Expert Systems with Applications, Volume 176, 15 August 2021, Article number 114897
DOI:10.1016/j.eswa.2021.114897

Adopting machine learning and condition monitoring P-F curves in determining and prioritizing high-value assets for life extension

Sunday Ochella[1], Mahmood Shafiee[2,*], Chris Sansom[1]

[1]*Department of Energy and Power, Cranfield University, Bedfordshire MK43 0AL, UK*
[2]*Mechanical Engineering Group, School of Engineering, University of Kent, Canterbury CT2 7NT, UK*
[*]*Corresponding author. E-mail: m.shafiee@kent.ac.uk*

Abstract: Many machine learning algorithms and models have been proposed in the literature for predicting the remaining useful life (RUL) of systems and components that are subject to condition monitoring (CM). However, in cases where data is ubiquitous, identifying the most suitable equipment for life-extension based on CM data and RUL predictions is a rather challenging task. This paper proposes a technique for determining and prioritizing high-value assets for life-extension treatments when they reach the end of their useful life. The technique exploits the use of key concepts in machine learning (such as data mining and *k*-means clustering) in combination with an important tool from reliability-centered maintenance (RCM) called the potential-failure (P-F) curve. The RCM process identifies essential equipment within a plant which are worth monitoring, and then derives the P-F curves for equipment using CM and operational data. Afterwards, a new index called the potential failure interval factor (PFIF) is calculated for each equipment or unit, serving as a health indicator. Subsequently, the units are grouped in two ways: (*i*) a regression model in combination with suitably defined PFIF window boundaries, (*ii*) a *k*-means clustering algorithm based on equipment with similar data features. The most suitable equipment for life-extension are identified in groups in order to aid in planning, decision-making and deployment of maintenance resources. Finally, the technique is empirically tested on NASA's Commercial Modular Aero-Propulsion System Simulation datasets and the results are discussed in detail.

Keywords: Machine learning; Data mining; Potential failure interval factor; *k*-means clustering; Life-extension; Remaining useful life; Condition monitoring.

1. Introduction

Engineering plants and systems have evolved progressively and have become significantly more intelligent in recent years and so have the demands made from these systems in terms of human dependence on their uptime and functionality. For instance, human activity is so dependent on power, as only few hours of downtime on the power grid will pose serious economic as well as safety risks [1]. Similarly, failure of offshore infrastructure such as oil and gas production facilities, marine renewable energy assets and other ship vessels and structures will affect not only the businesses but also a long trail of people along the value chain. This helps to emphasize the utmost importance of the need to ensure the safety and reliability of these systems. Therefore, as the evolution into the era of industry 4.0 continues, with an abundance of data being generated from engineering plants and installations, new ways of analyzing these data to make meaningful impacts, especially as regards asset life and integrity management, is exigent.

In practice, not all pieces of equipment within an engineering plant will benefit from tight inspection and maintenance regimes or life-extension treatments. As a practice, some equipment can actually be run until they fail because the safety, environmental and economic consequences of their failures are negligible. However, some other equipment might be incident-critical and, therefore, it will not be efficient to run them until they fail because of the huge safety and economic implications. For a plant with hundreds of equipment within its assets register, identifying the most vulnerable equipment for life-extension is a challenging task that, if carried out effectively, will help to ensure safe and cost-effective operations in later life of the facility. This is important for asset managers as it helps them assign resources towards life-extension in a more efficient and

effective manner. This study therefore directly contributes to the process of making life-extension decisions in a data-driven context, given an ecosystem where lots of operational, environmental and condition monitoring (CM) data are constantly gathered from plant operations.

An important development in the industry 4.0 era is the recent rapid advancement in sensor technologies and an attendant increase in the amount of data being collected from equipment on an operational facility. The resultant ease of collecting data from engineering assets has led to an increase in new ways of exploiting these data for asset monitoring purposes. One of the most popular approaches in recent time is the use of machine learning techniques and algorithms to develop models that provide insight into the underlying condition of equipment, based on data. This approach has proven to be popular because modern systems are complex and their failures cannot be simply modelled via physics of failure approaches.

The literature is awash with studies that use machine learning (ML) algorithms for remaining useful life (RUL) prediction. However, most of the proposed algorithms predict RUL with the intent of optimizing asset maintenance strategies and to aid in logistics support planning [2]. Furthermore, the existing methods have mainly been applied to structures and static mechanical equipment and include carrying out structural integrity assessments to determine the choice of a life extension strategy. Such approaches typically implement a life-extension program as a stand-alone project at the end of an asset's initial design life. A model that relates RUL prediction directly to life-extension decision-making was proposed by Vaidya and Rausand [3]. The model considered various factors such as future loading, system design information and expert opinions; however, the RUL prediction model was physics-based. Our study, instead, proposes the use of data-driven ML techniques to determine and prioritize the equipment for life-extension activities, strictly using sensor data gathered during operations as well as CM data, and not based on a formal structural integrity assessment. To the best of the authors' knowledge, and based on findings from literature search, this is the first attempt of looking at asset life extension as an ongoing series of activities and proposing strategies from a ML perspective, along with the use of tools from reliability centered maintenance (RCM). In that regard, the major contribution of this paper includes a unique attempt at combining a tool from RCM called the potential failure (P-F) curve and ML algorithms (e.g. data mining, *k*-means clustering) to prioritize vulnerable equipment for life-extension under an era of ubiquitous data. The study suggests a new technique of visualizing and exploiting P-F curves derived from CM data by assessing P-F curves from multiple equipment simultaneously, and then clustering equipment with similar degradation profiles, similar effort required during life-extension actions, similar spares philosophy and similar performance requirements in terms of safety and reliability. A novel index, called the potential failure interval factor (PFIF), is proposed to measure the health state of equipment. This new index, which has no unit, will enable the comparison of disparate pieces of equipment with dissimilar ranges of total lifetime, thereby fully exploiting the massive sensor data available to engineers in order to optimize life-extension planning and implementation.

The remaining part of this paper is organized as follows. Section 2 provides the theoretical background for P-F curves in CM and data mining, culminating in the choice of *k*-means clustering as the preferred algorithm. Section 3 provides details of the proposed technique, including data preprocessing and features selection, algorithm for fitting a regression model to processed data and applying a clustering algorithm to obtain groups. A demonstration of the applicability of the technique is presented in Section 4. Section 5 discusses the results obtained; and finally, Section 6 presents the conclusion and suggestions for further work.

2. Theoretical background

Different strategies for implementing life-extension actions in ageing engineering assets have been deployed by practitioners within different industry sectors as well as researchers in academia. Sharp *et al.* [4] proposed a framework that involved dividing the equipment on an offshore oil and gas facility into different functional groups such as structural components, process systems, marine systems and safety equipment, and

then developing performance indicators to determine an acceptable threshold for triggering life-extension actions. Essentially, the life-extension activities or remediation schemes proposed were under the broad category of repair, replace or upgrade. Some other approaches were proposed by Shafiee *et al.* [5], which include replacement/repowering, reconditioning, restoration (repair, remanufacture or retrofitting), reclaiming, retro-filling and use-up. In light of the diversity of the various life-extension strategies that have been proposed and implemented, it is important to establish a framework for prioritizing equipment under consideration for life-extension that not only fits into the operational philosophy of asset owners, but also duly takes into account the peculiarities of the information available to asset operators about the various pieces of equipment within the plant. Ersdal *et al.* [6] recognized that end-of-life assets can be grouped into four categories, namely: parts that cannot be inspected or maintained; parts with missing or incomplete inspection or maintenance data; parts with widespread deterioration; and technologically obsolete parts.

Irrespective of the approach used, the major considerations during asset life-extension include: safety, economics, regulatory requirements, serviceability, practicality of life-extension strategy implementation and, more recently, convergence with the new era of "smart systems". Pérez Ramírez *et al.* [7] proposed a systems engineering approach to the management of ageing oil and gas facilities such that the end-of-life strategies are incorporated into the maintenance philosophy of a facility with the overall aim of making equipment function well beyond their original design life. In another study, Shafiee *et al.* [8] proposed a techno-economic feasibility assessment framework for prioritizing safety critical elements (SCEs) within a plant for life-extension purposes. They showed that cost is a major driver in choosing a suitable end-of-life strategy by most asset managers. Animah *et al.* [9] developed a life-cycle cost-benefit approach that takes into account several categories of expenditures during the extended phase of operation of offshore assets, thus aiding asset managers to make informed choices based on calculated costs and benefits.

Most of the approaches mentioned so far ride on conventional methods of implementing a life-extension program which involves a project-like approach executed at the end of asset design life. This work takes a unique approach by viewing life-extension activities as an on-going series of activities, since different equipment within a fleet typically have varying design lives. The proposed approach involves mining data from each unit within the fleet and, based on strictly data, grouping units with similar time-to-failure indicators together for life-extension action. In the following subsections, a detailed background of the key tools used in this work are presented. This, in addition to relevant references, will aid easy understanding of the concepts used throughout the study.

2.1 *Reliability-centered maintenance (RCM)*

RCM, as a concept, was first proposed in the work by Nowlan and Heap [10], where they studied a fleet of aircraft at United Airlines and proposed changes to the existing maintenance program at the time. With RCM, they put forth a program that attempts to answer critical questions surrounding how failures occur, what the consequences of failures are, and what type of maintenance actions can prevent failures from occurring. Although the fundamental concepts have remained the same, the practice has evolved since then and has been adopted by maintenance engineers and asset managers across various industries. RCM was defined by Moubray [11] as a set of practices which must be carried out to ensure that any physical asset continues to perform its desired function. Failure of an equipment to meet pre-defined performance standards, within a given operational context, is therefore defined as a functional failure. The role of RCM is therefore to ensure that maintenance practices keep the identified equipment in such a state as to ensure that functional failure, with its attendant undesired consequences, is avoided. RCM practice asks the following key questions [12]:

i. Within an operational context, what are the functions of each equipment and the associated performance standards?
ii. In what ways does each equipment fail to perform its specified functions?
iii. What are the causes of each functional failure?

iv. What are the consequences of each failure?
v. What can be done to predict or prevent each failure?
vi. What should be done if a suitable proactive task cannot be found?

Questions (v) and (vi) are directly related to remaining useful life (RUL) estimation and life-extension considerations. This paper will therefore draw from the RCM concepts related to these two questions to help identify equipment for life-extension. The logical flow of the RCM decision process is illustrated in Fig. 1. This flow process specifies activities that intend to answer the key RCM questions mentioned above. The process involves a collection of all the assets within a plant in the form of an asset register/database, along with accompanying operational records and maintenance history for each equipment. The entire plant is then divided into systems, sub-systems and components, along with a definition of their operating contexts. Functional analyses are then carried out in order to define functional requirements and performance standards, thereby helping to establish what functional failure entails for each equipment. Based on the functional analyses and subsequent Failure Modes, Effects and Criticality Analyses (FMECA), the equipment are categorised, according to their criticalities, into different maintenance strategies. Table 1 specifies the categorization of the maintenance strategies and the application scenarios.

Figure 1

Fig. 1. RCM decision logic flowchart – adapted from Liang *et al.* [13].

Table 1

Table 1. RCM strategies and their associated application scenarios.

Obviously, there is no added benefit of extending the life of non-critical equipment or the ones designated for redesign. Thus, only equipment categorized under failure finding, scheduled restoration, scheduled discard and on-condition maintenance will typically be the focus for life-extension. In this study, the data from such equipment is mined, in combination with their potential failure curves, and a clustering algorithm is then applied to obtain clusters of equipment with similar health states. The ultimate aim of grouping equipment according to health states is to focus on the vulnerable groups which are likely to fail first, thereby aiding maintenance decision-making and the subsequent application of life-extension actions to equipment within the vulnerable groups.

2.1.1 *Predictive testing and inspection*

Equipment condition can be monitored through non-intrusive testing, supervisory control and data acquisition (SCADA), visual inspection and other testing methods, depending on the failure modes for the equipment being monitored. This practice is also referred to as condition monitoring (CM). Some important predictive testing and inspections, which are vital to the detection of incipient faults and performance deterioration, include vibration monitoring, infrared thermography, ultrasonic noise (acoustics) measurements, lubricant (oil) analyses, temperature measurements, flow characteristics, ultrasonic thickness measurements, eddy current testing and motor current signature analysis, amongst others. A detailed coverage of CM techniques is covered in [11]. Data from these inspections, when collected continuously or at intervals, and in combination with the baseline data, can be plotted against time to help reveal the performance characteristics. With enough historical data, the performance plot can be used to detect the point of incipient failure, also known as potential failure point. This point can only be detected when performance has started declining and potential failure is possible, hence the name potential failure (or P-F) curve.

There are a few papers in the literature which have used the P-F curve as a tool for evaluating performance and modelling degradation of equipment. Van Horenbeek *et al.* [14] studied the added value of implementing an imperfectly performing CM system into a wind turbine gearbox by using the P-F curve. The associated secondary damage, which can be prevented with early detection of potential failures, was also factored into the model. The methodology was tested on a wind turbine gearbox dataset selected from a manufacturer with a fleet of more than 800 onshore wind turbines operating over a time span of eight years. The approach can be extended to offshore wind energy applications but with more stringent detectability and efficiency parameters due to the logistical complexities of maintaining offshore assets. Lorenzoni *et al.* [15] modeled the degradation of components using Dynamic Bayesian Networks, with the P-F curve representing the degradation pattern which was modeled as a reversed exponential function. The characteristic of the P-F curve in the study was susceptible to maintenance activities as well as operating conditions, thus factoring these uncertainties in to derive the health state of equipment. Five different health states were used in their study to characterize operating equipment, including: new or as good as new, very slight indication of degradation, serious degradation, stage of rapid decline, and finally, stage with very high probability of failure.

2.1.2 *Potential failure curve (P-F curve)*

Based on the information gathered from predictive testing and inspection tasks, the condition of an equipment when plotted against time will yield the potential failure or P-F curve. Fig. 2 shows typical P-F curves.

Figure 2

Fig. 2. (a) A typical P-F curve, **(b)** A P-F curve for fatigue crack propagation (adapted from [16]).

The P-F curve is so named because it indicates the point at which the failure of an equipment being monitored becomes detectable. This point is indicated as the potential failure point, P, in Fig. 2(a). From commencement of the service life of an equipment up to a certain point, failure is undetectable because all the parameters of the equipment being monitored, like temperature, vibration, lube oil analysis, etc., indicate that the equipment is in a health state that is devoid of detectable faults. However, incipient failure becomes detectable at a certain time when deviations start to occur. The time from the actual point of detection of potential failure to the point of functional failure is referred to as the P-F interval. It is desirable that the P-F interval is sufficient for both decision-making and actual maintenance and life-extension activity, in order for the whole endeavor to be worthwhile.

2.1.3 *P-F interval determination*

Fig. 2(b) vividly illustrates how equipment performance degrades over time for a single failure mode and also how different CM techniques can detect the failure at different stages. If a visual inspection is conducted at point P_2, the exact size of the crack will not be detected. If, however, an appropriate and more accurate inspection technique, say radiography, is performed just after point P_2 but before point P_3, then it gives a P-F interval within the range $(t - t_2)$ to $(t - t_3)$, during which a maintenance intervention should be planned and implemented. Hence, for critical equipment, continuous monitoring using the right technologies and techniques is essential, in order to ensure early detection. The right data acquisition frequencies are also important in order for the P-F curve to serve as an effective tool to identify equipment undergoing deterioration in health state. In simple terms, it is desirable for the inspection interval to be less than the P-F interval in order for faulty conditions to be captured before failure occurs.

In practice, it is difficult to determine the P-F interval for most equipment. For some age-related degradations, the P-F curve could be linear from the point of occurrence of incipient failure to the point of

functional failure. For such cases, determination of the P-F interval can be performed by a straightforward extrapolation using the slope of the straight line degradation curve. However, in reality, most equipment exhibit non-linear degradation characteristics. Thus, estimating time-to-failure becomes an arduous but critical exercise.

2.1.4 *Relationship between P-F interval, useful life and asset life*

Moubray [11] defined useful life, L_u, as the period from commencement of service to the age at which the conditional probability of failure significantly increases. This may or may not coincide with the point at which incipient failure is first noticed. Jardine *et al.* [17] defined RUL as "the time left before observing a failure given current machine age and condition, and past operation profile". Fundamentally, an asset's lifetime can be subdivided into the useful life (normal operating state) and the faulty state during which the asset operates with an existing fault. Goode *et al.* [18] termed these two operating zones as "stable zone" and "failure zone" respectively. The entire asset life, L_a, is therefore defined as the sum of the times when the asset is in a good health state and the time when it operates in an unhealthy state until it fails. This is illustrated in Fig. 3. The asset life is therefore given by Eq. (1):

$$Asset\ life = Useful\ life + faulty\ zone \tag{1}$$

In Fig. 3, the faulty zone comprises the P-F interval (PF_{int}) and t_d, which represents the time difference between when the incipient failure actually started and when it is detected using sensor devices. So, the asset life is given in Eq (2) as:

$$L_a = L_u + (t_d + PF_{int}) \tag{2}$$

The ultimate goal of life-extension actions is to extend the service life of an asset beyond its original design lifetime. Upon detection of a fault, a life-extension action is carried out (labelled as on-condition maintenance in Fig. 3.) and the condition of the equipment returns to almost as good as new condition. This action potentially increases the lifetime of the equipment from "averted failure point 1" to "averted failure point 2". Effective monitoring and life-extension can therefore potentially continue in such cycles until a cut-off point called maximum lifetime, L_{max}, is reached, beyond which the asset owner, either as a matter of policy or for some other reasons, decommissions the equipment or plant.

Figure 3

Fig. 3. Effect of a life-extension action on P-F curve.

2.1.5 *P-F interval factor (PFIF)*

In this study, we define an index, called the P-F interval factor, for degrading components. This index is given by Eq. (3):

$$P-F\ Interval\ Factor_{i,t} = \frac{P-F\ Interval_{i,t}}{Unit\ Lifetime_i} \tag{3}$$

where $P-F\ Interval\ Factor_{i,t}$ is the P-F interval factor of the unit i at time t, $P-F\ Interval_{i,t}$ is the P-F interval of unit i at time t, and $Unit\ Lifetime_i$ is the operational or design life of the unit i. This indicator is important because by normalizing the P-F interval with the lifetime of each unit, a scale-independent value is obtained, which enables the grouping of disparate pieces of equipment with different ranges of total lifetime or P-F intervals. This is a very useful index that will also be used for health stage division, thereby serving as an indicator of the state of health of any unit under operation. For illustration purposes, consider a hypothetical

case where one equipment, A, has a typical lifetime duration of 20 years and another, B, a lifetime duration of 6 months. In order to group these equipment for life-extension action, if CM and sensor data suggest that A has two years left (which is the P-F interval) and B has half a month left, using the P-F interval alone produces two different timelines, which will not be useful for the purpose of grouping them together as equipment that are *soon-to-fail*. However, the PFIF index in case A is 0.1 and in case B is 0.083. Thus, depending on the clustering criteria, the ML algorithm will cluster both equipment in the same group: *soon-to-fail.*

2.2 *Data mining concepts and cluster analysis*

Data mining involves the extraction of embedded, hitherto unknown but essentially insightful and valuable information from data. Key features in data mining include the use of computer-based tools and algorithms, and the availability of big data, such that conventional methods of statistical analysis become unreasonable to implement. Two practical goals of data mining are prediction and description [19]. In the context of this paper, clustering will be used as a descriptive function to help group equipment that are at a similar stage of health with the aim of subsequently performing proactive or predictive tasks on the derived groups.

Cluster analysis generally entails using a set of methodologies to automatically group or classify observations using linkage rules such that observations similar to each other are in the same group while dissimilar observations come under different groups [20]. Cluster analyses are of two broad types, hierarchical and partitional clustering. Other clustering types are density-based, grid-based or model-based [21]. The two broad types are briefly discussed below and the rational for using *k*-means clustering for this work is thereby highlighted.

2.2.1 *Hierarchical clustering*

Hierarchical clustering groups data using a cluster tree or *dendrogram*. It is subdivided into agglomerative hierarchical clustering and divisive clustering, as shown in Fig. 4. Hierarchical agglomerative clustering is a bottom-up approach that starts with each data point as a member of a cluster and recursively merges clusters until a final single cluster is obtained. On the other hand, the divisive clustering process, which is a top-down approach, is procedurally the direct opposite of agglomerative clustering. It begins with the entire dataset as one cluster and progresses by dividing each cluster until a final stage where each data point stands on its own.

Figure 4

Fig. 4. Dendrogram for the two types of hierarchical clustering – adapted from [21].

For both methods, similarity rules are applied to merge data points into clusters. Zhao *et al.* [22] extracted latent variables that are not directly measured by sensors and also their correlation coefficients. An agglomerative hierarchical clustering algorithm was then used to group the extracted variables as well as the sensor readings using similarity measures, with the aim of identifying equipment for predictive maintenance. The method was applied to an electrical generator and its subsystems. Abdelhadi [23] used an agglomerative hierarchical clustering approach to cluster repairable machines into virtual cells for maintenance tasks. The study developed a machine failure incidence matrix from which an eigenvector for each failure is derived. Afterwards, a similarity matrix was generated such that the relation between failures and equipment in terms of relative weights were captured. Machine cells were then developed and failures were assigned to suitable cells via a complete linkage agglomerative algorithm.

2.2.2 *Partitional clustering*

The main type of partitional clustering is the *k*-means clustering, and its variants. The *k*-means clustering groups the points in a dataset by assigning observations to a predefined number of clusters. The step-by-step procedure for a typical *k*-means clustering algorithm is given below:

i. Initialize by determining number of clusters (i.e. *k*) containing randomly allocated data points or observations.
ii. Compute the centroids of each cluster in step (i) and compare all data points to the centroids by the use of a distance metric, moving data points to the closest centroids thereby adjusting the initial clusters.
iii. Compute the new centroids.
iv. Repeat steps (ii) and (iii) until there is no further movement of data points between clusters.

Three important parameters in the *k*-means algorithm are the number of clusters, *k*, cluster initialization and a distance metric [24]. While a number of distance measures like Euclidean distance, the Jaccard distance, the Mahalanobis distance, Manhattan distance, cosine distance, and so on, have been used for *k*-means and other clustering algorithms, the Euclidean distance is the most commonly used for the *k*-means algorithm. This is because, amongst other reasons, the *k*-means algorithm clusters data points represented in a multidimensional Euclidean space. So, the algorithm takes input parameter, *k*, and partitions *m* data points so that the resulting intra-cluster similarity is high but the inter-cluster similarity is low. This objective is achieved by minimizing the squared error in the distance between each data point in a cluster and its centroid. Given *m* samples of multidimensional data in a multidimensional space, which are to be partitioned into *k* clusters, the sum of squared errors is given by:

$$E^2 = \sum_{j=1}^{k} \sum_{i=1}^{n(C_j)} \left\| p_{i,j} - \mu_j \right\|^2 \tag{4}$$

where $n(C_j)$ is the number of data points in cluster C_j, where j ranges from 1 to k, $p_{i,j}$ is a vector representing the i^{th} data point within cluster C_j (i.e. $p_{i,j} \in C_j$) and μ_j is the mean vector representing the centroid of cluster C_j, which is obtained as:

$$\mu_j = \frac{1}{n(C_j)} \sum_{i=1}^{n(C_j)} p_{i,j} \tag{5}$$

and

$$\sum_{j=1}^{k} n(C_j) = m \tag{6}$$

As opposed to *k*-means clustering where each data point is assigned to a single cluster (hard assignment), a variation where each data point can be a member of multiple clusters with a membership value (soft assignment) is referred to as the fuzzy *c*-means clustering. Other variations of *k*-means clustering are highlighted in [24]. Table 2 provides the pros and cons of the two broad types of clustering.

Table 2

Table 2. Pros, cons and application cases for the two broad classes of clustering algorithms.

Regarding research in the area of maintenance scheduling, Gholami and Hafezalkotob [25] used *k*-means clustering to group equipment based on similarity of maintenance activities and then the rules were extracted to characterize the derived clusters. The method was applied to data from ten pumps under functional failure conditions. The data comprised pump factor values for the ten pumps for 250 different failures recorded. Lahrache *et al*. [26] used both *k*-means and hierarchical clustering to group faulty and unfaulty knives in a cutting tool machine. Also, Abdelhadi [27] proposed a method to use *k*-means clustering to group repairable machines into virtual groups based on their need for maintenance according to the time to failure and according to the location of the machines. Wakiru *et al.* [28] used a fuzzy cluster analysis to group multiple engines exhibiting similar lubricant performance characteristics based on the data collected from lubricant oil analysis for 17 medium speed engines of a thermal power plant.

3. Methodology

The proposed technique is intended to group equipment within a fleet into clusters with similar health states, enabling life-extension engineers to prioritize equipment approaching their end-of-life. A fleet may consist of a collection of several units of whole systems, or a collection of several units of subsystems or components. Units within a fleet may be identical, similar or heterogeneous [30]. Identical units imply the same system with identical technical features and under the same usage and operational conditions; similar units share almost identical technical features and operational conditions but may have slightly varying usage; while heterogeneous units have varying technical features, usages and operational conditions albeit they share some similarity in data traits that can be exploited for decision-making. This section describes the steps involved in the technique which was developed for a homogenous fleet (i.e. identical and similar units under the same operational conditions). The steps are broken down into two broad parts, phase 1 and phase 2.

3.1 *Phase 1 – data preparation and sensor selection*

Given a dataset of run-to-failure data for m units or pieces of equipment within a homogeneous fleet, let X_i represent the run-to-failure data for the i^{th} unit, where $i = 1, \ldots . m$. Since each unit will have a distinct lifetime, T_i, the data X_i is an array of the order T_i by N, where N represents the number of variables or sensor measurements from each unit. The followings are the steps involved in phase 1 of the methodology:

i. For ease of application of the algorithm, the data is prepared as an ensemble, containing the data for each unit vertically concatenated on each other, to give an overall dataset array, X. The combined dataset X will be an M by N array where M is given by:

$$M = \sum_{i=1}^{m} T_i \tag{7}$$

ii. The raw run-to-failure data, X, which is taken in bulk as the training data, is then cleaned, pre-processed and useful features are extracted. Data preprocessing and feature engineering techniques depend on the nature of the data and the use for which the data is intended [31]. Features can be extracted in the time domain, the frequency domain or the time-frequency domain, depending on the nature of the signals and the specific application. For simple time-domain degradation data, the mean and standard deviation or variance of a signal may change progressively as the equipment degrades. For rotating machinery such as gears, bearings and shafts, common features extracted for health state construction include root-mean-square error (RMSE), kurtosis, peak-to-peak, crest factor, skewness, etc. [32]. Features extraction does not only help to determine which signals are useful indicators of degradation, but also help in dimensionality reduction for the multivariate data. Other techniques of dimensionality reduction like principal component analysis (PCA) can also be used to reduce the dimension of the data from the fleet [33]. Signals with constant values (i.e. no variance) are not useful indicators and are as such eliminated, resulting in a reduced dataset, $X_{reduced}$.

iii. Next, the reduced data is normalized, unit-wise, so as to make the attributes from the different sensors comparable to one another. One approach is standardization. For that purpose, let s be the index representing the sensor number, with s ranging from 1 to N; and, let ℓ be the index corresponding to the number of data points for unit i, with ℓ ranging from 1 to T_i. If the s^{th} sensor for data $X_{ireduced}$ for the unit i has a mean value $\mu_{i,s}$ and standard deviation $\sigma_{i,s}$, then each value $x_{\ell,s}$ of each data point of $X_{ireduced}$ is transformed to:

$$z_{\ell,s} = \frac{x_{\ell,s} - \mu_{i,s}}{\sigma_{i,s}} \tag{8}$$

Another approach is the min-max scaling, which maps the attributes to the range [0,1] using the transformation given in Eq. (9):

$$x_{\ell,s} \rightarrow \frac{x_{\ell,s} - \min(x_{i,s})}{\max(x_{i,s}) - \min(x_{i,s})} \tag{9}$$

where $\min(x_{i,s})$ and $\max(x_{i,s})$ are the minimum and maximum values, respectively, of the s^{th} sensor or feature for unit i. Standardization is used for this work as it is more robust and not susceptible to outliers or extreme values [34].

iv. The normalized data is smoothed using a suitable algorithm, depending on the characteristics of the data such as noise level, presence of outliers, etc. For this work, we adopt a local regression smoothing algorithm, called the robust locally weighted scatterplot smoothing (RLOWESS) [35, 36], due to its effectiveness in handling outliers.

v. To gain further insight into the data $X_{reduced}$, monotonicity, trendability and prognosability metrics are computed as presented in the work of Coble and Hines [37, 38]. The fundamental concept is that features of data important for degradation prediction must be monotonically increasing or decreasing and, in addition, be trendable. This assumption of continuous degradation is mostly true for systems with a combination of electronic and mechanical components and may not be entirely correct for systems that exhibit some level of self-restoration when left temporarily without use, e.g. batteries [39]. Monotonicity, which characterizes the underlying positive or negative trend of a feature, is obtained as the average difference of the fraction of positive and negative derivatives for each run-to-failure data or trajectory. This is given by Eq. (10):

$$Monotonicity = mean\left(\left|\frac{no.\ of\ {}^{d}/_{dx}>0}{T_i - 1} - \frac{no.\ of\ {}^{d}/_{dx}<0}{T_i - 1}\right|\right) \tag{10}$$

In more precise mathematical terms, it can be expressed as follows:

$$Monotonicity = \frac{1}{m}\sum_{i=1}^{m}\left|\sum_{\ell=1}^{T_i-1}\frac{sgn(x_{i,s}(\ell+1) - x_{i,s}(\ell))}{T_i - 1}\right| \tag{11}$$

All symbols are as previously defined while $x_{i,s}(\ell)$ represents the value of the s^{th} sensor or feature for unit i corresponding to the index ℓ. The trendability metric is calculated as:

$$Trendability = \min_{i,h}\left|corr(x_{i,s}, x_{h,s})\right|, \quad i,h = 1,\dots,m \tag{12}$$

where $x_{i,s}$, $x_{h,s}$ represents any pair of vectors for the data from the s^{th} sensor or feature for units i and h respectively.

The prognosability metric gives a measure of the variance of the features towards end-of-life. This is an intuitive metric since a wide variance towards end-of-life can make it difficult to extrapolate a feature to the failure point. Prognosability is calculated by Eq. (13):

$$Prognosability = \exp\left(-\frac{std(failurevalues)}{\text{mean}(|feature_{end} - feature_{start}|)}\right) \tag{13}$$

where $failurevalues$ imply the population of the values of all the features at failure and $|feature_{end} - feature_{start}|$ stand for the difference between the start and end values of each individual feature. This is given in precise mathematical terms as:

$$Prognosability = \exp\left(-\frac{std_i(x_i(T_i))}{mean_i|x_i(T_i) - x_i(1)|}\right) \tag{14}$$

where $x_i(T_i)$ is a vector of the last data values from each sensor for unit i (i.e. just before unit i fails) and $x_i(1)$ is a vector of the first data values from corresponding sensors for the same unit (i.e. at the beginning of operations).

To select the optimal set of features, the three metrics are combined to obtain a fitness value defined as:

$$Combined\ value\ or\ fitness\ =\ w_m Monotonicity + w_t Trendability + w_p Prognosability \quad (15)$$

The weights w_m, w_t and w_p indicate the importance of each metric and should sum up to one. For this work, each metric is weighted equally. The exclusion criterion for each feature is then defined as *fitness* $> \tau$, where τ is a carefully selected threshold based on the values of the three metrics. Values for monotonicity, trendability and prognosability all lie in the range [0, 1], with 0 representing non-trendable features and 1 representing perfectly trendable features. The individual algorithms are implemented as MATLAB in-built functions and subsequently combined, thus selecting the most trendable sensors and obtaining a further reduced dataset, $X_{clusterdata}$, which is ready for use in phase 2 of this methodology.

3.2 *Phase 2, route 1 – fit linear model, construct health indicator and implement health stage division*

To obtain a single health indicator, the selected features are fused together to produce a single degradation trend that represents the instantaneous health states of each unit. There are various studies that propose different methods of doing this [40, 41]. Other methods are presented in a review by Lei *et al.* [2]. Fundamentally, the process involves two stages: health indicator (HI) construction and health stage (HS) division. HI construction can be further categorized into two: physics HI, which is related to the physics of failure and virtual HI, which involves fusing multiple sensor signals together to give a virtual description of the degradation trends of complex systems based on data. Having established a suitable HI, the HI profile is then subdivided into different health stages. Again, there are two broad ways of achieving this: a two-stage division into healthy and faulty states and a multi-stage division which assigns different health states as the unit progressively degrades from a healthy towards a failed state. Figure 5 shows the overall classification described in this section.

Figure 5

Fig. 5. Broad classification of health indicator construction and health stage division approaches.

Although it is useful to extract features from the data in order to gain insight into underlying trends, some original data can be used as features if they exhibit good trendability and monotonicity traits [42]. Bektas *et al.* [43] established a single health indicator trajectory by fitting a linear model using multiple linear regression directly on multi-regime degradation data, thereby performing features selection, dimensionality reduction and sensor fusion in one step. For this work, a linear model is fit onto the data output from phase 1, described in Section 3.1. To achieve this, we will calculate the PFIF, which essentially provides information regarding the state of health of each unit at any time instance, t. For this purpose, t corresponds to ℓ, the index of any given data point as operation progresses from $\ell = 1$ until failure at $\ell = T_i$. For any run-to-failure data, the PFIF for unit i at any time index, ℓ, is therefore given by Eq. (16):

$$PFIF_{i,\ell} = \frac{T_i - \ell}{T_i} \quad (16)$$

Using the values of the vector $PFIF_i$ as a response variable and also the variables in the data $X_{clusterdata,i}$, a simple linear model, which is given by Eq. (17), is fit to the data:

$$PFIF_i = \theta_0 + \theta_1 x_{i,1} + \theta_2 x_{i,2} + \theta_3 x_{i,3} + \ldots + \theta_s x_{i,s} \quad (17)$$

where θ_0 is the bias term, $\theta_1, \ldots, \theta_s$ are the model coefficients and $x_{i,s}$ are vectors representing the columns of $X_{clusterdata,i}$. In a vectorized form, we have:

$$PFIF_i = \theta_0 + X_{clusterdata,i}\theta, \quad \theta = [\theta_1\ ;\ \theta_2\ ;\ \ldots ;\ \theta_s]\ ,\ X_{clusterdata,i} \text{ is an } m \text{ by } N \text{ array of data} \quad (18)$$

The test data represents data from presently running units similar to those whose run-to-failure data were used to train a linear model and construct health indicators. Preparing the test data in a similar way as described in Section 3.1 yields the data, $XTest_{clusterdata,i}$. Applying the trained linear model on this data produces the health states at every time instance up till the present time index, ℓ_i, for each individual unit in the fleet. The health states at the present time can then be extracted and grouped together based on similar health states. For this study, four health states are defined based on the PFIF values, which mostly lie in the range [0, 1], with one being perfectly healthy units and 0 being failed equipment. A multi-scale health stage (HS) division was adopted using the following criteria: above 0.75 – "healthy"; above 0.50 up to 0.75 – "good - no action"; above 0.30 up to 0.50 – "good – monitor"; and, 0.30 and below – "soon-to-fail". A three-stage HS division was also implemented with the following window boundaries: above 0.75 – "healthy"; above 0.45 up to 0.75 – "good"; and, 0.45 and below – "soon-to-fail". Life-extension engineers may use expert judgment, and based on the peculiarity of the fleet, to assign different HS divisions. Equipment grouped together based on similar HS assignments can then be prioritized together for life-extension action and other associated logistics purposes.

3.3 *Phase 2, route 2 – k-means clustering using fleet data*

As an alternative to fitting a linear model to the data, a clustering algorithm can be used to group the units. Clustering is implemented after feature engineering and dimensionality reduction on the training data, thus identifying the trendable variables that are important condition indicators. The data that provides information regarding the current health state for each equipment is the last entry in the time-series for each unit. As such, the last row for each unit, i, corresponding to the operational stage or time index, $\max(\ell_i)$, is extracted, producing a reduced data, $XTest_{data4k-means}$, which is an m by N array, where m is the number of units in the fleet and N is the number of selected trendable sensors. A k-means algorithm is then applied on the data $XTest_{data4k-means}$, specifying the number of clusters to be equal to the desired number of health stages. The overall flow of the proposed technique, covering phase 1, phase 2 route 1 and phase 2 route 2, is illustrated in Fig. 6. It is important to note that route 2 of phase 2 in this technique is a not as amenable to user specification as route 1, where users can make choices regarding the type of algorithm to use for fitting the regression model and the level of accuracy to aim for, including the use of non-linear models to obtain model parameters that yield better predictions. Using route 2, only the number of clusters (and their respective centroids) can be specified, which corresponds to the number of divisions in the multi-stage HS division.

Figure 6

Fig. 6. Methodological approach for determining the most vulnerable equipment for life-extension.

4. Case studies

To demonstrate the feasibility and applicability of the proposed technique, it is tested on the NASA's publicly available database called C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) [44].

4.1 *Data description*

C-MAPSS comprises four different run-to-failure datasets under varying combinations of fault modes and operational conditions. The training sets all start from a healthy state and terminate at the failure point of each unit. The test set starts from a healthy state and is terminated at some unknown point during each unit's lifetime. For more details about the dataset, the readers can refer to Saxena *et al.* [45]. One of the datasets, FD001, is for a homogeneous fleet comprising run-to-failure data from 100 identical turbofan engines, with one failure

mode and under one set of operating conditions. Each of the 100 engine units has a distinct lifetime, t_i, three columns representing operating conditions settings and another 21 columns representing sensor data. The dataset, which comes as a numerical array organized as described in section 3 phase 1, is ordered as presented in Table 3.

****Table 3****

Table 3. C-MAPSS dataset parameters and corresponding variables assigned

4.2 *Application of the proposed technique*

This section describes the application of the proposed technique on the dataset C-MAPSS FD001.

4.2.1 Phase 1 – data preparation and sensor selection

Data from some sensors are directly eliminated by observing some features of the data, such as the mean and the variance. Constant value data with near zero variances are eliminated as they do not provide any useful information regarding the condition of the units under observation. This step reduces the data, X, from 21 sensors to the data $X_{reduced}$, comprising 14 sensors. The sensors that exhibit some variance, which are contained in $X_{reduced}$, are sensors 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20 and 21. The data, $X_{reduced}$, is then organized unit-by-unit as an ensemble of data for each unit, after which they are normalized using the standardization approach to obtain $X_{ireduced}$ and then smoothed using the RLOWESS algorithm.

To achieve further dimensionality reduction while ensuring that the most trendable sensors are retained for the construction of health indicators for each unit, the trendability, monotonicity and prognosability metrics are computed using the formulae in Eq. (11), Eq (12) and Eq. (15) respectively. Fig. 7(a), (b) and (c) show respectively three plots of trendability, monotonicity and prognosability metrics values obtained for 16 sensors. Figure 7(d) also shows the combined values, which are the sum of trendability, monotonicity and prognosability values.

****Figure 7****

Fig. 7. The values for (a) trendability (b) monotonicity (c) prognosability and the combined metrics for 16 sensors.

To arrive at the final set of sensors to be fused to obtain the health indicators, the values of the three metrics are combined to obtain the plot showed in Fig. 7(d). The individual plots, as well as the combined plot, show that sensors 8, 9, 13, and 14 consistently exhibit the lowest trendability traits. Consequently, based on the exclusion criterion defined in Eq. (15), these sensors were discarded using the exclusion criterion *fitness* > 2.0, yielding the data $X_{clusterdata}$, comprising the 10 selected sensors of 2, 3, 4, 7, 11, 12, 15, 17, 20 and 21. Fig. 8 shows the degradation trend of the 10 selected sensors for the first three units.

****Figure 8****

Fig. 8. Degradation trend for 10 selected sensors on units 1, 2 and 3.

4.2.2 Phase 2, route 1 – construct health indicator and implement HS division

In order to fit a regression model to the preprocessed training data, $X_{clusterdata}$, the PFIF is computed using the formula provided in Eq. (16). The degradation trajectory for each unit, i, runs from $\ell = 1$ cycle to $\ell = T_i$ cycles, where T_i corresponds to the time index at which the trajectory is terminated (i.e. upon failure of the unit). These values are used to calculate the P-F interval and then the PFIF index, which is added as a column to the data $X_{clusterdata}$ and used as the response variable for fitting the regression model to the data.

A least squares regression model is fit to the data using MATLAB, to obtain the bias term, θ_0 and the model coefficients, θ_s. Values were averaged from two runs of the MATLAB code that produced good fits of the model, to give θ_o = 0.5019 and θ = [-0.0300; -0.0199; -0.0471; 0.0466; -0.0622; 0.0573; -0.0365; -0.0188; 0.0314; 0.0369]. Using the model, the ten selected sensors are fused together to construct a single health indicator. The health indicators, some of which were predominantly monotonically increasing while others were predominantly monotonically decreasing, were all offset to start from one and then decrease progressively until failure. A visualization of the constructed HIs for all 100 units within the fleet is shown in Fig. 9.

Figure 9

Fig. 9. Constructed HIs using trained data for all 100 units within the fleet.

Following the procedure outlined in Section 3.2, the test data, which comprises data for the 100 units up to an undefined time, are imported into MATLAB and pre-processed to obtain $XTest_{clusterdata,i}$. The trained linear regression model is then used on $XTest_{clusterdata,i}$ to predict the HIs for each of the 100 units in the test dataset. A plot of the HIs for the first 20 units in the test data is shown in Fig. 10. It can be observed from Fig. 10 that the trajectories for most of the units end abruptly. Extracting the HIs at the end of each trajectory gives the current health state of each unit. Equipment with the same health state can then be grouped together for the purpose of life-extension decision-making.

Figure 10

Fig. 10. Constructed HIs for the 20 units using the test dataset.

4.2.3 Phase 2, route 2 – Group units using k-means clustering

The fundamental goal of the proposed technique is to achieve grouping of equipment with similar health states so as to prioritize the most vulnerable equipment for life-extension actions. An alternative way to achieve this grouping is to apply a clustering algorithm, after pre-processing the data and selecting the most trendable features or sensors. From Fig. 10, it was established that the important indicator of the current health state for each unit is the last point in the data for each unit, corresponding to the point where each degradation trajectory ends. So, by extracting the last data point for each unit from the pre-processed test data, $XTest_{clusterdata,i}$ we obtain the data $XTest_{data4k-means}$. A *k*-means algorithm is then run using random initialization for ten replicates with 100 iterations in each replicate and square Euclidian distance as the distance measure. The number of clusters is set to four and then to three, for four-stage HS division and three-stage HS division respectively. This will produce groups of units that should have similar health states and thus help to prioritize life-extension decision-making. Section 5 presents the results obtained and discusses the findings.

5. Results and discussion

This section presents the results obtained for algorithms implemented to perform three-stage and four-stage HS divisions. The results obtained using phase 2, route 1 of the technique (i.e. using a linear regression model) are compared with those obtained using phase 2, route 2 (i.e. using *k*-means clustering). Since the dataset comes with ground truth RUL values, the predicted PFIF results, which were mostly in the range [0, 1], were easily compared to the normalized values of the true PFIF values which were calculated as follows.

$$True\ PFIF_{i,\ell} = \frac{True\ RUL_i}{True\ RUL_i + \ell} \tag{19}$$

$$Normalized\ True\ PFIF_{i,\ell} = \frac{True\ PFIF_{i,\ell} - min\ (True\ PFIF_{i,\ell})}{max(True\ PFIF_{i,\ell}) - min\ (True\ PFIF_{i,\ell})} \quad (20)$$

Fig. 11 is a plot of the predicted PFIF using the regression model, against the normalized true PFIF for each unit, and it shows a very good match between the predicted values and the ground truth values.

Figure 11

Fig. 11. Comparison of predicted and true health indices.

5.1 *Three-stage HS division*

The grouping of equipment was implemented by setting window boundaries based on the predicted PFIF values in order to establish health states. The sub-sections below will present the results for the different health states.

5.1.1 Healthy units

The results obtained using both the regression model and *k*-means clustering are presented side by side in Table 4 for healthy units, for both the 3-stage and 4-stage HS divisions. For the three-stage HS division, it can be observed from the results that both the linear model and the *k*-means clustering algorithm grouped 27 out of 29 units as healthy. The other two units were grouped by the *k*-means algorithm as "good". Also, the predicted as well as the normalised true PFIF values show that 83% of the healthy units have true PFIF values above 0.65; this corresponds to the units having spent only 35% of their lifetimes, with 65% of their lifetimes left. Given that the application of this work is for life-extension, and that healthy units are grouped as mostly "healthy", with a few as "good", this translates to 100% acceptable grouping.

Table 4

Table 4. Healthy units grouping for both 3-stage and 4-stage HS division (Number of units: 29)

For both the three-stage and four-stage HS divisions, it can be observed from Table 4 that group 2 of the *k*-means clustering corresponds to the "healthy" units. For the four-stage HS division, it was observed that the match between the group assignments when using the regression model as compared to when using the *k*-means clustering approach was not consistent. This is because many of those units grouped as "good" and "healthy" were assigned to one of the groups when using the regression model and to other groups when *k*-means clustering is used. This is completely okay since the intent of this grouping in particular, and of prognostics in general, is to identify equipment that are about to fail before they actually fail. In that regard, equipment in a good state of health identified as such is not a cause for concern.

5.1.2 *Good units*

Table 5 presents the results for "good" units' assignments for the three-stage HS division. In this case, group 1 of the *k*-means clustering corresponds to "good" units from the regression model. 17 out of 31 units were clustered as "good" by both approaches, while the *k*-means algorithm grouped another 13 as "healthy." Only one unit was grouped by the *k*-means algorithm as "soon-to-fail." Again, in the context of life-extension, if an equipment in a "good" state is wrongly categorised as "soon-to-fail," there are no serious safety implications, even though there may be some associated logistics or cost implications. In terms of PFIF accuracy, 5 out of 31 units grouped as "good" have true PFIF values below 0.4 (i.e. less than 40% of their lifetime is left). This gives a grouping "accuracy" of about 84%.

Table 5

Table 5. Good units grouping for 3-stage HS division (Number of units: 31)

5.1.3 Soon-to-fail

There is a good match between units grouped as "soon-to-fail" by using both approaches. The results presented in Table 6 for "soon-to-fail" units show that of the 40 units assigned to this group by the regression model, 31 were also assigned to the same group by the *k*-means clustering approach. The *k*-means approach assigned the other 9 units to group 1, which corresponds to "good" units. This is the main area of concern in terms of safety, reliability and availability; when "soon-to-fail" units are grouped as "good" units. However, the true PFIF values show that 39 out of the 40 units have values below 0.4 (i.e. all units have less than 40% of their lifetimes left). As such, the regression model has 97.5% "accuracy" in grouping. Looking at assignments using only the *k*-means approach, 35 units were actually grouped as "soon-to-fail," with only two of them having true PFIF values above 0.4. This gives an "accuracy" of about 94% in grouping.

Table 6

Table 6. "Soon-to-fail" units grouping for 3-stage HS division (Number of units: 40).

5.2 *Four-stage HS division*

For the four-stage HS divisions, a different set of window boundaries, defined in Section 3.2, was set for the regression model while the parameter, *k*, was assigned a value of 4 for the *k*-means clustering approach. The results obtained are presented in the following sub-sections.

5.2.1 Healthy units

Given that the cut-off threshold for healthy units was set at values of predicted PFIF > 0.75 for both the three-stage and the four-stage HS divisions, the results obtained for "healthy" units for the regression model were the same. However, since the *k*-means clustering approach now has $k = 4$, the expectation was that a slightly different unit assignments will be obtained. As such, while the regression algorithm grouped 29 units as "healthy," the *k*-means approach grouped 22 units as "healthy". Details of the results have been presented and discussed in Section 5.1.1.

5.2.2 Good units – no action

One of the intents behind the four-stage HS division is to distinguish between units that have recorded very minimal degradation and those that have significant degradation but are still okay to be operated. Units with minimal degradation are grouped as "good – no action." Using the specified window boundaries, 26 out of the 100 units were extracted and grouped as "good – no action." Out of these, the *k*-means approach grouped 10 in the same category, nine as "good – monitor," four as "healthy" and one as "soon-to-fail. However, an analysis of the true PFIF values for the units show that 2 units have PFIF values below 0.4, giving an "accuracy" of about 92% in grouping. Considering only the results for the *k*-means approach, there was no clear distinction between the groups "good – no action" and "healthy" as equipment having true PFIF values within the appropriate ranges were almost equally grouped into both health stages.

Table 7

Table 7. "Good – no action" groupings for 4-stage HS division (Number of units: 26)

5.2.3 Good units – monitor

Table 8 presents the results obtained from using the regression model to group equipment as "good – monitor" along with the assignments using *k*-means for the same set of units. One of the units has a very low true PFIF value of 0.1850, implying that it was wrongly grouped and should have been grouped as "soon-to-fail." The *k*-means clustering approach assigned 11 out of the 16 units to the same group, four units were assigned as "soon-to-fail" while one was grouped as "good – no action". Assignment accuracies based on true PFIF values are presented in summary in Table 8.

Table 8

Table 8. "Good – monitor" groupings for 4-stage HS division (Number of units: 16)

5.2.4 Soon-to-fail

Similar to the three-stage HS division, there is a good match between both approaches in grouping units as "soon-to-fail." The results in Table 9 show that 26 out of 29 units were assigned to this group by both approaches, while the *k*-means algorithm assigned 3 equipment to the group "good – monitor." This is an undesirable result given that all units due to fail soon should be identified. Considering only the *k*-means assignments, 32 units were assigned as "soon-to-fail," out of which only one unit had a true PFIF value above 0.4. This translates to about 97% "accuracy" in grouping.

Table 9

Table 9. "Soon-to-fail" groupings for 4-stage HS division (Number of units: 29)

As mentioned earlier, it is very important that no unit close to failure is grouped as either "healthy" or "good" as it will lead to unexpected failures. The results for both the three-stage and the four-stage HS division using both the regression model and the *k*-means algorithm show reasonably high classification accuracies based on the true PFIF values.

5.3 *Summary of results*

The summary of the entire groupings using both approaches and for the different multi-stage HS divisions is presented in Table 10.

Table 10

Table 1. Summary of group assignments and accuracies (Note that percentages are based on number of units with true PFIF values within suitable thresholds).

In general, the *k*-means algorithm performed better for three-stage HS divisions. The *k*-means approach could not clearly distinguish between the division of "good units" into "no action" and "monitor" categories. However, to attain a better-defined grouping accuracy, the regression model is the proffered approach, since the window boundaries are user-defined. What must be noted is the importance of defining the window boundaries for different health states based on sound understanding on the technical details of the units.

5.4 *Importance of experts' judgements and other considerations*

While the proposed technique has been demonstrated to produce consistent results, it is important to note a few important points. Machine learning approaches to solving engineering problems have been generally considered as black-box approaches due to the fact that it is difficult to explain the models used in clear and

specific mathematical terms. However, the reality of complex systems and the ubiquitous availability of data make their use inevitable. Therefore, experts' judgements must be used to gauge the results before implementation. Fig. 6, which gives the overall flow of the proposed approach, factors in the important role of experts' judgments. For instance, an equipment from a particular manufacturer which is known to have certain maintenance requirements, in spite of available data, must be considered irrespective of its grouping. Additionally, in order to cluster the equipment for life-extension actions (repair, upgrade or replacement), some other operational realities such as minimum downtime required to execute actions, safety implications, economic implications, etc. also need to be considered. Moreover, operational and environmental uncertainties like terrain (whether onshore or offshore), lead times for ordering of spares parts and logistics requirements for repairs all need to be factored into the decision-making framework.

6. Conclusion and future work

The fundamental theory behind data mining concepts have been around for a while now. Also, the practice of RCM and the use of P-F curves by maintenance and reliability engineers and specialists are well established. This work developed and implemented a technique that harnessed concepts from both fields, factoring in the recent rapid advances in sensor technologies and data collection capabilities, to help group and prioritize equipment within a homogeneous fleet for life-extension actions. This is a novel combination of both concepts and the results presented a remarkable consistency. For asset managers and decision makers, this is potentially an important tool that will help with better-informed and data-driven logistics planning and spare parts management. Much better grouping results can be achieved by using more accurate models which may include adding regularization to the regression model or formulating a more rigorous approach for establishing the window boundaries for use with the potential failure interval factor (PFIF).

The methodology for assessing the accuracy or suitability of unit assignments into groups can be formulated via a mathematically rigorous approach rather than just mere counts and comparison to the true PFIF as used in this work. Such a mathematical formulation, which is an area for future research, may in fact include the modelling of uncertainties into the accuracy of unit assignments. Furthermore, this work only considered identical units under the same operational settings for a single failure mode. It can be further extended and applied to a heterogeneous fleet with dissimilar units under varying loading conditions, different operational settings and multiple failure modes. Another important area of work will be a look at how life-extension actions carried out for any unit or group of units influence the continuous and ongoing use of the model. If, for instance, a life-extension action involves an upgrade and a replacement, it will be interesting to know how it affects the model in terms of base data availability for the affected unit and availability of specific sensors for additional or continuous data acquisition.

In terms of application, this work is essential for identifying and prioritizing vulnerable equipment for life-extension. It adds to the repertoire of models, tools and decision support systems available to asset managers and reliability engineers. Feedback from the proposed process can potentially serve as useful input for plant and equipment design for longevity and also influence original equipment manufacturer (OEM) sensor placement philosophies.

Acknowledgement

The first author wishes to thank the Petroleum Technology Development Fund (PTDF) in Nigeria for sponsoring his PhD research.

References

[1] M. Shafiee, Modelling and analysis of availability for critical interdependent infrastructures, Int. J. Risk Assess. Manag. 19 (2016) 299–314. https://doi.org/10.1504/IJRAM.2016.079608.

[2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition

to RUL prediction, Mech. Syst. Signal Process. 104 (2018) 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016.
[3] P. Vaidya and M. Rausand, Remaining useful life, technical health, and life extension, Proc. Inst. Mech. Eng. Part O: J. Risk Reliab. 225 (2011) 219-231. https://doi.org/10.1177/1748007810394557.
[4] J. V. Sharp, E.G. Terry, J. Wintle, A framework for the management of ageing of safety critical elements offshore, in: ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering, June 19–24, Rotterdam, The Netherlands, pp. 141–153. https://doi.org/10.1115/omae2011-49203.
[5] M. Shafiee, I. Animah, Life extension decision making of safety critical systems: An overview, J. Loss Prev. Process Ind. 47 (2017) 174–188. https://doi.org/10.1016/j.jlp.2017.03.008.
[6] G. Ersdal, E. Hörnlund, H. Spilde, Experience from Norwegian programme on ageing and life extension, in: ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering, June 19–24, Rotterdam, The Netherlands, pp. 517–522. https://doi.org/10.1115/omae2011-50046.
[7] P.A.P. Ramírez, I.B. Utne, C. Haskins, Application of systems engineering to integrate ageing management into maintenance management of oil and gas facilities, Syst. Eng. 16 (2013) 329–345. https://doi.org/10.1002/sys.21257.
[8] M. Shafiee, I. Animah, N. Simms, Development of a techno-economic framework for life extension decision making of safety critical installations, J. Loss Prev. Process Ind. 44 (2016) 299–310. https://doi.org/10.1016/j.jlp.2016.09.013.
[9] I. Animah, M. Shafiee, N. Simms, J.A. Erkoyuncu, J. Maiti, Selection of the most suitable life extension strategy for ageing offshore assets using a life-cycle cost-benefit analysis approach, J. Qual. Maint. Eng. 24 (2018) 311–330.
[10] S.F. Nowlan, Planning and operational aspects of “on condition” philosophies, Aircr. Eng. Aerosp. Technol. 44 (1972) 26–28. https://doi.org/10.1108/eb034882.
[11] J. Moubray, Reliability-centered maintenance, Second Edi, Butterworth-Heinemann, 1997.
[12] M. Shafiee, Maintenance strategy selection problem: An MCDM overview, J. Qual. Maint. Eng. 21 (2015) 378–402. https://doi.org/10.1108/JQME-09-2013-0063.
[13] W. Liang, L. Pang, L. Zhang, J. Hu, Reliability-centered maintenance study on key parts of reciprocating compressor, in: Proc. 2012 Int. Conf. Qual. Reliab. Risk, Maintenance, Saf. Eng. 15-18 June 2012, Chengdu, China. https://doi.org/10.1109/ICQR2MSE.2012.6246265.
[14] A. Van Horenbeek, J. Van Ostaeyen, J.R. Duflou, L. Pintelon, Quantifying the added value of an imperfectly performing condition monitoring system - Application to a wind turbine gearbox, Reliab. Eng. Syst. Saf. 111 (2013) 45–57. https://doi.org/10.1016/j.ress.2012.10.010.
[15] A. Lorenzoni, M. Kempf, O. Mannuß, Degradation model constructed with the aid of dynamic Bayesian networks, Cogent Eng. 4 (2017) 1–12. https://doi.org/10.1080/23311916.2017.1395786.
[16] N. Regan, RCM Solution - A practical guide to starting and maintaining a successful RCM program, Industrial Press, Inc. ISBN-13: 978-0831134242, 2012.
[17] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (2006) 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012.
[18] K.B. Goode, J. Moore, B.J. Roylance, Plant machinery working life prediction method utilizing reliability and condition-monitoring data, Proc. Inst. Mech. Eng. Part E J. Process Mech. 214 (2000) 109–122. https://doi.org/10.1243/0954408001530146.
[19] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press, Third Edition, ISBN: 978-1-119-51604-0, 2019.
[20] G.J. Myatt, Making Sense of Data: A practical guide to exploratory data analysis and data mining, John Wiley and Sons, Inc., ISBN: 978-0470074718, 2006. https://doi.org/10.1002/0470101024.
[21] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Third Edit, Elsevier, ISBN 978-0-12-381479-1, 2012. https://doi.org/10.1016/C2009-0-61819-5.
[22] P. Zhao, M. Kurihara, T. Noda, H. Kashiwa, M. Hiyama, T. Suzuki, Equipment sub-system extraction and its application in predictive maintenance, in: 2018 IEEE Int. Conf. Progn. Heal. Manag., 2018: pp. 1–5. https://doi.org/10.1109/ICPHM.2018.8448603.
[23] A. Abdelhadi, Preventive maintenance operations scheduling based on eigenvalue and clustering methods, 2019 IEEE 6th Int. Conf. Ind. Eng. Appl. ICIEA 2019. (2019) 183–186. https://doi.org/10.1109/IEA.2019.8715069.
[24] A.K. Jain, Data clustering: 50 years beyond *k*-means, Pattern Recognit. Lett. 31 (2010) 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.
[25] P. Gholami, A. Hafezalkotob, Maintenance scheduling using data mining techniques and time series models, Int. J. Manag. Sci. Eng. Manag. 13 (2018) 100–107. https://doi.org/10.1080/17509653.2017.1314201.
[26] A. Lahrache, M. Cocconcelli, R. Rubini, Anomaly detection in a cutting tool by *k*-means clustering and Support Vector Machines, Diagnostyka. 18 (2017) 21–29.
[27] A. Abdelhadi, Heuristic approach to schedule preventive maintenance operations using *k*-means methodology, Int. J. Mech. Eng. Technol. 8 (2017) 300–307.
[28] J.M. Wakiru, L. Pintelon, V. V. Karanović, M.T. Jocanović, M.D. Orošnjak, Analysis of lubrication oil towards

maintenance grouping for multiple equipment using fuzzy cluster analysis, IOP Conf. Ser. Mater. Sci. Eng. 393 (2018) 012011. https://doi.org/10.1088/1757-899X/393/1/012011.
[29] S. Al-Dahidi, F. Di Maio, P. Baraldi, E. Zio, Remaining useful life estimation in heterogeneous fleets working under variable operating conditions, Reliab. Eng. Syst. Saf. 156 (2016) 109–124. https://doi.org/10.1016/j.ress.2016.07.019.
[30] G. Medina-Oliva, A. Voisin, M. Monnin, J.B. Leger, Predictive diagnosis based on a fleet-wide ontology approach, Knowledge-Based Syst. 68 (2014) 40–57. https://doi.org/10.1016/j.knosys.2013.12.020.
[31] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, F. Herrera, A survey on data preprocessing for data stream mining: Current status and future directions, Neurocomputing. 239 (2017) 39–57. https://doi.org/10.1016/j.neucom.2017.01.078.
[32] J. Zhu, T. Nostrand, C. Spiegel, B. Morton, Survey of condition indicators for condition monitoring systems, PHM 2014 - Proc. Annu. Conf. Progn. Heal. Manag. Soc. 2014. (2014) 1–13.
[33] Y. Liu, X. Hu, W. Zhang, Remaining useful life prediction based on health index similarity, Reliab. Eng. Syst. Saf. 185 (2019) 502–510. https://doi.org/10.1016/j.ress.2019.02.002.
[34] C.C. Aggarwal, Data mining: the textbook, Springer, 2015. https://doi.org/10.1007/978-3-319-14142-8.
[35] W.S. Cleveland, S.J. Devlin, E. Grosse, Regression by local fitting: Methods, properties, and computational algorithms, J. Econom. 37 (1988) 87–114. https://doi.org/https://doi.org/10.1016/0304-4076(88)90077-2.
[36] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (1979) 829–836. https://doi.org/10.1080/01621459.1979.10481038.
[37] J. Coble, J.W. Hines, Fusing data sources for optimal prognostic parameter selection, Transactions of the American Nuclear Society 100 (2009) 211–212.
[38] J. Coble, J.W. Hines, Identifying optimal prognostic parameters from data: A genetic algorithms approach, in: Annu. Conf. Progn. Heal. Manag. Soc. PHM 2009.
[39] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing. 240 (2017) 98–109. https://doi.org/10.1016/j.neucom.2017.02.045.
[40] D. Wang, K.L. Tsui, Q. Miao, Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators, IEEE Access. 6 (2017) 665–676. https://doi.org/10.1109/ACCESS.2017.2774261.
[41] V. Atamuradov, K. Medjaher, F. Camci, N. Zerhouni, P. Dersin, B. Lamoureux, Machine health indicator construction framework for failure diagnostics and prognostics, J. Signal Process. Syst. 92, (2020) 591–609. https://doi.org/10.1007/s11265-019-01491-4.
[42] T. Wang, J. Yu, D. Siegel, J. Lee, A similarity-based prognostics approach for remaining useful life estimation of engineered systems, in: 2008 Int. Conf. Progn. Heal. Manag. 6-9 Oct. 2008, Denver, CO, USA. https://doi.org/10.1109/PHM.2008.4711421.
[43] O. Bektas, A. Alfudail, J.A. Jones, Reducing dimensionality of multi-regime data for failure prognostics, J. Fail. Anal. Prev. 17 (2017) 1268–1275. https://doi.org/10.1007/s11668-017-0368-2.
[44] A. Saxena and K. Goebel, Turbofan engine degradation simulation dataset, NASA Ames Progn. Data Repos. (2008). http://ti.arc.nasa.gov/project/prognostic-data-repository (accessed February 4, 2020).
[45] A. Saxena, K. Goebel, D. Simon, N. Eklund, Damage propagation modeling for aircraft engine run-to-failure simulation, in: 2008 Int. Conf. Progn. Heal. Manag. 6-9 Oct. 2008, Denver, CO, USA. https://doi.org/10.1109/PHM.2008.4711414.

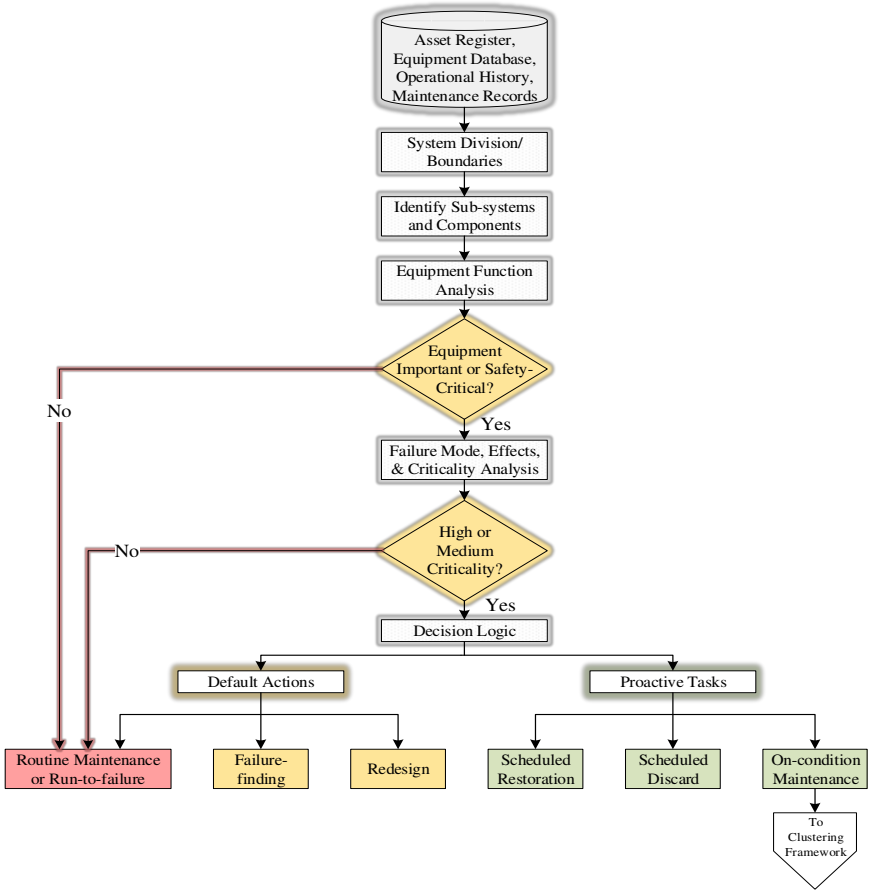

Fig. 1. RCM decision logic flowchart – adapted from Liang *et al.* (2012).

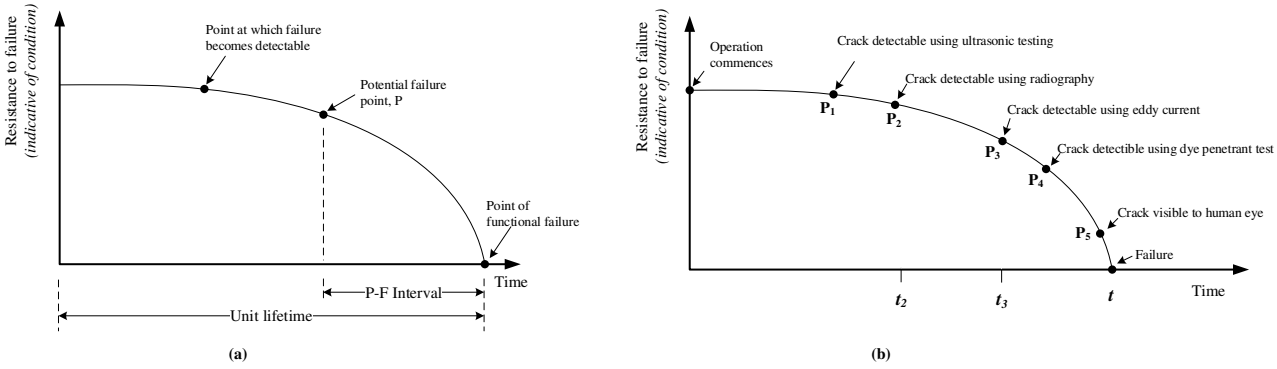

Fig. 2. (a) A typical P-F curve, **(b)** A P-F curve for fatigue crack propagation (adapted from Regan (2012))).

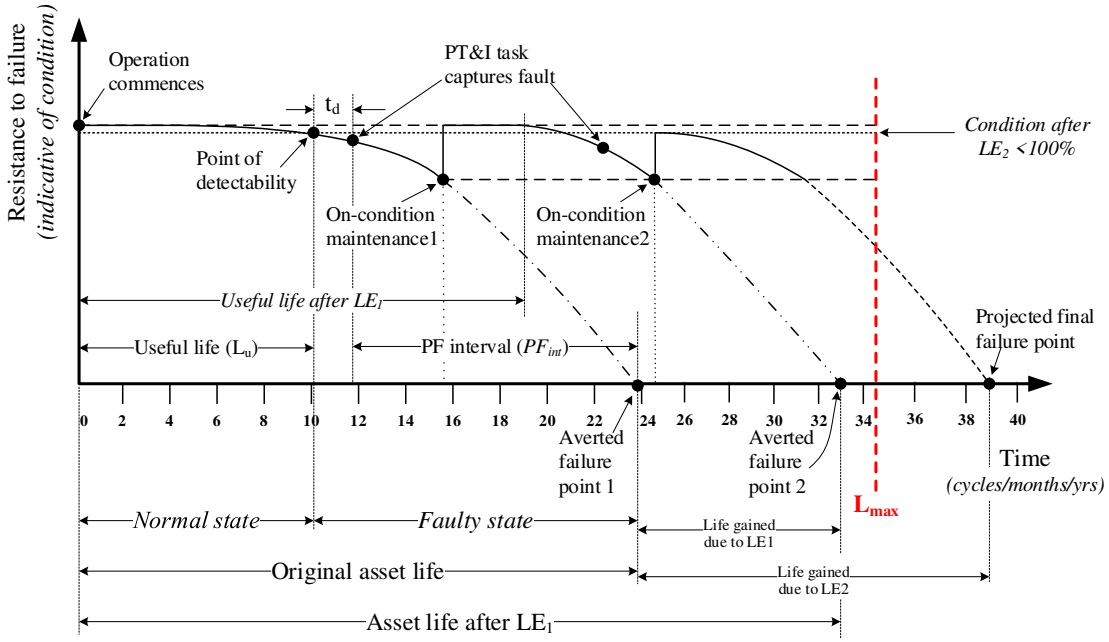

Fig. 3. Effect of a life-extension action on P-F curve.

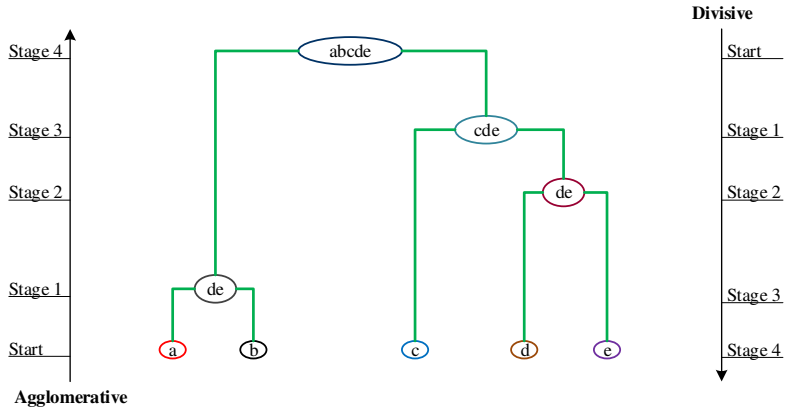

Fig. 4. Dendrogram for the two types of hierarchical clustering – adapted from Han *et al.* (2012).

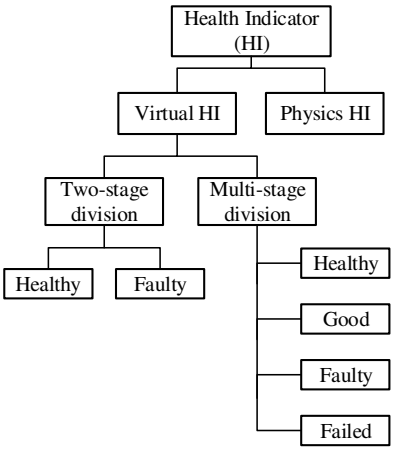

Fig. 5. Broad classification of health indicator construction and health stage division approaches.

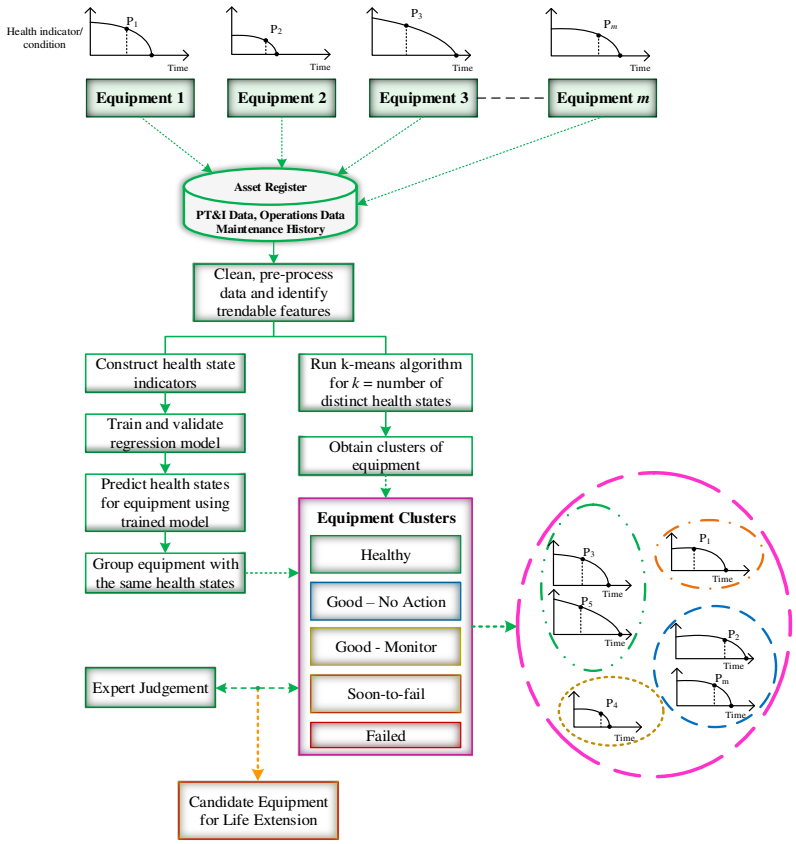

Fig. 6. Methodological approach for determining the most vulnerable equipment for life-extension.

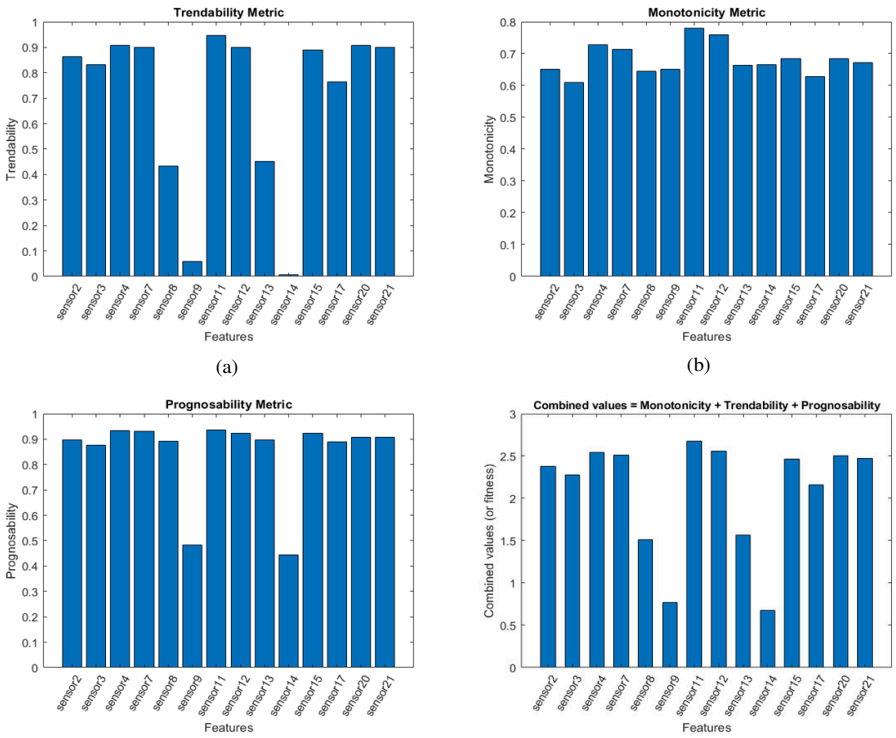

Fig. 7. The values for (a) trendability (b) monotonicity (c) prognosability and the combined metrics for 16 sensors.

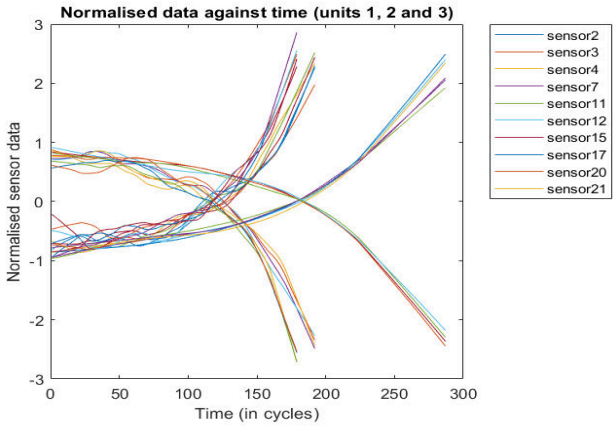

Fig. 8. Degradation trend for 10 selected sensors on units 1, 2 and 3.

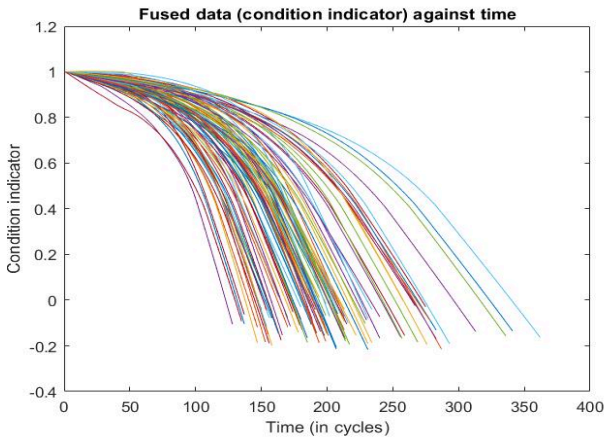

Fig. 9. Constructed HIs using trained data for all 100 units within the fleet.

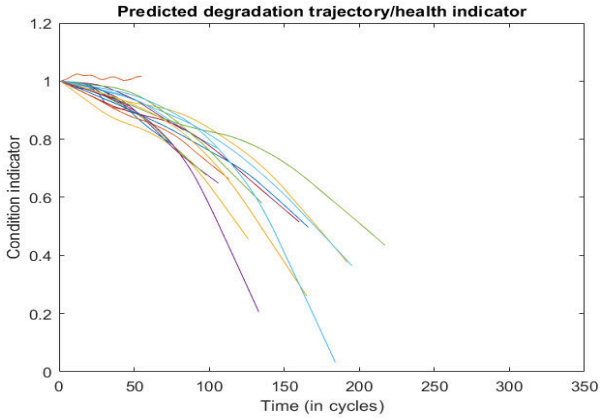

Fig. 10. Constructed HIs for the 20 units using the test dataset.

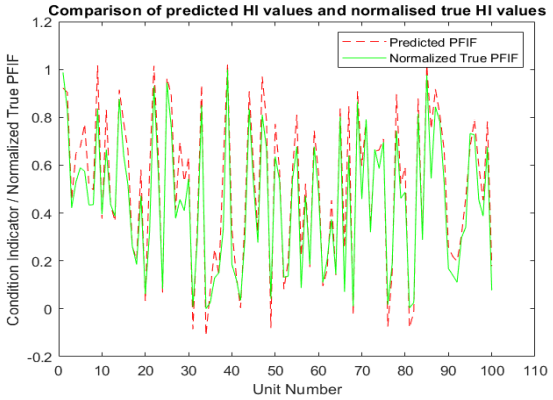

Fig. 11. Comparison of predicted and true health indices.

Table 1. RCM strategies and their associated application scenarios.

Maintenance strategy	Application scenario
No maintenance or run-to-failure	a. Failure of equipment/item has no safety or environmental consequences. b. The economic consequence of failure is also negligible or tolerable.
Failure-finding	a. Failure of equipment has no immediate obvious consequence. b. Equipment typically has a backup protective safety device which can fail without being immediately evident.
Redesign	a. Equipment whose behavior may not be fully known. b. No known maintenance action will reduce the probability of failure. c. Cost of known maintenance action outweighs economic consequence of failure and failure is not negligible or tolerable – redesign or redundancy becomes the option.
Scheduled discard	a. Non-repairable items, e.g. pump impellers, seals, valve seats, etc. b. Involves replacing and discarding equipment without regard to condition (as in conventional preventive maintenance).
Scheduled restoration	a. Repairable equipment/items. b. Suitable on-condition tasks cannot be devised to avert potential failure. c. Involves overhauling or repairing items without regard to condition (as in conventional preventive maintenance).
On-condition maintenance	a. Degradable equipment/items. b. Condition indicators are known and can be monitored using sensor data or other PT&I techniques.

Table 2. Pros, cons and application cases for the two broad classes of clustering algorithms

Clustering type	Pros	Cons	Application cases
Hierarchical clustering (agglomerative and divisive)	a. No overlaps between clusters. b. Can be applied to more variety of data than *k*-means. c. Typically yields a unique dendrogram (repeatable).	a. Applicable to relatively small datasets (<10,000 observations). b. Generating the hierarchical tree can be slow. c. Can handle outliers well. d. Does not follow a scale.	System and subsystems predictive maintenance Zhao *et al.* (2018); grouping maintainable equipment Abdelhadi (2019).
Partitional clustering (*k*-means and its variants)	a. Computationally faster. b. Can handle a larger number of observations than hierarchical clustering. c. Clusters are clearly defined without overlaps. d. Scalable as it is based on actual numerical data.	a. Difficulty in predefining optimal number of clusters. b. Can be distorted by outliers. c. Works only with numerical data. d. Not repeatable. Random initialization potentially results in varying clusters.	Grouping maintenance activities Gholami & Hafezalkotob (2018); Fault type clustering Lahrache *et al.*, (2017); Maintenance planning optimization (Jain, 2010; Gholami & Hafezalkotob, 2018). RUL estimation for heterogeneous fleet (Al-Dahidi *et al.*, 2016)

Table 3. C-MAPSS dataset parameters and corresponding variables assigned

# column	Measured parameter	Unit of measurement	Variable assigned *(for this work)*
1	Unit number	--	unit_num
2	Time	cycles	Time
3	Operational setting 1	--	ops_set1
4	Operational setting 2	--	ops_set2
5	Operational setting 3	--	ops_set3
6	Total temperature at fan inlet	°R	sensor1
7	Total temperature at LPC outlet	°R	sensor2
8	Total temperature at HPC outlet	°R	sensor3
9	Total temperature at LPT outlet	°R	sensor4
10	Pressure at fan inlet	psia	sensor5
11	Total pressure in bypass-duct	psia	sensor6
12	Total pressure at HPC outlet	psia	sensor7
13	Physical fan speed	rpm	sensor8
14	Physical core speed	rpm	sensor9
15	Engine pressure ratio (P50/P2)	--	sensor10
16	Static pressure at HPC outlet	psia	sensor11
17	Ratio of fuel flow to Ps30	pps/psi	sensor12
18	Corrected fan speed	rpm	sensor13
19	Corrected core speed	rpm	sensor14
20	Bypass Ratio	--	sensor15
21	Burner fuel-air ratio	--	sensor16
22	Bleed Enthalpy	--	sensor17
23	Demanded fan speed	rpm	sensor18
24	Demanded corrected fan speed	rpm	sensor19
25	HPT coolant bleed	lbm/s	sensor20
26	LPT coolant bleed	lbm/s	sensor21

Table 4. Healthy units grouping for both 3-stage and 4-stage HS division (Number of units: 29)

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS (3-stage)	*k*-means HS (4-stage)
1	0.9225	0.9867	112	Healthy	Group 2	Group 2
2	0.9039	0.8334	98	Healthy	Group 2	Group 2
6	0.7707	**0.5742**	**93**	Healthy	Group 2	Group 2
9	1.0169	0.8360	111	Healthy	Group 2	Group 2
11	0.8313	0.6652	97	Healthy	**Group 1**	**Group 4**
14	0.9155	0.8764	107	Healthy	Group 2	Group 2
15	0.7795	0.6430	83	Healthy	Group 2	Group 2
22	1.0152	0.9299	111	Healthy	Group 2	Group 2
25	0.9653	0.9447	145	Healthy	Group 2	**Group 4**
26	0.8915	0.7591	119	Healthy	Group 2	Group 2
33	0.9367	0.8502	106	Healthy	Group 2	Group 2
39	1.0229	1.0000	142	Healthy	Group 2	Group 2
44	0.9086	0.8361	109	Healthy	Group 2	Group 2
47	0.9703	0.8102	135	Healthy	Group 2	Group 2
48	0.7833	0.6682	92	Healthy	Group 2	**Group 4**
50	0.7721	0.6356	79	Healthy	**Group 1**	**Group 4**
55	0.8107	0.6772	137	Healthy	Group 2	**Group 4**
65	0.8385	0.8025	128	Healthy	Group 2	**Group 4**
67	0.8464	0.6407	77	Healthy	Group 2	Group 2
69	0.9089	0.8660	121	Healthy	Group 2	**Group 4**
71	0.7906	0.7909	118	Healthy	Group 2	**Group 4**
78	0.8961	0.7427	107	Healthy	Group 2	**Group 4**
83	0.8785	0.8146	137	Healthy	Group 2	***Group 4***
85	1.0173	0.9777	118	Healthy	Group 2	Group 2
86	0.7517	**0.5446**	**89**	Healthy	Group 2	**Group 4**
87	0.9169	0.8436	116	Healthy	Group 2	**Group 4**
88	0.8135	0.7830	115	Healthy	Group 2	Group 2
96	0.7888	0.7265	137	Healthy	Group 2	Group 2
99	0.7825	0.6756	117	Healthy	Group 2	Group 2

Table 5. Good units grouping for 3-stage HS division (Number of units: 31)

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS
3	0.4564	0.4217	69	Good	Group 1
4	0.6494	0.5301	82	Good	Group 1
5	0.6767	0.5897	91	Good	**Group 2**
7	0.5153	0.4332	91	Good	Group 1
8	0.4969	0.4351	95	Good	Group 1
16	0.6635	0.5172	84	Good	**Group 2**
19	0.5808	0.4718	87	Good	Group 1
21	0.5798	**0.3220**	**57**	Good	Group 1
23	0.6426	0.5680	113	Good	Group 1
28	0.6954	0.4567	97	Good	**Group 2**
29	0.5316	0.4099	90	Good	**Group 2**
30	0.6292	0.5427	115	Good	Group 1
38	0.6067	**0.3321**	**50**	Good	**Group 2**
45	0.5936	0.5201	114	Good	Group 1
51	0.5792	0.5376	114	Good	**Group 2**
54	0.5838	0.5416	97	Good	**Group 2**
57	0.5224	0.4715	103	Good	Group 1
59	0.7438	0.6773	114	Good	**Group 2**
60	0.5648	0.4889	100	Good	**Group 2**
63	0.4536	**0.3735**	**72**	Good	Group 1
70	0.5967	0.4589	94	Good	Group 1
73	0.6591	0.6655	131	Good	**Group 2**
74	0.6635	0.5865	126	Good	Group 1
75	0.7094	0.6959	113	Good	**Group 2**
79	0.5214	0.4616	63	Good	Group 1
80	0.5907	0.4872	90	Good	Group 1
89	0.6639	0.5279	136	Good	**Group 2**
94	0.4627	**0.3411**	**55**	Good	Group 1
95	0.6753	0.7323	128	Good	**Group 2**
97	0.5991	0.4557	82	Good	***Group 3***
98	0.4534	**0.3874**	**59**	Good	Group 1

Table 6. "Soon-to-fail" units grouping for 3-stage HS division (Number of units: 40).

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS
10	0.3764	0.3948	96	Soon-to-fail	**Group 1**
12	0.4358	**0.4346**	**124**	Soon-to-fail	Group 3
13	0.3651	0.3872	95	Soon-to-fail	Group 3
17	0.2602	0.2622	50	Soon-to-fail	Group 3
18	0.2058	0.1850	28	Soon-to-fail	Group 3
20	0.0324	0.0614	16	Soon-to-fail	Group 3
24	0.0678	0.0839	20	Soon-to-fail	Group 3
27	0.4133	0.3777	66	Soon-to-fail	**Group 1**
31	-0.0859	0.0077	8	Soon-to-fail	Group 3
32	0.3381	0.2834	48	Soon-to-fail	**Group 1**
34	-0.1119	0.0000	7	Soon-to-fail	Group 3
35	0.0754	0.0254	11	Soon-to-fail	Group 3
36	0.2490	0.1286	19	Soon-to-fail	**Group 1**
37	0.1405	0.1507	21	Soon-to-fail	Group 3
40	0.3335	0.1850	28	Soon-to-fail	Group 3
41	0.1418	0.1241	18	Soon-to-fail	Group 3
42	0.0032	0.0354	10	Soon-to-fail	Group 3
43	0.3733	0.2922	59	Soon-to-fail	Group 3
46	0.3231	0.2766	47	Soon-to-fail	**Group 1**
49	-0.0802	0.0414	21	Soon-to-fail	Group 3
52	0.0819	0.1312	29	Soon-to-fail	Group 3
53	0.1987	0.1362	26	Soon-to-fail	Group 3
56	0.2212	0.0869	15	Soon-to-fail	**Group 1**

58	0.1746	0.1847	37	Soon-to-fail	Group 3
61	0.0945	0.1097	21	Soon-to-fail	Group 3
62	0.1633	0.2046	54	Soon-to-fail	Group 3
64	0.1404	0.1441	28	Soon-to-fail	Group 3
66	0.2531	0.0706	14	Soon-to-fail	Group 3
68	-0.0256	0.0101	8	Soon-to-fail	Group 3
72	0.3421	0.3196	50	Soon-to-fail	**Group 1**
76	-0.0740	0.0173	10	Soon-to-fail	Group 3
77	0.1484	0.1844	34	Soon-to-fail	Group 3
81	-0.0803	0.0038	8	Soon-to-fail	Group 3
82	-0.0128	0.0254	9	Soon-to-fail	Group 3
84	0.3014	0.2880	58	Soon-to-fail	Group 3
90	0.2574	0.1679	28	Soon-to-fail	Group 3
91	0.2184	0.1400	38	Soon-to-fail	**Group 1**
92	0.1977	0.1109	20	Soon-to-fail	Group 3
93	0.3072	0.2961	85	Soon-to-fail	**Group 1**
100	0.1791	0.0769	20	Soon-to-fail	Group 3

Table 7. "Good – no action" groupings for 4-stage HS division (Number of units: 26)

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS
4	0.6494	0.5301	82	Good - no action	***Group 3***
5	0.6767	0.5897	91	Good - no action	**Group 2**
7	0.5153	0.4332	91	Good - no action	***Group 3***
16	0.6635	0.5172	84	Good - no action	**Group 2**
19	0.5808	0.4718	87	Good - no action	***Group 3***
21	0.5798	**0.3220**	**57**	Good - no action	***Group 3***
23	0.6426	0.5680	113	Good - no action	Group 4
28	0.6954	0.4567	97	Good - no action	Group 4
29	0.5316	0.4099	90	Good - no action	Group 4
30	0.6292	0.5427	115	Good - no action	***Group 3***
38	0.6067	**0.3321**	**50**	Good - no action	Group 4
45	0.5936	0.5201	114	Good - no action	***Group 3***
51	0.5792	0.5376	114	Good - no action	Group 4
54	0.5838	0.5416	97	Good - no action	**Group 2**
57	0.5224	0.4715	103	Good - no action	Group 4
59	0.7438	0.6773	114	Good - no action	**Group 2**
60	0.5648	0.4889	100	Good - no action	Group 4
70	0.5967	0.4589	94	Good - no action	Group 4
73	0.6591	0.6655	131	Good - no action	Group 4
74	0.6635	0.5865	126	Good - no action	***Group 3***
75	0.7094	0.6959	113	Good - no action	**Group 2**
79	0.5214	0.4616	63	Good - no action	***Group 3***
80	0.5907	0.4872	90	Good - no action	***Group 3***
89	0.6639	0.5279	136	Good - no action	Group 4
95	0.6753	0.7323	128	Good - no action	Group 4
97	0.5991	0.4557	82	Good - no action	***Group 1***

Table 8. "Good – monitor" groupings for 4-stage HS division (Number of units: 16)

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS
3	0.4564	0.4217	69	Good - monitor	***Group 4***
8	0.4969	0.4351	95	Good - monitor	Group 3
10	0.3764	0.3948	96	Good - monitor	Group 3
12	0.4358	0.4346	124	Good - monitor	Group 3
13	0.3651	0.3872	95	Good - monitor	**Group 1**
27	0.4133	0.3777	66	Good - monitor	Group 3
32	0.3381	0.2834	48	Good - monitor	Group 3
40	0.3335	**0.1850**	**28**	Good - monitor	**Group 1**
43	0.3733	0.2922	59	Good - monitor	**Group 1**

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k*-means HS
46	0.3231	0.2766	47	Good - monitor	Group 3
63	0.4536	0.3735	72	Good - monitor	Group 3
72	0.3421	0.3196	50	Good - monitor	Group 3
84	0.3014	0.2880	58	Good - monitor	**Group 1**
93	0.3072	0.2961	85	Good - monitor	Group 3
94	0.4627	0.3411	55	Good - monitor	Group 3
98	0.4534	0.3874	59	Good - monitor	Group 3

Table 9. "Soon-to-fail" groupings for 4-stage HS division (Number of units: 29)

Unit	Predicted PFIF	Normalized True PFIF	True RUL	Model HS	*k* -means HS
17	0.2602	0.2622	50	Soon-to-fail	Group 1
18	0.2058	0.1850	28	Soon-to-fail	Group 1
20	0.0324	0.0614	16	Soon-to-fail	Group 1
24	0.0678	0.0839	20	Soon-to-fail	Group 1
31	-0.0859	0.0077	8	Soon-to-fail	Group 1
34	-0.1119	0.0000	7	Soon-to-fail	Group 1
35	0.0754	0.0254	11	Soon-to-fail	Group 1
36	0.2490	0.1286	19	Soon-to-fail	**Group 3**
37	0.1405	0.1507	21	Soon-to-fail	Group 1
41	0.1418	0.1241	18	Soon-to-fail	Group 1
42	0.0032	0.0354	10	Soon-to-fail	Group 1
49	-0.0802	0.0414	21	Soon-to-fail	Group 1
52	0.0819	0.1312	29	Soon-to-fail	Group 1
53	0.1987	0.1362	26	Soon-to-fail	Group 1
56	0.2212	0.0869	15	Soon-to-fail	**Group 3**
58	0.1746	0.1847	37	Soon-to-fail	Group 1
61	0.0945	0.1097	21	Soon-to-fail	Group 1
62	0.1633	0.2046	54	Soon-to-fail	Group 1
64	0.1404	0.1441	28	Soon-to-fail	Group 1
66	0.2531	0.0706	14	Soon-to-fail	Group 1
68	-0.0256	0.0101	8	Soon-to-fail	Group 1
76	-0.0740	0.0173	10	Soon-to-fail	Group 1
77	0.1484	0.1844	34	Soon-to-fail	Group 1
81	-0.0803	0.0038	8	Soon-to-fail	Group 1
82	-0.0128	0.0254	9	Soon-to-fail	Group 1
90	0.2574	0.1679	28	Soon-to-fail	Group 1
91	0.2184	0.1400	38	Soon-to-fail	**Group 3**
92	0.1977	0.1109	20	Soon-to-fail	Group 1
100	0.1791	0.0769	20	Soon-to-fail	Group 1

Table 2. Summary of group assignments and accuracies (Note that percentages are based on number of units with true PFIF values within suitable thresholds)

Three-stage HS division					**Four-stage HS division**				
Category	**Number of units and percentage accuracy**				***Category***	**Number of units and percentage accuracy**			
	Model	***Accuracy***	***k-means***	***Accuracy***		***Model***	***Accuracy***	***k-means***	***Accuracy***
Healthy	29	83%	40	75%	Healthy	29	83%	22	95%
Good	31	84%	28	97%	Good – no action	26	100%	23	96%
					Good - monitor	16	94%	24	67%*
Soon-to-fail	40	97%	32	100%	Soon-to-fail	29	100%	31	84%

**Low because k-means algorithm could not clearly distinguish this group; some were assigned to the group above it and others to the group below.*

Adopting machine learning and condition monitoring P-F curves in determining and prioritizing high-value assets for life extension

Ochella, Sunday

2021-03-13

Ochella S, Shafiee M, Sansom C. (2021) Adopting machine learning and condition monitoring P-F curves in determining and prioritizing high-value assets for life extension. Expert Systems with Applications, Volume 176, August 2021, Article number 114897

https://doi.org/10.1016/j.eswa.2021.114897