layer to produce a predicted value. An ANN uses activation functions to find the weights between prior and subsequent nodes as well as the bias in each node in a backpropagated or feedforward fashion. The network is trained repeatedly and eventually converges when the errors are minimized. Individual feature vectors (x_i), such as sensor values, are multiplied by their weights (w_m), and the results are then transmitted to a hidden layer. Similarly, neurons (j)

belonging to a hidden layer are multiplied by their weights (w_n), and their results are transmitted to an output layer. Finally, the output layer derives the predicted RUL (y_k). Note that a specific machine learning technique does not always outperform the remaining techniques. The model performance largely depends on the quantity and quality of the historical process data of the systems. Data quantity and quality are essentially case-specific owing to the diversity of data environments that are affected by the system performances, operations, and environments. Hence, RUL prediction models should be validated and verified to identify the best model for failure-time prediction.

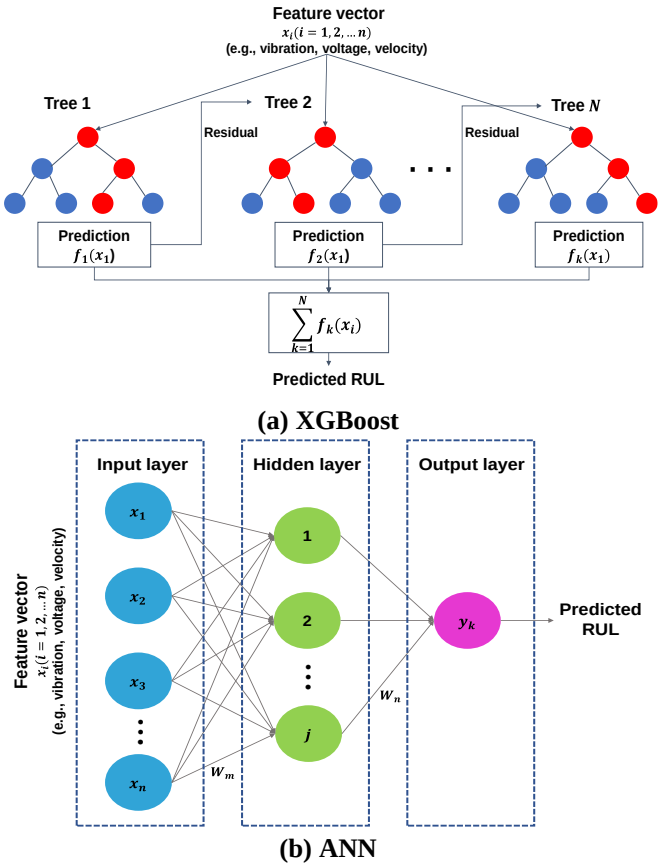


Figure 6. Structures of XGBoost and ANN-based RUL prediction models

3.1.2. Model evaluation

Model evaluation is performed to identify the best RUL prediction model among models by measuring and comparing model performances. The model performance can be evaluated using the difference between the actual and predicted RULs. The actual and predicted RULs were obtained from the real measurements and the best model, respectively. In this study, the article, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used as performance metrics. These two metrics are commonly used to measure the performance of RUL prediction models. Equation (15) expresses the RMSE, where Y is an actual RUL and $\hat{Y}$ is a predicted RUL. RMSE is suitable for representing model precision because it is less sensitive in terms of errors than the mean absolute error, which is sensitive to outliers. Equation (16) expresses the MAPE, as it allows a simple and intuitive interpretation in evaluating models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2}{n}} \tag{15}$$

$$MAPE(\%) = \frac{100}{n} \times \sum_{i=1}^{n} \frac{\left|Y_i - \widehat{Y_i}\right|}{Y_i} \tag{16}$$

The best RUL prediction model is selected and applied to predict the RUL of a system when it shows excellence in RMSE and MAPE; the lower the value, the better the prediction. Finally, the predicted failure time is derived as the output of the RUL prediction modeling phase. The predicted failure time should be used instead of the RUL in the next phase because a Weibull distribution can be fitted only when the failure time is specified. The predicted failure time was calculated as the time required to add the predicted RUL to the current time, as shown in Figure 5 (b).

3.2. Replacement cycle optimization

The replacement cycle optimization derives an optimal preventive replacement time and cost based on the predicted failure time and two replacement models. For problem simplification, the assumptions are as follows:1) the target system comprises multiple machines, 2) the age replacement model does not conduct repairs for preventive replacement and failure replacement,3) the block replacement model conducts minimum repairs (Case III), and 4) the predicted failure time follows a Weibull distribution for each machine. These assumptions are commonly used in reliability engineering to streamline complex systems (Rebaiaia and Ait-kadi, 2020). For example, the assumption of minimal repair is often applied in industries where components are refurbished for short-term use, such as the aviation sector, where remanufactured parts temporarily extend operational life. Similarly, the Weibull distribution is widely adopted due to its flexibility in modeling different failure rates and its practical feasibility in reliability analyses of mechanical systems.

The Weibull distribution is acceptable for predicting failure time because of its common use in reliability engineering. This distribution enables the flexible modeling of various types of failure rates in mechanical systems. Furthermore, it retains practical feasibility in reliability engineering based on the weakest-link principle, wherein the lifespan of a system depends on the weakest component that may initially cause failure. Equations (17), (18), and (19) express the

failure probability density function $f(t)$, failure probability distribution function $F(t)$, and reliability function $R(t)$ of the Weibull distribution, respectively. α denotes a scale parameter that affects the stretch-out of a distribution. β denotes a shape parameter that takes on various shapes of a distribution.

α and β can be estimated using the maximum likelihood estimation (MLE) among several techniques. MLE is a method of estimating a parameter θ to maximize the likelihood function $L = \prod_{i=1}^{n} f(x_1, \theta)$, wherein x_n is a data observation, $f(x)$ is a density function, and θ is an unknown parameter. θ is estimated by substituting the likelihood function with the Weibull probability density function and then differentiating the Weibull function with respect to α and β.

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{(\beta-1)} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{17}$$

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{18}$$

$$R(t) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \tag{19}$$

Figure 7 illustrates the types of failure rates in the Weibull distribution. The failure rate types are the decreasing failure rate (DFR), constant failure rate (CFR), and increasing failure rate (IFR), depending on the distribution of $h(t)$. The failure rate becomes DFR, CFR, and IFR if $\beta < 1$, $\beta = 1$, and $\beta > 1$, respectively. Determining α and β is critical to fit $f(t)$, $F(t)$, and $R(t)$ in a Weibull distribution. This is because α and β values are instantiated to the equations of the following replacement models so as to obtain the optimal replacement time and cost rate per unit time.

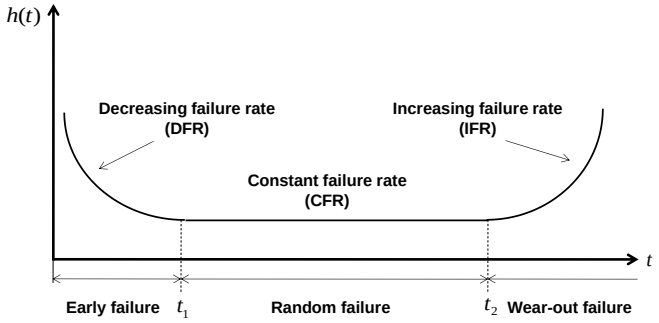


Figure 7. Types of failure rates

3.2.1. Age replacement model

It is necessary to determine $t_0{}^*$ and $C_A(t_0{}^*)$ once the predicted failure time is fitted to a Weibull distribution. Three functions, $f(t)$, $F(t)$, and $R(t)$, in the Weibull distribution were applied to the age replacement model. Equation (20) expresses $C_A(t_0)$, where Equations (18) and (19) are substituted into Equation (3). In case of IFR ($\beta > 1$), Equation (21) expresses a Weibull distribution $h(t_0{}^*)$, and Equation (22) expresses $C_A(t_0{}^*)$, where $h(t_0{}^*)$ is substituted to Equation (5).

$$C_A(t_0) = \frac{C_p\left(\exp\left[-\left(\frac{t_0}{\alpha}\right)^{\beta}\right]\right) + C_f\left(1 - \exp\left[-\left(\frac{t_0}{\alpha}\right)^{\beta}\right]\right)}{\int_0^{t_0}\left(\exp\left[-\left(\frac{t_0}{\alpha}\right)^{\beta}\right]\right)dt} \tag{20}$$

$$h(t_0{}^*) = \frac{f(t_0{}^*)}{1 - F(t_0{}^*)} = \frac{\beta}{\alpha}\left(\frac{t_0{}^*}{\alpha}\right)^{\beta-1} \tag{21}$$

$$C_A(t_0{}^*) = \left[\frac{\beta}{\alpha}\left(\frac{t_0{}^*}{\alpha}\right)^{\beta-1}\right]\left(C_f - C_p\right) \tag{22}$$

The cost efficiency should be analyzed when an age replacement model is applied (Rausand and Hoyland, 2004). In other words, preventive replacement must ensure high-cost efficiency to provide more benefits than failed replacements. Otherwise, the age replacement model may become obsolete. The cost efficiency can be measured as the ratio of preventive to failure replacement costs (Rausand and Hoyland, 2004). The preventive replacement cost is derived from Equation (3). The failure replacement cost can be obtained using the following process: the divergence of the replacement time to infinity ($t_0 \to \infty$) signifies that failure replacement occurs while preventive replacement does not occur. In Equation (3), when $t_0 \to \infty$, the numerator $F(\infty)$ becomes 1 and the denominator $E(c)$ becomes the mean time to failure (MTTF). MTTF refers to the case in which a system cannot be repaired. Equation (23) expresses $C_A(t_0 \to \infty)$. Equation (24) expresses MTTF of a Weibull distribution, where a gamma function is $\Gamma(\beta) = \int_0^{\infty} t^{\beta-1}e^{-t}dt$.

$$\begin{aligned} C_A(t_0) &= \frac{C_p + \left(C_f - C_p\right)F(t_0)}{\int_0^{t_0}\left(1 - F(t)\right)dt} \\ \to\ C_A(\infty) &= \frac{C_p + \left(C_f - C_p\right)F(\infty)}{\int_0^{\infty}\left(1 - F(t)\right)dt} \\ &= \frac{C_p + \left(C_f - C_p\right)}{MTTF} \end{aligned} \tag{23}$$

$$MTTF = \alpha \cdot \Gamma\left(1 + \frac{1}{\beta}\right) \tag{24}$$

Equation (25) expresses the cost efficiency $C_A(t_0)/C_A(\infty)$, where the numerator and denominator are divided individually by C_p, when $r = \frac{(C_f - C_p)}{C_p}$ (Rausand and Hoyland, 2004). The cost efficiency becomes higher as $C_A(t_0)/C_A(\infty)$ becomes lower. In other words, a change in r can affect the determination of preventive replacement

strategies owing to its influence on cost efficiency (Rausand and Hoyland, 2004). C_p is generally predetermined. C_f varies depending on the system environment, where an indirect cost is added to C_p due to the occurrence of a failure. Accordingly, changes in C_f and C_p affect the optimal replacement time. Therefore, a sensitivity analysis needs to be performed to determine the relevance of C_f and C_p to an optimal replacement time.

$$\frac{C_A(t_0)}{C_A(\infty)} = \frac{C_p + (C_f - C_p)F(t_0)}{\int_0^{t_0}(1-F(t))dt} \times \left(\frac{C_p + (C_f - C_p)}{MTTF}\right)^{-1} = \frac{1 + r \cdot F(t_0)}{\int_0^{t_0}(1-F(t))dt} \times \frac{MTTF}{1+r} \quad (25)$$

3.2.2. Minimum-repair block replacement model

In this study, we adopted a block replacement model with minimal repair, which corresponds to Case III in Section 2.2.2, as the block replacement model. The model immediately conducts minimum repair of a system in which the occurrence of a failure precedes a scheduled replacement interval and then replaces all machines belonging to the block at the scheduled replacement time.

Figure 8 shows the concept of the block replacement model with minimal repair. Because t_p indicates a scheduled preventive replacement time, t_p is generally determined as the near-average value of failure times on individual machines. Minimum repairs were conducted on the two machines when failure (i) occurred in machine 1 and failure (ii) occurred in machine 2 because the two failures occurred before t_p. Machine 1 can be re-run with minimum repair during failure (i) to t_p although failure (i) may break down machine 1. When current t_p is applied, the minimum repair range for failure (i) corresponds to the time from failure (i) to t_p.

When $t_p{}^*$ is applied, all machines belonging to a block are replaced by $t_p{}^*$. In this case, minimum repair is conducted only for failure (i) on machine 1 and not for failure (ii) on machine 2. The probability of failure (ii) dramatically decreases because all machines, including machine 2, are repaired and replaced at $t_p{}^*$ prior to failure (ii). The minimum repair range is the duration from the time point of failure (i) to $t_p{}^*$. In this regard, $t_p{}^*$ can positively affect failure prevention because it preemptively decreases the potential for failure in different machines. In contrast, $t_p{}^*$ may increase expenditure as a negative impact owing to more frequent replacements of machines. Hence, $t_p{}^*$ should be optimally determined using economic analysis as a tradeoff problem.

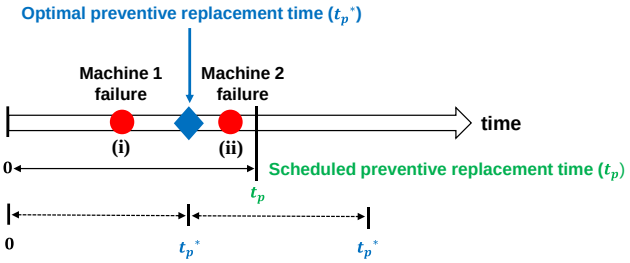


Figure 8. Minimum-repair block replacement model

Equation (26) expresses $C_{B3}(t_p)$ derived from Equation (12). This equation is valid when the system failure time follows a Weibull distribution. In Equation (12), the failure rate function is $h(t) = \frac{f(t)}{R(t)} = -\frac{d}{dt} lnR(t)$. Equation (27) expresses the integral of $h(t)$, where Equation (19) is substituted with $R(t)$. Equation (28) expresses the differential of Equation (26) with respect to t_p. This equation is set to zero when $t_p{}^*$ needs to be determined. Equation (29) expresses $t_p{}^*$, which is derived from Equation (28) (Rebaiaia and Ait-kadi, 2020).

Once C_k and C_p are determined, economic analysis is performed through a cost comparison between the current and optimal preventive replacements determined by t_p and $t_p{}^*$, respectively. The economic analysis leads to establish a good strategy for PM by validating the comparative cost advantage of the minimum repair block replacement model against the current preventive replacement.

$$C_{B3}(t_p) = \frac{C_p + C_k(\frac{t_p}{\alpha})^\beta}{t_p} \quad (26)$$

$$\int h(t)dt = -lnR(t) = -lne^{-(\frac{t_p}{\alpha})^\beta} = (\frac{t_p}{\alpha})^\beta \quad (27)$$

$$\frac{dC_{B3}(t_p)}{dt_p} = \frac{\beta C_k(\frac{t_p}{\alpha})^{\beta-1} - C_p + C_k(\frac{t_p}{\alpha})^\beta}{{t_p}^2} = 0 \quad (28)$$

$$t_p{}^* = \alpha\left(\frac{C_p}{C_k(\beta-1)}\right)^{\frac{1}{\beta}} \quad (29)$$

4. CASE STUDY

We hypothesize that the proposed method, which incorporated RUL prediction into block replacement scheduling, leads to cost efficiency in maintenance strategy. This hypothesis is induced from that the proposed method helps reduce maintenance costs and improve system reliability by using values applied to a preventive maintenance model tailored to the system's characteristics, instead of utilizing the arithmetic mean of predicted values for multiple machines. A case study was conducted using

the CMAPSS dataset to validate our hypothesis. The objectives were 1) to create RUL prediction models for turbofan engine systems and 2) to find an optimal replacement time and cost based on RUL prediction models and age and minimum-repair block replacement models. The CMAPSS dataset records operational parameters, sensor data, and RULs of turbofan engine systems, provided by the National Aeronautics and Space Administration (NASA) (Lei *et al*., 2018). While the dataset is widely used in RUL prediction studies for its completeness and reproducibility, its reliance on simulated data limits its direct applicability to real-world industrial environments, where data are often incomplete, noisy, or inconsistent. To address this, future studies could validate the proposed methodology using real-world data from industries such as aviation, energy, or manufacturing. Collaborations with industrial partners could provide access to operational and maintenance data, enabling a more comprehensive evaluation of the model's performance in real scenarios. In addition, the dataset is useful for generating and validating models because it contains the real RULs of all engines and is subdivided into a training or testing dataset. However, the CMAPSS dataset excludes cost data relevant to replacement and maintenance. Accordingly, certain costs must be inevitably assumed. In this study, the cost of a single turbofan engine was estimated at $200,000. This assumption was based on interviews with two experts in the rotating machinery domain. It is important to note that actual turbofan engine prices are typically not disclosed due to business and security policies. A sensitivity analysis is conducted to investigate the changes in the total replacement cost with regard to changes in C_p and C_f, in the case study. While this analysis provides valuable insights, future studies could expand on this by incorporating sensitivity analyses for additional factors, such as variations in system reliability, maintenance schedules, or operational environments. These analyses would allow the methodology to be tailored to a wider range of industries with diverse cost structures and maintenance requirements, enhancing its practical applicability.

The following tools were used for data analysis: Python for programming language and environment; scikit-learn library for LR, SVR, RF, and Ridge regression-based modeling; XGBoost library for XGBoost-based modeling; Keras library for ANN-based modeling; Reliability library for Weibull distribution fitting and replacement-cycle optimization.

4.1. Data specification

The CMAPSS dataset comprises four sub-datasets, as listed in Table 1. We used only two sub-datasets, named FD001 and FD003, because they provide complete datasets that contain twenty-one sensor data individually on one-hundred turbofan engines. Refer that Aydemir and Acar (2020) and Liu *et al*. (2021) used these two sub-datasets for the same reason. Figure 9 illustrates the main parts of the turbofan engine. Table 2 lists the sensor attributes for the CMAPSS dataset. Note that "use" indicates whether the attribute is used or not in the case study (will be explained in Section 4.2). The details of the CMAPSS dataset are as follows (Chao *et al*., 2021; Saxena *et al*., 2008):

- A turbofan engine is primarily composed of rotating parts, including a fan, low-pressure compressor (LPC), high-pressure compressor (HPC), low-pressure turbine (LPT), and high-pressure turbine (HPT).
- The sensor data include numerical values that were measured and recorded historically without data missing during the run-to-failure of each turbofan. The data may reflect the behavior of failures and deteriorations owing to continuous rotations in the parts.
- The sensor data are measured not on a timestamp, but on a cycle unit. Hence, "cycle" is used as the time-unit for RUL prediction.
- Real RULs were included, which served as ground truth. The ground truth enables the measurement of model performance by comparing the difference between the real and predicted RULs.
- Each sub-dataset was subdivided into training and

Item	Sub-dataset			
	FD001	FD002	FD003	FD004
Engines in training dataset (unit)	100	260	100	249
Engines in testing dataset (unit)	100	259	100	248
Min/Max in training dataset (cycle)	128/362	128/378	145/525	128/543
Min/Max in testing dataset (cycle)	31/303	21/367	38/475	19/486

Table 1. Overview of the CMAPSS dataset

testing datasets. This allowed dataset partitioning to be skipped

- The four sub-datasets comprising FD001–FD004 involved different fault modes. FD001 and FD002 incorporated one fault mode (HPC fault), whereas FD003 and FD004 incorporated two fault modes (HPC and fan fault).
- Each turbofan engine begins with different degrees of initial wear and manufacturing conditions; however, these are unknown.
- A turbofan engine operates in a normal state from the start but develops a fault at some point.

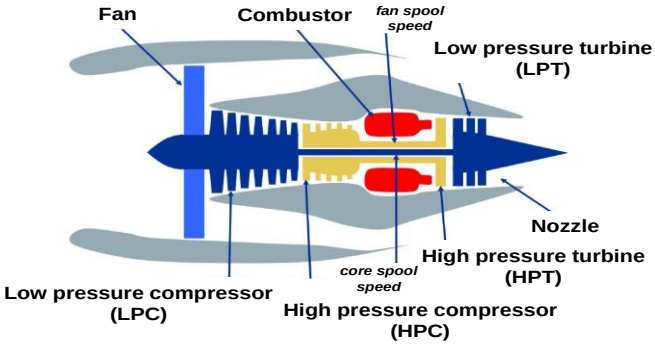


Figure 9. Scheme of a turbofan engine (re-edited from Frederick *et al.*, 2007)

Symbol	Attribute	Unit	Use
T2	Total temperature at fan inlet	°R	
T24	Total temperature at LPC outlet	°R	o
T30	Total temperature at HPC outlet	°R	o
T50	Total temperature at LPT outlet	°R	o
P2	Pressure at fan inlet	psia	
P15	Total pressure in bypass-duct	psia	
P30	Total pressure at HPC outlet	psia	o
Nf	Physical fan speed	rpm	o
Nc	Physical core speed	rpm	o
epr	Engine pressure ratio	-	
Ps30	Static pressure at HPC outlet	psia	o
phi	Ratio of fuel flow to Ps30	pps/psi	o
NRf	Corrected fan speed	rpm	o
NRc	Corrected core speed	rpm	o
BPR	Bypass ratio	-	o
farB	Burner fuel-air ratio	-	
htBleed	Bleed Enthalpy	-	o
Nf_dmd	Demanded fan speed	rpm	
PCNfR_dmd	Demanded corrected fan speed	rpm	
W31	HPT coolant bleed	lbm/s	o
W32	LPT coolant bleed	lbm/s	o

Table 2. Sensor attributes in CMAPSS

4.2. Result of RUL prediction

We generated and evaluated RUL prediction models by following the RUL prediction modeling procedure, as described in Section 3.1. The CMAPSS dataset files were acquired from their data sources. These files were converted from the original text (.txt) into comma-separated value (.csv) formats for convenient data analysis. In feature selection, we excluded operational parameters because they were not used for RUL prediction. In addition, we excluded the sensor attributes that are irrelevant to the RUL prediction in our feature analysis. Figure 10 shows the three types of sensor attributes. In Figure 10 (a) and (b), “T2” and “P15” sensor attributes need to be excluded because the former has a constant value, and the latter has categorical values. In Figure 10 (c), “Nf” sensor attribute needs to be selected as a feature because it may represent health states of engines while it shows increasing trends. In this regard, we include the 14 sensor attributes as the features for RUL prediction (marked ‘o’ in Table 2). We exclude “T2”, “P2”, “epr”, “farB”, “Nf_dmd”, “PCNfR_dmd”, and “P15”. During data preprocessing, we performed min-max normalization to scale all the values of each sensor attribute in the range of 0-1. Data cleaning was not conducted because missing data were not found, and noise data were unknown.

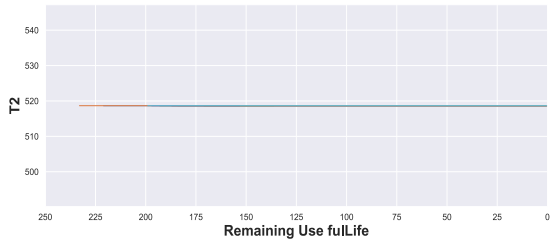
For machine learning modeling, we generated RUL prediction models using LR, SVR, RF, ridge regression (hereafter, Ridge), XGBoost (XGB), and ANN. We heuristically determined the hyperparameters of individual models. The hyperparameters were determined as follows: radial basis function as a kernel in SVR; 100 as the number of trees, and 6 as the maximum depth in RF; 500 as an alpha (a penalty term to represent a constraint) in Ridge, 100 as the number of trees, and 6 as the maximum depth in XGB; 14 as the number of nodes in an input layer, 10 as the number of hidden layers, and rectified linear unit as an activation function in the ANN.

We evaluated six RUL prediction models based on performance metrics, including RMSE and MAPE, as expressed in Equations (15) and (16), respectively. Tables 3 and 4 present the performance results for the FD001 and FD003 sub-datasets, respectively. As listed in Table 3, XGB achieved the lowest RMSE and MAPE, indicating the best prediction accuracy. As shown in Table 4, XGB had the lowest RMSE, although ANN had the lowest MAPE.

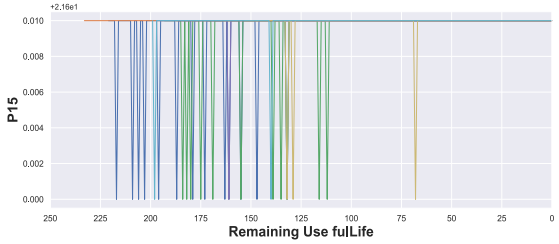
Additionally, the coefficient of determination (R^2) was calculated to further evaluate the predictive performance of the models. The results showed that XGB achieved the highest R^2 value of 0.815, followed by ANN (0.808) and RF (0.791). SVR also attained 0.785, while LR and Ridge both recorded 0.722. These results indicate that XGB demonstrated the strongest correlation between predicted and actual RUL values, reinforcing its superior predictive capability.

Consequently, the XGB-based model in FD001 showed the best prediction accuracy, with an RMSE of 17.86 and a MAPE of 11.04%. This result demonstrates the suitability of XGB for RUL prediction in systems with limited but meaningful features, such as turbofan engines. The practical implications of this accuracy are significant: industries implementing this model can expect improved reliability in failure forecasts, leading to fewer unexpected downtimes. For instance, in the aviation sector, this model could enable airlines to optimize engine replacement schedules, minimizing both the risk of in-flight failures and operational disruptions. Similarly, manufacturing plants could adopt this approach to align maintenance actions with production schedules, improving overall equipment effectiveness. We infer that XGB generates decision trees correctly to reduce errors at individual branches sequentially and effectively during the training of datasets, where the number of features is relatively small. Li *et al*. (2018) also demonstrated that a decision-tree-based approach performs well in numerical prediction, particularly when the number of features is small.

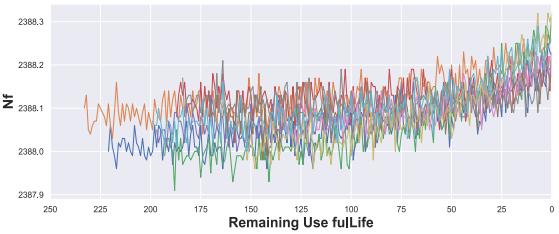
Hence, the XGB-based model in FD001 was used as the RUL prediction model in the case study because it outperformed the other models in both FD001 and FD003. Figure 11 shows a visualization of Alpha-Lambda-RMSE for different RUL prediction models in the FD001. The X-axis represents bias (Alpha), the Y-axis indicates error magnitude (Lambda), and the Z-axis shows overall prediction accuracy (RMSE). The red marker represents the actual RUL as a reference point. Models closer to this reference exhibit higher prediction accuracy. Among them, XGB and ANN perform best, showing minimal bias and lower RMSE values.



(a) Constant type



(b) Categorical type



(c) Changeable type

Figure 10. Types of sensor attributes

Performance	LR	SVR	RF	Ridge	XGB	ANN	Average
RMSE	21.90	19.25	18.97	21.90	17.86	18.20	19.72
MAPE (%)	32.28	16.75	19.58	32.30	11.04	12.27	20.73

Table 3. Prediction result in the FD001 sub-dataset

Performance	LR	SVR	RF	Ridge	XGB	ANN	Average
RMSE	23.88	22.90	22.05	23.88	20.69	22.42	22.64
MAPE (%)	30.47	19.97	21.46	30.64	16.56	16.42	22.59

Table 4. Prediction result in the FD003 sub-dataset

Number of engines	Average (cycle)	Standard deviation	Range (cycle)	Minimum (cycle)	Maximum (cycle)	Kurtosis	Skewness
100	206.97	39.31	173.09	140.98	314.07	-0.01	0.61

Table 5. Descriptive statistics of the predicted failure times

Interestingly, the six models in FD001 obtained better prediction accuracy, on average, than FD003. It is conjectured that the two fault modes in FD003 affected these features. Deteriorations can result in the occurrence of two fault modes; thus, clearly differentiating and extracting the features from the sensor data that compositely possess the characteristics of the two fault modes is challenging. This phenomenon can negatively affect the prediction accuracy of the models. Meanwhile, deterioration induced a singular fault mode in FD001, and its features were more distinguishable and extractable than those in FD003.

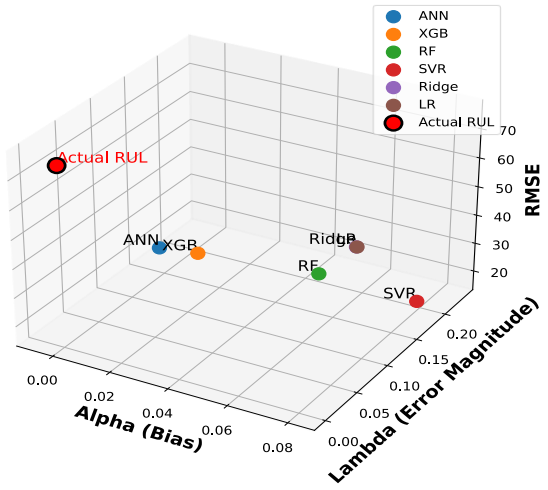


Figure 11. Alpha-Lambda-RMSE for different RUL prediction models in the FD001 sub-dataset

Now, we can calculate the predicted failure time by adding the predicted RUL to the current time on the turbofan engine, as shown in Figure 5 (b). Equation (30) expresses the predicted failure time. For example, the predicted failure time becomes 148.81 cycles for Engine 1 when its predicted RUL is 117.81 and its current cycle is 31. In this manner, we obtained all the predicted failure times for the one-hundred turbofan engines. Table 5 lists the descriptive statistics of the predicted failure times. In our prediction, the average of the predicted failure times is 206.97 cycles, the earliest time is 140.98 cycles on Engine 85, and the latest time is 314.07 cycles on Engine 49.

$$Predicted\ failure\ time\ (cycle) = Predicted\ RUL + Current\ cycle \quad (30)$$

4.3. Result of replacement time optimization

We optimized the preventive replacement times by following the replacement cycle optimization procedure, as suggested in Section 3.2. For this purpose, we derived a preventive replacement time to minimize the expected cost rate per unit time using the predicted failure times that were obtained in Section 4.2. The Weibull distribution was applied to obtain the failure times of the 100 engines. We estimated α and β using MLE, as explained in Section 3.2. Table 6 presents the α and β values estimated in the case study. Figure 12 presents a probability plot of the Weibull distribution for all engines. The failure times exhibited a desirable fit with the Weibull distribution, as the majority of the black dots were located within the confidence bounds (sky blue area), although the two tails of the distribution deviated slightly.

Dataset	Scale parameter (α)	Shape parameter (β)
100	223.46	5.41

Table 6. Parameters determined for Weibull distribution

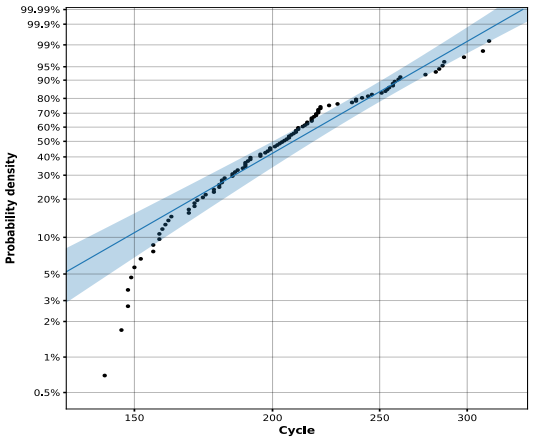


Figure 12. Weibull probability plot for predicted failure times on turbofan engines

4.3.1. Age replacement model

We applied the age replacement model to determine the optimal replacement times and their expected cost rate per

unit time for individual engines. For this, we predetermined and assumed C_p and C_f because the CMAPSS dataset excludes cost data, as described in Section 4. C_p was assumed to be \$200000 based on an engine's approximate price, whereas C_f varied from \$220000 to \$2200000. As described in Section 2.2.1, C_f depends on the operational environment and is higher than C_p owing to the additional indirect cost of post-treatment after failure. Accordingly, a sensitivity analysis was conducted to determine the change in the total replacement cost with respect to the changes in C_f and r under the given assumptions.

Table 7 presents the results of the age–replacement model. ${t_0}^*$ is determined using the Reliability library in Python because ${t_0}^*$ is difficult to obtain mathematically, as described in Section 2.2.1. Once ${t_0}^*$ has been determined, $C_A({t_0}^*)$ and $C_A(\infty)$ are calculated using Equation (22) and Equation (23), respectively. In addition, cost-effectiveness was analyzed to compare the benefits of the model between preventive replacement and failure replacement. The cost-effectiveness, i.e., $C_A({t_0}^*)/C_A(\infty)$, is obtained using Equation (25). Here, the MTTF is 206.11 cycles as it is derived from Equation (24).

As shown in Table 7, ${t_0}^*$ decreases, and $C_A({t_0}^*)$ increases as r increases. This result indicates that an increase in indirect costs is accompanied by faster replacement cycles and higher total maintenance costs. In other words, sustaining a small difference between C_p and C_f can be a good strategy in PM because it makes the engines operate longer as well as expends lower maintenance cost. In contrast, $C_A({t_0}^*)/C_A(\infty)$ decreases as r increases. A lower $C_A({t_0}^*)/C_A(\infty)$ implies superior cost-effectiveness of preventive replacement to failure replacement, as $\frac{C_A({t_0}^*)}{C_A(\infty)} = 1$ stands for no difference between the two replacements.

C_p	C_f	r	${t_0}^*$	$C_A({t_0}^*)$	$C_A(\infty)$	$C_A({t_0}^*)/C_A(\infty)$
200	220	0.1	267.24	1.066	1.067	0.999
200	240	0.2	232.16	1.147	1.164	0.985
200	260	0.3	214.42	1.210	1.261	0.959
200	280	0.4	202.77	1.263	1.358	0.929
200	300	0.5	194.27	1.307	1.456	0.898
200	320	0.6	187.64	1.346	1.553	0.867
200	340	0.7	182.29	1.380	1.650	0.837
200	360	0.8	177.73	1.411	1.747	0.808
200	380	0.9	173.78	1.439	1.844	0.781
200	400	1	170.44	1.465	1.941	0.755
200	1200	5	126.25	1.951	5.822	0.335
200	2200	10	110.99	2.214	10.674	0.207

Table 7. Result of age replacement model
($C_p, C_f, C_A({t_0}^*), C_A(\infty)$: thousands \$, ${t_0}^*$: cycles)

4.3.2. Minimum-repair block replacement model

We applied the block replacement model to determine the optimal replacement times and their expected cost rate per unit time on a block comprising 100 engines. We adopted the minimum-repair block replacement model, which is defined as Case III in Section 2.2.2. This model allows failed engines to be recovered with minimal repair by reusing second-hand or remanufactured engines; thus, the model is useful when it is difficult to detach and repair broken parts in a system, such as an engine system. The model has an advantage in maintenance compared with Cases I and II. However, it may force inventory management of second-hand systems and change management of C_k.

Similar to the age replacement model, we predetermined and assumed C_p and C_k for individual engines. C_k was assumed to be \$120000, considering a 40% discount on the replacement cost of a new engine (C_p=200000). Here, t_p was set to 205 cycles as the rounded-down value of the average predicted failure time (206.97 cycles in Table 5). As shown below, $C_{B3}(t_p)$ was obtained using Equation (26), given t_p=205, C_k=120000, and C_p=200000. The values of t_p and $C_{B3}(t_p)$ act as ground values for cost comparison in the following optimization.

${t_p}^*$ was determined using Equation (29), and $C_{B3}({t_p}^*)$ was obtained using Equation (26). Here, the difference between t_p=205 and ${t_p}^*$=186.68 is revealed with 18.32 cycles. The difference in maintenance costs between $C_{B3}(t_p)$ and $C_{B3}({t_p}^*)$ was \$29. Expectedly, these reduced ${t_p}^*$, and $C_{B3}({t_p}^*)$ contributed to saving 29\$/cycle on each engine as the average maintenance cost per unit time.

$$C_{B3}(205) = \frac{C_p + C_k\left(\frac{t_p}{\alpha}\right)^{\beta}}{t_p}$$

$$= \frac{200{,}000 + 120{,}000\left(\frac{205}{223.46}\right)^{5.41}}{205} = \$1343$$

$${t_p}^* = \alpha\left(\frac{C_p}{C_k(\beta-1)}\right)^{\frac{1}{\beta}} = 223.46\left(\frac{200{,}000}{120{,}000(5.41-1)}\right)^{\frac{1}{5.41}}$$

$$= 186.68\ cycles$$

$$C_{B3}(186.68) = \frac{C_p + C_k\left(\frac{t_p}{\alpha}\right)^{\beta}}{t_p}$$

$$= \frac{200{,}000 + 120{,}000\left(\frac{186.68}{223.46}\right)^{5.41}}{186.68} = \$1314$$

We conducted a sensitivity analysis to investigate the changes in ${t_p}^*$ and $C_{B3}({t_p}^*)$ with respect to the changes in

C_k and C_p. Table 8 presents the sensitivity analysis results. $t_p{}^*$ decreased, whereas $C_{B3}(t_p{}^*)$ continually increased with increasing C_k. In addition, $t_p{}^*$ and $C_{B3}(t_p{}^*)$ increased as C_p increase. From the PM perspective, this result implies that a decrease in C_k is necessary to reduce $C_{B3}(t_p{}^*)$ although it induces an increase in $t_p{}^*$. In addition, C_p needs to be decreased to reduce maintenance costs when C_k remains the same. Moreover, decreasing C_k and C_p leads to a reduction in the total maintenance cost when the replacement time is fixed.

Table 9 lists the changes in $f(t_p{}^*)$, $R(t_p{}^*)$, and $h(t_p{}^*)$ in the Weibull distribution with respect to the change in $t_p{}^*$, given a constant C_p =200000. $f(t_p{}^*)$ and $R(t_p{}^*)$ were obtained using Equations (17) and (19), respectively. $h(t_p{}^*)$ was derived from $f(t_p{}^*)/\ R(t_p{}^*)$. As $t_p{}^*$ decreases, $R(t_p{}^*)$ increases and $h(t_p{}^*)$ decreases. This indicates that $t_p{}^*$ is proportional to $h(t_p{}^*)$ and inversely proportional to $R(t_p{}^*)$. In other words, a short preventive replacement time with minimal repair can positively affect the robustness of a system owing to its contribution to increasing the system's reliability and decreasing its failure rate. From the PM perspective, determining $t_p{}^*$ makes the system reliability manageable because of its quantitative influence on the failure rate and reliability. In the meantime, it turns out that the failure rate follows IFR because β(=5.41) is larger than 1 in the Weibull distribution. This is also reasonable in terms of physical aspects because each engine is gradually worn out as a typical mechanical system. Therefore, $C_{B3}(t_p{}^*)$ satisfies Equation (14) and can be obtained using Equation (14). For example, $C_{B3}(t_p{}^*)$ is \$1314 when $t_p{}^*$ is 186.68 cycles, as follows:

$$C_{B3}(186.68) = C_k \times h(186.68) = 120{,}000 \times 0.01095 = 1314\$$$

4.3.3. Cost analysis in sub-block segmentation

In Section 4.3.2, the optimization results indicate that short replacement times result in a reduction in the minimum number of repair engines. However, wastage of expenses can occur and, in turn, have a negative impact on the total operating cost if all engines must be replaced simultaneously; nevertheless, some engines possess sufficient RULs. Conversely, long replacement times positively affect the preventive maintenance cost; however, they may induce maintenance inefficiency, as more failed engines should be repaired immediately before the replacement time arrives. Such a trade-off relationship can occur in the minimum-repair block replacement model, wherein the "block" is treated as a group of maintenance. This trade-off results in an optimization problem. We apply block segmentation to overcome this, which partitions an entire block into several sub-blocks. Block segmentation enables the derivation of $t_p{}^*$ and $C_{B3}(t_p{}^*)$ diversely and flexibly on sub-blocks, thereby achieving cost optimization and improving the maintenance efficiency.

We segment the one-hundred engines into three sub-blocks. The predicted failure cycles of the individual engines are used as criteria for segmentation. This segmentation approach offers several advantages:

- Precision in maintenance strategies: Tailors maintenance schedules to the specific RUL characteristics of each engine, ensuring efficient allocation of resources.
- Cost efficiency: Minimizes unnecessary expenses by avoiding excessive repairs for engines with sufficient RUL.
- Reduced operational downtime: Limits disruptions caused by simultaneous failures or repairs, enhancing system reliability.
- Flexibility in scheduling: Provides adaptability to varying operational demands, improving the resilience of maintenance plans.

Thus, a short replacement cycle needs to be applied to engines in which RULs remain short because it can reduce maintenance costs by decreasing the number of minimum-repair engines. In contrast, a long replacement cycle needs to be applied to engines that possess sufficient RULs to prevent dissipative replacement expenses. In this regard, $t_p{}^*$ (=186.68) is set as the lower threshold, and the top 20% of the predicted failure cycles is set as the upper threshold for segmentation into three sub-blocks. Table 10 presents the descriptive statistics for the three sub-blocks. Sub-block 1 contains engines that belong under the lower threshold, Sub-block 2 contains engines that belong over the lower threshold and under the upper threshold, and Sub-block 3 contains engines that belong over the upper threshold.

Table 11 presents the optimization results of the sub-block segmentation. In each sub-block, α and β are estimated using MLE. $t_p{}^*$ was derived using Equation (29), and $C_{B3}(t_p{}^*)$ was calculated using Equation (26), given C_p=200000 and C_k=120000. In Sub-block 1, the shortened $t_p{}^*$ requires five engines to undergo minimum repairs, resulting in an increase (+119 \$/cycle) in cost to 1433 \$/cycle compared to the original cost of 1314 \$/cycle. This demonstrates that while frequent replacements can improve reliability, they may incur higher costs due to the increased repair frequency. Meanwhile, in Sub-block 2, the extension of $t_p{}^*$ to 188.23 cycles reduces the number of engines requiring minimum repairs to one, reflecting a balance between extending operational cycles and controlling maintenance costs. The result shows a decrease of 183

$/cycle in cost to 1250 $/cycle compared to the original cost (1314 $/cycle). In Sub-block 3, no engines require minimum repairs under the optimized strategy, resulting in the most significant cost reduction of 402 $ per cycle, highlighting the efficiency of applying longer replacement cycles for engines with extended RULs. The analysis suggests that maintaining the original $t_p{}^*$ of 186.68 cycles for engines in Sub-block 1 is cost-effective, as it minimizes unnecessary expenses while maintaining repair efficiency. In contrast, adjusting $t_p{}^*$ for Sub-blocks 2 and 3 achieves significant cost savings of up to $183 and $402 per cycle, respectively, demonstrating the benefits of segmentation in optimizing maintenance strategies for diverse engine conditions.

Consequently, these optimization results demonstrate the monetary benefit from the sub-block segmentation. The total maintenance cost was reduced to $23,106,745 (derived from optimized $C_{B3}(t_p{}^*)$ in three sub-blocks, calculated as (166.33 cycles × $1314 × 33) + (209.33 cycles × $1131 × 47) + (268.48 cycles × $912 × 20)), compared to $27,195,858 (derived from fixed $C_{B3}(t_p{}^*)$, calculated as (206.97 cycles × $1314 × 100)), under the conventional method. Algorithm 1 expresses the pseudo-code of the cost optimization procedure based on the sub-block segmentation. The inputs of the algorithm are predicted failure cycles (derived from the RUL prediction model), optimal replacement time (derived from the minimum-repair block replacement model), Weibull parameters (derived from MLE), cost parameters (derived from assumption). The outputs of the algorithm are the optimal replacement time and its corresponding maintenance cost.

Algorithm 1: Sub-block Segmentation and Cost Optimization

Input: Predicted failure cycles {T_i}, Optimal replacement time $t_p{}^*$ = 186.68, Weibull parameters (α, β), Cost parameters (C_p, C_f, C_k)

Output: Optimal replacement time $\boldsymbol{t_p}^*$ and its maintenance cost $C_{opt} = C_{B3}(t_p{}^*)$

1: **Sort {T_i} in ascending order**

2: **Determine segmentation thresholds:**

$t_p{}^*$_U ← 80th percentile of {T_ i } ▷ Upper threshold (Top 20% of engines)

3: **Segment engines into three sub-blocks:**

Sub-block 1: {T_i | T_$i \le \boldsymbol{t_p}^*$} ▷ Engines with lower endurance

Sub-block 2: {T_i | $\boldsymbol{t_p}^*$ < T_$i \le t_p{}^*$_U} ▷ Medium endurance engines

Sub-block 3: {T_i | T_$i > t_p{}^*$_U} ▷ High endurance engines (Top 20%)

4: For each sub-block $j \in \{1,2,3\}$ do:

5: Fit Weibull distribution using (α, β) for sub-block j

6: Compute optimal replacement time $\boldsymbol{t_p}^*\_j$ using:

Solve Equation (28): $dC_{B3}(t_p)/\, dt_p = 0$ for $\boldsymbol{t_p}^*\_j$, Equation (29) ▷ $\boldsymbol{t_p}^*\_j$

7: Compute expected maintenance cost C_{opt}, j using:

Apply Equation (26)

8: **End for**

9: Compare C_{opt} 1, C_{opt} 2, C_{opt} 3 and select $\boldsymbol{t_p}^*$ with minimum C_{opt}

Return: Optimal replacement time $\boldsymbol{t_p}^*$ and its maintenance cost C_{opt}

C_k	$C_p = 180000$		$C_p = 200000$		$C_p = 220000$	
	$t_p{}^*$	$C_{B3}(t_p{}^*)$	$t_p{}^*$	$C_{B3}(t_p{}^*)$	$t_p{}^*$	$C_{B3}(t_p{}^*)$
80,000	197.32	1119	201.20	1219	204.78	1318
100,000	189.35	1166	193.07	1271	196.51	1373
120,000	183.08	1206	186.68	1314	189.99	1421
140,000	177.93	1241	181.43	1352	184.66	1462
160,000	173.59	1272	177.01	1386	180.15	1498
180,000	169.86	1300	173.20	1417	176.27	1531

Table 8. Result of sensitivity analysis ($C_k, C_p, C_{B3}(t_p{}^*)$: $, $t_p{}^*$: cycle)

$t_p{}^*$	C_k	$f(t_p{}^*)$	$R(t_p{}^*)$	$h(t_p{}^*)$
201.20	80,000	0.00865	0.56732	0.01524
193.07	100,000	0.00807	0.63542	0.01271
186.68	120,000	0.00751	0.68525	0.01095
181.43	140,000	0.00699	0.72330	0.00966
177.01	160,000	0.00653	0.75317	0.00866
173.20	180,000	0.00612	0.77726	0.00787

Table 9. Result of Weibull distribution functions at $C_p = 200000$ (C_k, C_p : \$, $t_p{}^*$: cycle)

Sub-block	Number of engines	Average (cycle)	Range (cycle)	Minimum (cycle)	Maximum (cycle)	Standard deviation
1	33	166.33	44.97	140.98	185.94	13.31
2	47	209.33	50.44	187.95	238.39	13.15
3	20	268.48	73.18	240.88	314.07	21.70

Table 10. Descriptive statistics of block segmentation

Sub-block	α	β	$t_p{}^*$	$C_{B3}(t_p{}^*)$	Cost comparison	Optimization selection
1	172.35	15.00	149.55	1433	+ 119/per cycle	X
2	215.51	16.45	188.23	1131	- 183/per cycle	O
3	278.73	12.36	238.64	912	- 402/per cycle	O

Table 11. Optimization in sub-block segmentation ($\boldsymbol{t_p{}^*}$: cycle, $\boldsymbol{C_{B3}(t_p{}^*)}$, comparison: \$)

5. DISCUSSION

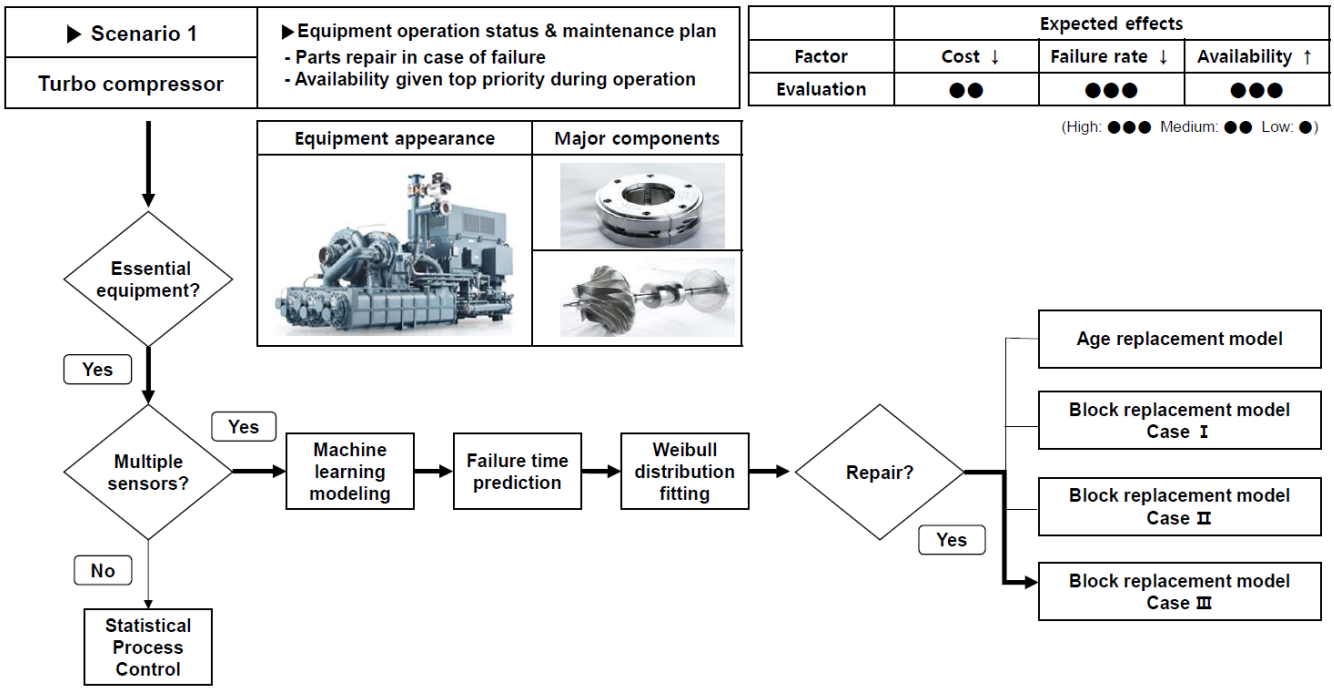
5.1. Industrial application strategies

In this section, industrial application strategies are proposed based on the suggested method and case study. To illustrate these strategies, scenarios are provided demonstrating RUL prediction methods and preventive replacement model applications for equipment and components commonly used in industrial settings.

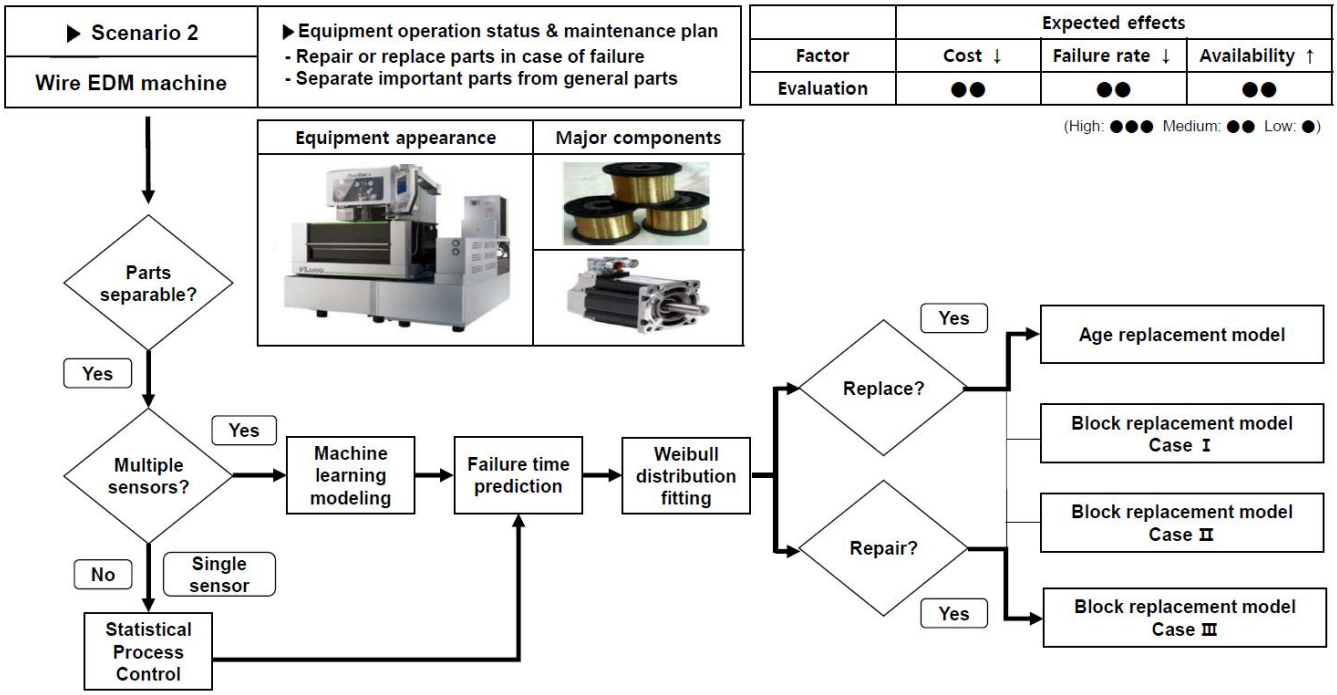
Figure 13(a) presents the prediction and preventive maintenance methodology for turbo compressors, depicted as a flowchart for better comprehension. Turbo compressors are predominantly used in large-scale industries such as steel manufacturing, petrochemicals, and gas power plants. These devices compress atmospheric air using the centrifugal force of the impeller and utilize the compressed air to drive or operate various equipment.

The key characteristic of turbo compressors is their critical role in industrial operations; unexpected failures can halt the entire production line. Moreover, turbo compressors are large, heavy, and difficult to replace with alternative equipment due to their unique functionality. Functionally, turbo compressors can be divided into components such as coolers, motors, pumps, and filters, each comprising multiple subparts. Typically, turbo compressors, as one of the most vital pieces of equipment on the production line, are managed with numerous sensors attached. Given these characteristics, machine learning modeling is recommended to predict the RUL and failure time of the equipment. Considering the nature of the equipment, delays caused by component replacements result in increased losses, making it crucial to enhance equipment availability. Therefore, in terms of component replacement, Case III of the block replacement model, which considers availability, is prioritized. However, adjustments and refinements can be made based on field process conditions, environmental and operational variables, and priority shifts.

Figure 13(b) illustrates the predictive and preventive maintenance approach for a wire EDM (Electrical Discharge Machining) machine. A wire EDM machine cuts a metal workpiece using a thin brass wire. This machine is primarily used for precision and complex metal processing in industries such as automotive, industrial robot, and large-scale machinery manufacturing. They are often a part of metal processing production lines in manufacturing plants, where multiple machines exist and run simultaneously. Compared with a turbo compressor, the failures of wire EDM machines do not halt the entire production line



(a) Turbo compressor



(b) Wire EDM machine

Figure 13. Industrial application scenarios

because substitutes of the wire EDM machine exist in the line. The replacement model is applied differently based on the importance and characteristics of the components. An age replacement model is applied when the failure rate gradually increases over time or when replacement is preferred over repair. For components with a large quantity or those subject to repair, a block replacement model can be used.

In practice, the adaptability and scalability of the proposed method can be flexibly adjusted depending on operational constraints and failure rate distributions in various industries. For instance, stochastic variations must be considered in wind turbines, hierarchical optimization is required for railway infrastructure, and conservative maintenance strategies are essential in highly regulated industries like energy power plants and aerospace. Therefore, the proposed method can be effectively tailored to different industrial environments, enabling dynamic and cost-efficient maintenance planning across various sectors, including manufacturing and large-scale production systems.

5.2. Potential impacts in industries

The highlight of the proposed method is to minimize maintenance cost and enhance system availability simultaneously by integrating machine learning-based RUL prediction with a cost-optimal block replacement model. Machine learning reduces uncertainty in RUL estimation through providing data-driven intelligence beyond statistics.

The proposed method aimed to optimize preventive maintenance by reducing the uncertainty in RUL estimation. In industrial environments where multiple machines operate simultaneously, traditional maintenance strategies determine maintenance schedules typically using the arithmetic mean of the RUL values predicted for multiple machines. However, such arithmetic mean is limited to account for the variability and uncertainty inherent from individual machine conditions, thereby leading to premature or delayed replacements in traditional maintenance strategies. On the contrary, the proposed method achieved a cost reduction of 2.2% per cycle in the case study, compared with the arithmetic mean of predicted RUL values. For large-scale manufacturing industries, where maintenance costs can reach millions of dollars annually, a 2.2% per cycle cost reduction represents substantial economic benefits. Additionally, under optimal operating conditions, the proposed method theoretically reduced maintenance costs by up to 32% (from $1,343 to $912 per cycle). It should be noted that cost saving can vary in industrial applications depending on system configurations and maintenance policies.

In addition, the proposed method aimed to improve system availability toward downtime minimization by optimizing maintenance schedules and reducing MTTR (Mean Time To Repair). Such schedule optimization in preventive replacements can minimize unnecessary downtime and improve overall system reliability. Our data-driven method aids in achieving more system-efficient operations by transitioning from traditional static maintenance strategies.

6. CONCLUSION

This study proposes a PM method to predict the RULs of multiple mechanical systems to decide their optimal replacement times for maintenance cost minimization. First, machine learning models were generated to predict the RULs of mechanical systems and were evaluated to select the best model based on several regression evaluation metrics. Second, preventive replacement cycles were optimized to minimize maintenance costs based on the best RUL prediction model and minimum-repair block replacement model. Optimization could be achieved by utilizing the predicted RULs to fit the Weibull distribution that represented the probability of failure and reliability in the PM.

The contributions of this study are as follows. First, this work bridges the theoretical and practical gaps in maintenance optimization by integrating predictive analytics with actionable preventive strategies. Unlike traditional RUL prediction studies that remain at the prediction stage, this study demonstrates how to apply RUL predictions to optimize maintenance schedules and reduce costs in real-world industrial settings. Second, the proposed method is an intermediator between PM and PdM, that is, it is a PdM-based PM, enabling industries to transition from static to dynamic maintenance strategies. Third, by targeting multiple systems grouped as blocks, this study enhances the scalability and applicability of maintenance optimization methods to complex industrial environments.

While our method demonstrated cost saving and improved maintenance efficiency, some limitations remain. First, the proposed method was validated only using the CMAPSS dataset, which might not fully capture the variability of real industrial settings. Second, the proposed method assumed relatively stable failure distributions. Future work will focus on validating the proposed method using actual industrial datasets to reinforce its applicability in industries. The proposed method needs to validate applicability by analyzing actual datasets in diverse industrial sectors, such as power generation, railways and manufacturing. Another future work will increase adaptability of the proposed method through integrating a dynamic method in highly variable operation conditions. Expectedly, dynamic maintenance scheduling needs to be developed in the systems with markedly different failure rates or highly variable operational conditions owing to uncertainty or instability in the environment of the systems. A dynamic PdM-PM method would provide deeper insights into enhancing adaptability in complex and heterogeneous industrial environments.

ACKNOWLEDGEMENT

This research was supported by the Ministry of SMEs and Startups, Republic of Korea, under 'Continuous Process Manufacturing Standardization of Shared Data between Facilities/Factories/Businesses in Characteristic Industries' in 'Smart Manufacturing Innovation R&D Program' (RS-2022-00140694).

NOMENCLATURE

PdM	Predictive maintenance
CM	Corrective maintenance
PM	Preventive maintenance
CBM	Condition-based maintenance
t_0	Replacement time
t_0^*	Optimal replacement time
T	Failure time
$F(t)$	Probability distribution function
$f(t)$	Probability density function
$R(t)$	Reliability function
$h(t)$	Failure rate function
$E(t)$	Expected time
$E(c)$	Expected total cost
C_p	The total cost of preventive replacement
C_f	The total cost of failure replacement
C_d	Downtime cost
C_k	Minimum-repair cost
$C_A(t_0)$	Expected cost rate per unit time of age replacement model
t_p	Interval of block replacement model
t_p^*	Interval of optimal block replacement model
$M(t_p)$	Renewal function
$m(t_p)$	Renewal density function
$C_{B1}(t_p)$	Expected cost rate per unit time for Case I in block replacement model
$C_{B2}(t_p)$	Expected cost rate per unit time for Case II in block replacement model
$C_{B3}(t_p)$	Expected cost rate per unit time for Case III in block replacement model
$RMSE$	Root mean square error
$\hat{Y}$	Predicted value
Y	Actual value
α	Scale parameter of Weibull distribution
β	Shape parameter of Weibull distribution
DFR	Decreasing failure rate
CFR	Constant failure rate
IFR	Increasing failure rate
$MTTF$	Mean time to failure
$MTTR$	Mean time to repair
Γ	Gamma function
r	$(C_f - C_p)/C_p$
MLE	Maximum likelihood estimation

REFERENCES

Aydemir, G., & Acar, B. (2020). Anomaly monitoring improves remaining useful life estimation of industrial machinery. *Journal of Manufacturing Systems*, *56*, 463–469. https://doi.org/10.1016/j.jmsy.2020.06.014

Biggio, L., & Kastanis, I. (2020). Prognostics and health management of industrial assets: Current progress and road ahead. *Frontiers in Artificial Intelligence*, *3*, 578613. https://doi.org/10.3389/frai.2020.578613

Cailian, L., & Chun, Z. (2020). Life prediction of battery based on random forest optimized by genetic algorithm. *2020 IEEE International Conference on Prognostics and Health Management* (ICPHM), 19950220. https://doi.org/10.1109/ICPHM49022.2020.9187060

Chao, M.A., Kulkarni, C., Goebel, K., & Fink, O. (2021). Aircraft Engine Run-to-Failure Dataset Under Real Flight Conditions for Prognostics and Diagnostics. *Data*, *6*(1), 5. https://doi.org/10.3390/data6010005

Chen, C., Shi, J., Lu, N., Zhu, Z.H. & Jiang, B. (2022). Data-driven predictive maintenance strategy considering the uncertainty in remaining useful life prediction. *Neurocomputing*, *494*, 79-88. https://doi.org/10.1016/j.neucom.2022.04.055

Dong, W., Liu, s., Cao, Y., & Bae S. (2020). Time-based replacement policies for a fault tolerant system subject to degradation and two types of shocks. *Quality and Reliability Engineering International*, *36*(7), 2338–2350. https://doi.org/10.1002/qre.2700

Ferreira, C., & Gonçalves, G. (2022). Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. *Journal of Manufacturing Systems*, *63*, 550–562. https://doi.org/10.1016/j.jmsy.2022.05.010

Frederick, D., DeCastro, J., & Litt, J. (2007). *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)*. NASA/TM-2007-215026.

Jin, L., & Yamamoto, W. (2017). Adaptive age replacement using on-line monitoring. *Procedia Engineering*, *174*, 117–125. https://doi.org/10.1016/j.proeng.2017.01.177

Kraus M., & Feuerriegel, S. (2019). Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences. *Decision Support Systems*, Vol. 125, 113100. https://doi.org/10.1016/j.dss.2019.113100

Lei, Y., Li, N., Guo, L., Li N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

Liu, L., Wang, L., & Yu, Z. (2021). Remaining useful life estimation of aircraft engines based on deep

convolution neural network and lightGBM combination model. *International Journal of Computational Intelligence Systems, 14*, 165. https://doi.org/10.1007/s44196-021-00020-1

Lv., Y., Guo, X., Zhou, Q., Qian, L. & Liu, J. (2023). Predictive maintenance decision-making for variable faults with non-equivalent costs of fault severities. *Advanced Engineering Informatics, 56.* https://doi.org/10.1016/j.aei.2023.102011

Najdi, B., Benbrahim, M.& Kabbaj, M.N. (2024). Adaptive Res-LSTM attention-based remaining useful lifetime prognosis of rolling bearings. International *Journal of Prognostics and Health Management*, https://doi.org/10.36001/ijphm.2025.v16i1.4171

Nakagawa, T. (1979). A summary of block replacement policies. *Operations Research*, *4*, 351–361.

Narayanan, L.K., Loganayagi, S., Hemavathi, R., Jayalakshmi, D. & Vimal V.R. (2024). Machine Learning-Based Predictive Maintenance for Industrial Equipment Optimization. *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies.* https://doi.org/ 10.1109/TQCEBT59414.2024.10545280

Park, P., Jung, M., & Di Marco, P. (2020). Remaining useful life estimation of bearings using data-driven ridge regression. *Applied Sciences*, *10*(24), 8977. https://doi.org/10.3390/app10248977

Rausand, M., & Hoyland, A. (2004). *System reliability theory models, statistical methods, and applications*. Press, Wiley.

Rebaiaia, M.L., Ait-Kadi, D., & Jamshidi, A. (2017). Periodic replacement strategies: optimality conditions and numerical performance comparisons. *International Journal of Production Research*, *55*(23), 7135–7152. https://doi.org/10.1080/00207543.2017.1349953

Rebaiaia, ML., & Ait-kadi, D. (2020). Maintenance policies with minimal repair and replacement on failures: analysis and comparison. *International Journal of Production Research, 59*(23), 6995–7017. https://doi.org/10.1080/00207543.2020.1832275

Ross, S.M. (1980). *Introduction to probability models*. Academic, Press, N.Y.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 International Conference on Prognostics and Health Management*, 10423504. https://doi.org/10.1109/PHM.2008.4711414

Shi, J., Yu, T., Goebel, K., & Wu, D. (2021). Remaining useful life prediction of bearings using ensemble learning: The impact of diversity in base learners and features. *Journal of Computing and Information Science in Engineering*, *21*(2), 021004. https://doi.org/10.1115/1.4048215

Thyago P. Carvalho, Fabrízzio A. A. M. N. Soares, Roberto Vita, Roberto da P. Francisco, João P. Basto, & Symone G. S. Alcalá. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering 137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Tong, Z., Miao, J., Tong, S., & Lu, Y. (2021). Early prediction of remaining useful life for Lithium-ion batteries based on a hybrid machine learning method. *Journal of Cleaner Production*, *317*, 128265. https://doi.org/10.1016/j.jclepro.2021.128265

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, *139*, 469–489. https://doi.org/10.1016/S0377-2217(01)00197-7

Woo, J., Shin, S., Seo, W., & Meilanitasari, M. (2018). Developing a big data analytics platform for manufacturing systems: architecture, method, and implementation. *The International Journal of Advanced Manufacturing Technology*, *99*, 2193–2217. https://doi.org/10.1007/s00170-018-2416-9

Wu, J.Y., Wu, M., Chen, Z., Li, X., & Yan, R. (2021). A joint classification-regression method for multi-stage remaining useful life prediction. *Journal of Manufacturing Systems*, *58*, 109–119. https://doi.org/10.1016/j.jmsy.2020.11.016

Xue, Z., Zhang, Y., Cheng, C., & Ma, G. (2020). Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression. *Neurocomputing*, *376*, 95–102. https://doi.org/10.1016/j.neucom.2019.09.074

Zhao, X., A-Kalifa, KN., Hamouda, AM., & Nakagawa, T. (2017). Age replacement models: A summary with new perspectives and methods. *Reliability Engineering and System Safety*, *161*, 95–105. https://doi.org/10.1016/j.ress.2017.01.011

Zonta, T., Costa, C., Zeiser, F.A., Ramos, G., Kunst, R. & Righi, R. (2022). A predictive maintenance model for optimizing production schedule using deep neural networks. *Journal of Manufacturing Systems, 62.* 450-462. https://doi.org/10.1016/j.jmsy.2021.12.013