Neural Computing and Applications
https://doi.org/10.1007/s00521-021-06531-4

S.I: APPLICATIONS OF MACHINE LEARNING IN MAINTENANCE ENGINEERING AND MANAGEMENT (IFAC AMEST 2020)

A machine learning-based clustering approach to diagnose multi-component degradation of aircraft fuel systems

Haochen Liu[1] · Yifan Zhao[1] · Anna Zaporowska[1] · Zakwan Skaf[1]

Received: 25 March 2021 / Accepted: 8 September 2021
© The Author(s) 2021

Abstract
Accurate fault diagnosis and prognosis can significantly reduce maintenance costs, increase the safety and availability of engineering systems that have become increasingly complex. It has been observed that very limited researches have been reported on fault diagnosis where multi-component degradation are presented. This is essentially a challenging Complex Systems problem where features multiple components interacting simultaneously and nonlinearly with each other and its environment on multiple levels. Even the degradation of a single component can lead to a misidentification of the fault severity level. This paper introduces a new test rig to simulate the multi-component degradation of the aircraft fuel system. A machine learning-based data analytical approach based on the classification of clustering features from both time and frequency domains is proposed. The scope of this framework is the identification of the location and severity of not only the system fault but also the multi-component degradation. The results illustrate that (a) the fault can be detected with accuracy $> 99\%$; (b) the severity of fault can be identified with an accuracy of almost 100%; (c) the degradation level can be successfully identified with the R-square value > 0.9.

Keywords Fast Fourier transform · Clustering analysis · Fault diagnosis · Multi-component degradation

1 Introduction

One of the vital problems of the modern aviation industry is planning maintenance in a way that will ensure machine reliability and passengers safety. Major airlines spent over $40 billion each year on these operations including among others servicing, troubleshooting, equipment and spare parts warehousing, labour and technicians training [1]. The MRO (Maintenance, Repair and Overhaul) in the aviation market cost $64.3 billion in 2015 and is expected to reach almost 100 billion with a 4.1% growth rate after a decade [2, 3]. This raises the urgent need for efficient, robust fault detection and identification tools which will be able to detect anomalies at early stages to prevent fatal failure. Raising awareness of this matter led to the development of new approaches for maintenance operations such as Integrated Vehicle Health Management (IVHM) or Condition Based Maintenance (CBM) [4–6]. Instead of scheduled parts replacement based on operating time or performed cycles, systems are monitored to perform diagnosis tasks to check their current condition and take action when the need arises or assess the remaining useful lifetime of components [7, 8].

It has been found that data-driven solutions are useful, especially in the cases where it is over-sophisticated to develop numerical models of real-life multi-component machinery [9]. For a case study of the monitoring condition of military aircraft [10], the developed Prognosis and Health Management framework was divided into three levels. Initially, the member level stands for the condition of acquiring and processing the significant information monitored using multiple types of sensors [11], which can be represented as oil tanks or gear modules. Then based on the knowledge and feature extracted from these data, the

✉ Yifan Zhao
yifan.zhao@cranfield.ac.uk

[1] School of Aerospace, Transport and Manufacturing, Cranfield University, College Rd, Wharley End, Cranfield, Bedford MK 430AL, UK

Published online: 07 October 2021

health status of subsystems was diagnosed at the following regional level, which can be represented by avionics or electromechanical system. Finally, on the last platform level, which is similar to a whole advanced military aircraft, the comprehensive health status is assessed and reported to the operator, and decisions about maintenance are made and a historical database was created and stored [12]. By this strategy, the information and uncertainty within multiscale are included and the knowledge of how the fault of a single component can propagate and affect the whole system can be developed. It set an effective analysis architecture and guidance for diagnosis in complex systems [13].

Starting from the above foundation, different approaches assessing the remaining useful lifetime of engine systems was presented in [14], where data from sensors was collected and processed to analyse the trends and symptoms of different, predefined types of faults. The condition monitoring of safety–critical complex systems in [15] applied the combination of smart sensing and fast diagnostic software, supported by a model-based reasoning system, working together for mitigating anomalies as they occur. To efficiently monitor the condition of system components and predict the remaining useful lifetime, it is wise to combine high-quality knowledge from different sources which will foster good decision making [16]. These high-quality sources come from the pre-conducted maintenance, numerical simulations, multi-type sensing and maintenance history. For example, Klingelschmidt et al. [17] presented an approach using different types of sensing data captured on various parts of machinery for condition monitoring. This type of sensing combination and accompanying complex diagnosis and prognosis system can be solved using Input–Output Hidden Markov Model architecture [18, 19]. Therefore, the abovementioned strategy can be addressed to provide better input information for multi-component fault and degradation diagnosis.

The aircraft fuel system is one of the most complex subsystems, consisting of multiple cross-linked mechanical and electro-hydraulic elements [20, 21]. Current fault diagnostic methods employ the information from adjacent or nearby sensors as reinforcement. However, they only work well when assuming that the rest untargeted components are healthy, but fails in the realistic scenario with multiple degraded components. Lin et al. [22] proposed a probabilistic framework to incorporate multi-component degradation information for the aim of fault diagnosing in aircraft fuel systems, but the identification of degradation has not been addressed. According to the above-stated diagnosis architecture, how to reveal the mist from the whole faulty status and then break through to find the specific degraded component is quite challenging and related studies are limited. The identification strategy of the multi-component degradation and degradation severity for this type of complex system is critical and demanded [23].

Addressing the above challenges, this paper aims to develop an efficient data-driven framework to detect and identify faults and degradation within multi-components in the aircraft fuel system. The application is experimentally simulated on a test rig that allows replicating failure modes of different real-life components. Then, the decision-making of fault identification is implemented by a clustering method in cooperation with time and frequency feature analysis. It can not only detect the severity level of fault but also accurately estimate the severity of degradation for multiple components.

2 Materials and methods

2.1 Experiment setup

The aircraft fuel system has a great impact on flight safety since most of the accidents associated with the fuel system lead to hazardous and even catastrophic events [24]. The deployment of fault diagnosis into the fuel system can improve not only the aircraft safety and reliability but also the turn-around time to increase the aircraft's availability. This section provides a comprehensive description of the experimental fuel rig which includes the hydraulic system, the control and measuring system, and the fault injection mechanism.

2.1.1 Hydraulic system

Figure 1 illustrates the layout of the fuel rig, which consists of three fuel tanks, three gear pumps, five shut-off valves and six direct proportional valves (DPVs). All the components are connected using the pipe and mounted on an aluminium optical breadboard (1.8 m × 1.1 m × 5 cm) which is above a drip tray to catch any unintended leak in the system. Two main tanks act as the left-wing tank and right-wing tank, respectively, and a sump tank represents the engine that receives the fuel from the aircraft fuel system. Each gear pump is driven by an external motor drive and has a pressure-relief valve inside to prevent overstressing the gears. A list of the hydraulic system specifications is shown in Table 1. More specifically, the fuel test rig can be divided into three lines:

- As illustrated in Fig. 1, the engine fuel feed line consists of the Shut-off valve 1, a non-return sticking valve (emulated by the DPV1), two gear pumps (Gear pump 1 and Gear pump 2), a pressure-relief valve (a shut-off valve), a clogged filter (emulated by the

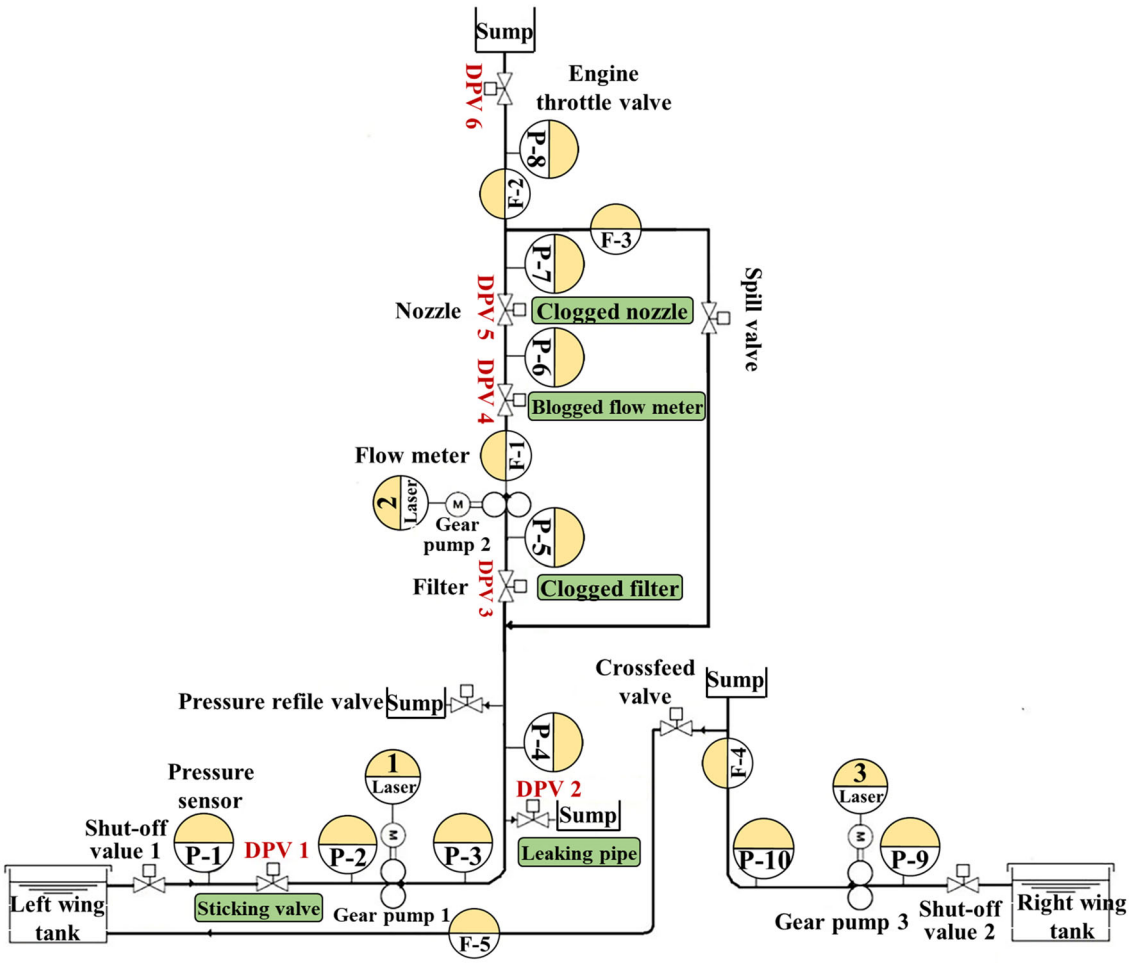

Fig. 1 Layout of the fuel rig system

Table 1 List of specifications of the hydraulic system

Item	Description
Gear Pump	Oberdorfer-N999R, Bare shaft, 150 PSI max, RPM range: 0–600 rpm, 12 teeth
Motor	3-Phase, 0.37 kW, 4 pole, 230/400 V, 50 H
Shut-off valve	Orifice: 4.5 mm, power supply: 24 V DC, 5.0 MPa max
DPV	Orifice: 3 mm, power supply: 24 V DC, 25 bar max
Pipe	Polyurethane, 4 mm internal diameter and 6 mm outside diameter

DPV3), a flow metre (emulated by the DPV4) and a clogged nozzle (emulated by the DPV5).

- The cross-feed line includes the shut-off valve 2, Gear pump 3 which transfers the fuel from the right-wing tank to the left-wing tank to keep maintaining the central gravity of the aircraft during flight, and a cross-feed valve (a shut-off valve).
- The spill line includes a spill valve (a shut-off valve) which is used to return the fuel when the engine requires less fuel from the aircraft fuel system, an engine throttle valve (emulated by the DPV6) which is used to generate the backpressure when the spill valve is opened.

2.1.2 Control and measuring system

The control and measuring system (CMS) includes ten pressure sensors (marked with P-x in), five flow metres (marked with F-x), three laser sensors, three AC inverters (ABB ACS150), nine National Instruments (NI) module cards that installed in two NI CDAQ-9172 8 slots USB chassis and a LabVIEW-based software. A snapshot of CMS is illustrated in Fig. 2. National Instruments (NI)

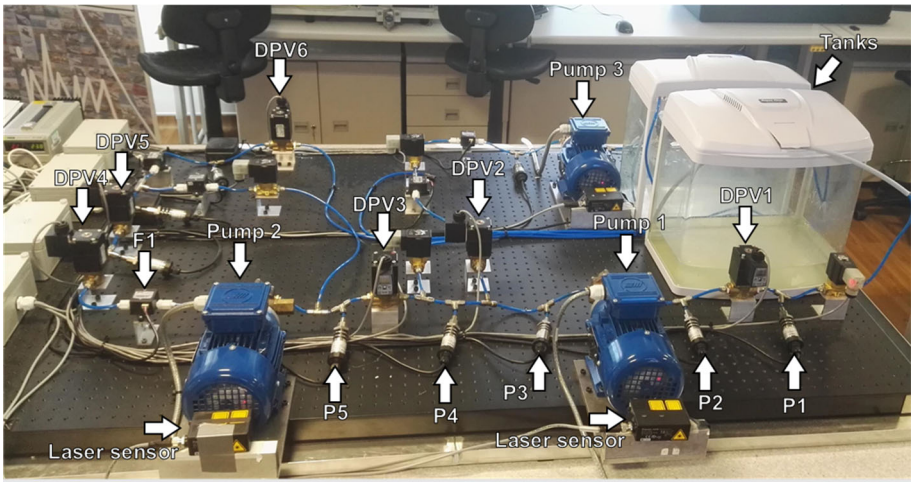

Fig. 2 A snapshot of the laboratory scale testing rig

LabView software 2014 was used to customise the control for the entire system. The NI modules used in the control and measuring system are chosen for their customisable and accurate features compared with other tools. More specifically, the nine NI module cards selected for the control and measuring system are NI 9485, NI 9205, NI 9264, three NI 9401 and three NI 9472. The specifications of the sensors are summarised in Table 2. The NI 9485 module is an 8-channel solid-state relay sourcing or sinking digital output module. It allows direct connection to a variety of industrial devices such as valves and motors. The NI 9485 module is chosen to control the open/close status of the five-solenoid shut-off valves in the hydraulic by providing access to the solid-state relay for switching the voltage applied to the shut-off valve. The NI 9205 module is a 250 kS/s, 32-channel voltage input module. It is chosen to receive the analogue voltage output from the ten pressure sensors and five flow metres, and convert this information using the calibration forms into digitised information readable on the developed software. The sampling rate is 1 kHz within the LabVIEW environment. The NI 9264 is a 25 kS/s, 16-channel module simultaneously updating the analogue output module which is chosen to enable the implementation of the six DPVs position control. The DPV position is modified by varying the voltage applied to the solenoid and is an open circuit. The NI 9401 module is a configurable digital I/O interface. It is chosen to receive the output from the laser sensor and convert them into a frequency for calculation of the pump speed. The NI 9472 module is an 8-channel 24 V logic, sourcing digital output module which is chosen to provide the signals to the pump inverter to implement the pump controls (start the pump, stop the pump, increase speed, maintain speed, decrease speed). The pump speed input from the control system is 0–5 V to the inverter drive, whereby the inverter drive determines the 3-phase motor control.

Dedicated software was developed using LabVIEW for controlling the system and collecting the data. The user has control over shut-off valves position, pump speed/volumetric flow rate, and direct proportional valves position (manually or automatically). For DPVs control, the user has two options: to set the opening percentage of the DPVs manually through the knobs or to set a couple of parameters (time, opening percentage) to control the DPVs through a defined profile. It should be noted that the control system is ready to accommodate all planned failure modes in a plug and play manner, in terms of both hardware and

Table 2 Specifications of the sensors

Sensor	Description
Absolute pressure sensor	Measurement range: 0–5 bar, output signal: 0–5 V, quoted accuracy: ± 0.25% (full scale), power supply: 12.8 V DC
Gauge pressure sensor	Measurement range: 0–4 bar, output signal: 0–5 V, quoted accuracy: ± 0.25% (full scale), power supply: 12.8 V DC
Flow meter	Measurement range: 0–2 L/min, output signal: 0–5 V, quoted accuracy: ± 3% (full scale), temperature sensitivity: ± 0.2% per °C, reference temperature: 23 °C, power supply: 12.8 V DC
Laser sensor	Sensing range: 0–10 m, power supply: 10–30 V DC

software. The data file, provided by the control and measuring system, contains the time of the experiment, the atmospheric pressure and temperature in the lab, the readings from the ten pressure sensors, five flow metres and three laser sensors, the status of the shut-off valves, and the opening percentage of the DPVs. For each sensor, the sample frequency was chosen as 1000 Hz considering the sensor specification and dynamics of this test rig.

2.1.3 Fault types

For an aircraft fuel system, the common hardware faults can be classified into three different types, namely process faults, actuator faults and sensor faults, according to the type of faulty hardware. Process faults include faults that affect the operational ability of the system itself such as a leaking pipe or cracked joint. Actuator faults include faults that affect the actuated parts of the system such as pump malfunction or a sticking valve, and sensor faults include faults that affect the sensor operation. The degradation of components causes most faults in the aircraft fuel system due to fouling, erosion, or corrosion. Table 3 defines the simulated faults corresponding to the designed test rig.

To inject various faults, with different degrees of severity into the fuel test rig, five DPVs were used. The failure of the shut-off valve (being stuck in amid range position) is implemented by using DPV1 that is initially fully open. fully opened DPV1 is equivalent to a healthy valve status while the partially closed DPV1 is equivalent to a sticking valve with a certain degree of fault severity. Different degrees of severity can be emulated by varying the opening percentage of the DPV which is controlled through the developed software and can be varied from 0 to 100%. DPV2 is used to emulate a leaking pipe fault into the system and set to be initially fully closed. The fully closed DPV2 is equivalent to a healthy pipe while the partially opened DPV2 is equivalent to a leaking pipe with a certain degree of severity. DPV3 is used to inject a clogged filter fault into the system. The fully opened DPV3 is equivalent to a healthy filter while the partially closed DPV3 is equivalent to a clogged filter with a certain degree of severity. DPV4 is used to inject a blocked flow metre fault into the system. The fully opened DPV4 is equivalent to a healthy flow metre while the partially closed DPV 4 is equivalent to a blocked flow metre with a certain degree of severity. DPV5 is used to inject a clogged nozzle fault into the system. The fully opened DPV5 is equivalent to a healthy nozzle while the partially closed DPV5 is equivalent to a clogged nozzle with a certain degree of severity.

Table 3 The faults injected into the test rig

Fault type	Fault
Process fault	Leaking pipe
Actuator fault	Sticking valve
	Clogged filter
	Clogged nozzle
Sensor fault	Blocked flow meter

2.2 Severity of fault and degradation

In this study, DPV2 was used to simulate the leaking pipe, as a system fault. The severity of this fault is simulated by varying the opening percentage of this valve with values of 0% (healthy), 30%, 40%, and 50%. It should be noted that in this study, the opening percentage of DPV2 less than 30% is not considered as a fault. This study aims to detect and identify the fault when the severity level is just above the threshold between fault and degradation (30%), which allows for preventing further development and propagation of the damage at the earliest possible stage. Additionally, faults with a severity level exceeding 50% are significantly easier to be detected and identified than the lower ones, therefore ignored.

For each fault severity of DPV2, the leaking percentage of four valves (Sticking valve, Clogged filter, Blocked flow meter, Clogged nozzle shown in Fig. 1) was changed from 0%, 10% to 20%, respectively, to simulate a multi-component degradation. The leaking percentage was used to simulate the degradation level. In other words, for each of these four valves, there are 3 possible states or degradation levels. There are $3^4 = 81$ possible combinations of multi-component degradation for each fault severity level. The corresponding between the combination index and percentage value of each of the four valves is shown in Table 4. The combination index presents an exclusive location and severity of this multi-component degradation. Therefore, the identification of this complex degradation problem is now simplified to the estimation of the combination index with associated faulty severity.

2.3 Data analysis

Due to the complexity of this system, applying a model-based approach to monitor the system condition is difficult due to the challenge to establish an analytical or numerical model to effectively represent this system. Even each component can be modelled successfully, at the system level, different components may interact simultaneously and nonlinearly, which leads that the modelling of the whole system is almost impossible. Additionally, any change of system design will require the model to be re-estimated, so the universality of such an approach is

Table 4 The combination of leaking percentage of each valve to simulate multiple-component degradation

Faulty severity (leaking pipe) (%)	Degradation level				
	Combination index	Sticking valve (%)	Clogged nozzle (%)	Blocked flow meter (%)	Clogged filter (%)
0	1	0	0	0	0
	2	0	0	0	10
	3	0	0	0	20
	4	0	0	10	0
	5	0	0	10	10
	6	0	0	10	20
	7	0	0	20	0
	…	…	…	…	…
	81	20	20	20	20
30	1	0	0	0	0
	…	…	…	…	…
	81	20	20	20	20
40	1	0	0	0	0
	…	…	…	…	…
	81	20	20	20	20
50	1	0	0	0	0
	…	…	…	…	…
	81	20	20	20	20

limited. This paper proposes to use a model-free clustering analysis based on two features extracted from the time and frequency domains, respectively, incorporated with machine learning classifiers for fault detection, classification and identification, as well as degradation identification.

The proposed methodology of data analysis can be illustrated in Fig. 3. There are four objectives (illustrated by four colour blocks) including (a) fault detection: determining if the system is faulty or healthy, (b) fault classification: dividing the sampled data into a number of groups using an unsupervised approach, which mainly evaluates how well the selected features differentiate severity levels of fault, (c) fault identification: determining the severity level of fault, and (d) degradation identification: determining the severity level of each degradation and combination, which is simplified to the determination of the combination index. Initially, all pressure sensors are considered, and the used sensors are then narrowed down gradually through three steps of channel selection or reduction (illustrated by red colour) to identify an optimal channel eventually. The details of each process are described below.

2.3.1 Feature extraction

Assuming that the system is stationary under a specific severity level of fault and degradation combination, the mean amplitude of each channel (P1, P2, …, P8 in Fig. 1) is calculated by

$$\overline{P_l} = \frac{1}{N}\sum_{n=0}^{N-1} P_l(n) \tag{1}$$

where l is the index of the channel and N is the number of sampling data. The channels P9 and P10 are not considered because they are in a different branch with the components that have fault and degradation. It has been observed that the amplitude of pressure measurements changes for different severity levels of fault, this value is therefore considered as the first feature from the time domain. The second feature is represented by the peak frequency with the maximum power of frequency response or Power Spectrum Density (PSD) of each channel. The frequency response for a considered channel P_l is given by

$$X_l(k) = \sum_{n=0}^{N-1} P_l(n) e^{\frac{-i2\pi kn}{N}} \tag{2}$$

where $k = 0, \ldots, N-1$. The peak frequency f_l is written as

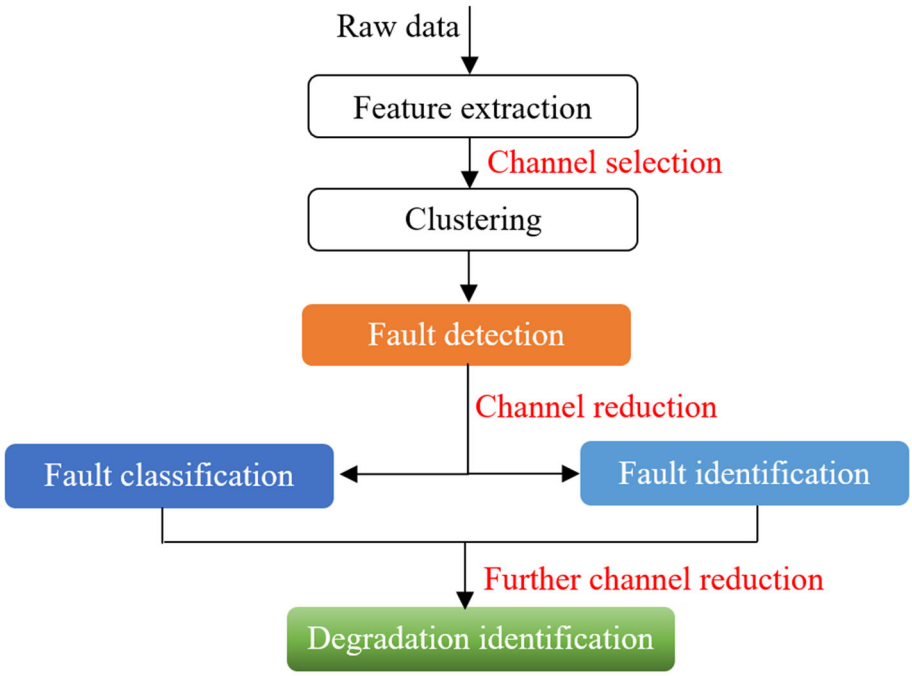

Fig. 3 The pipeline of the proposed methodology of data analysis

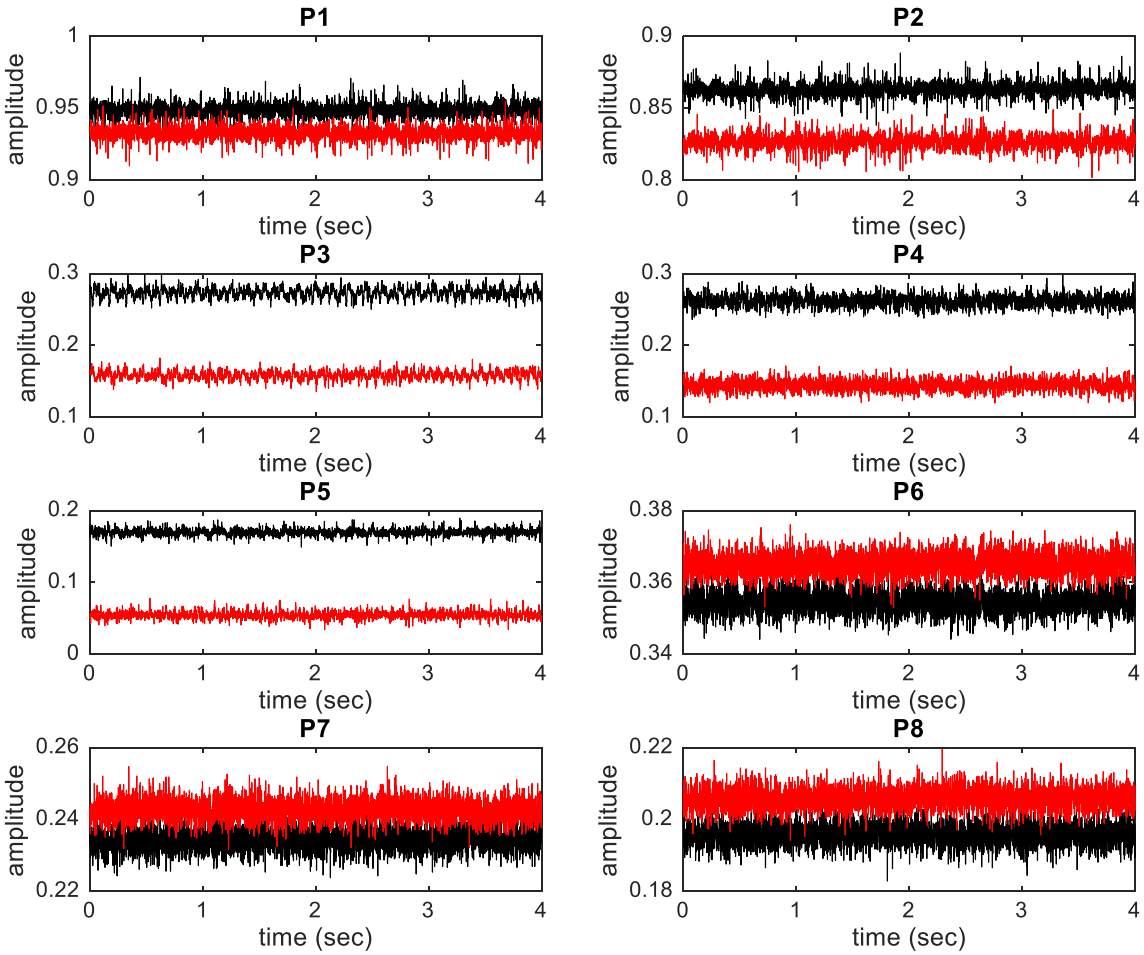

Fig. 4 A raw measurement comparison for 8 considered channels where the black indicates the case with no fault and no degradation and the red indicates the case with 30% fault severity and no degradation

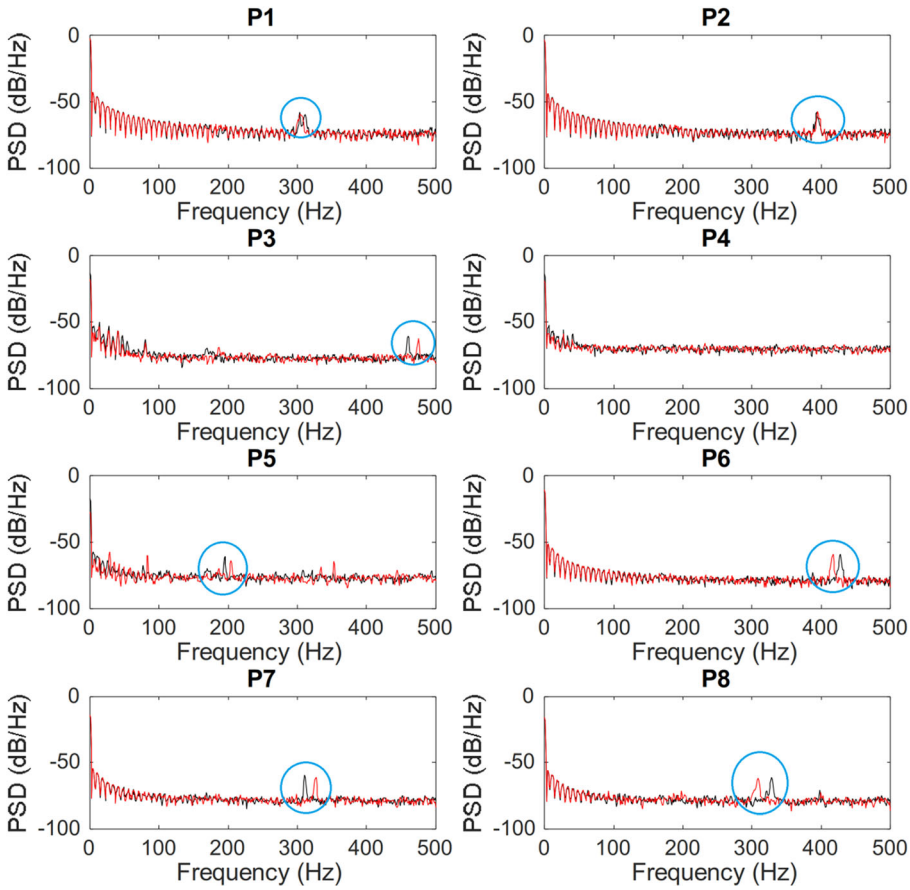

Fig. 5 A comparison of PSD for 8 considered channels where the black indicates the case with no fault and no degradation and the red indicates the case with 30% fault severity and no degradation

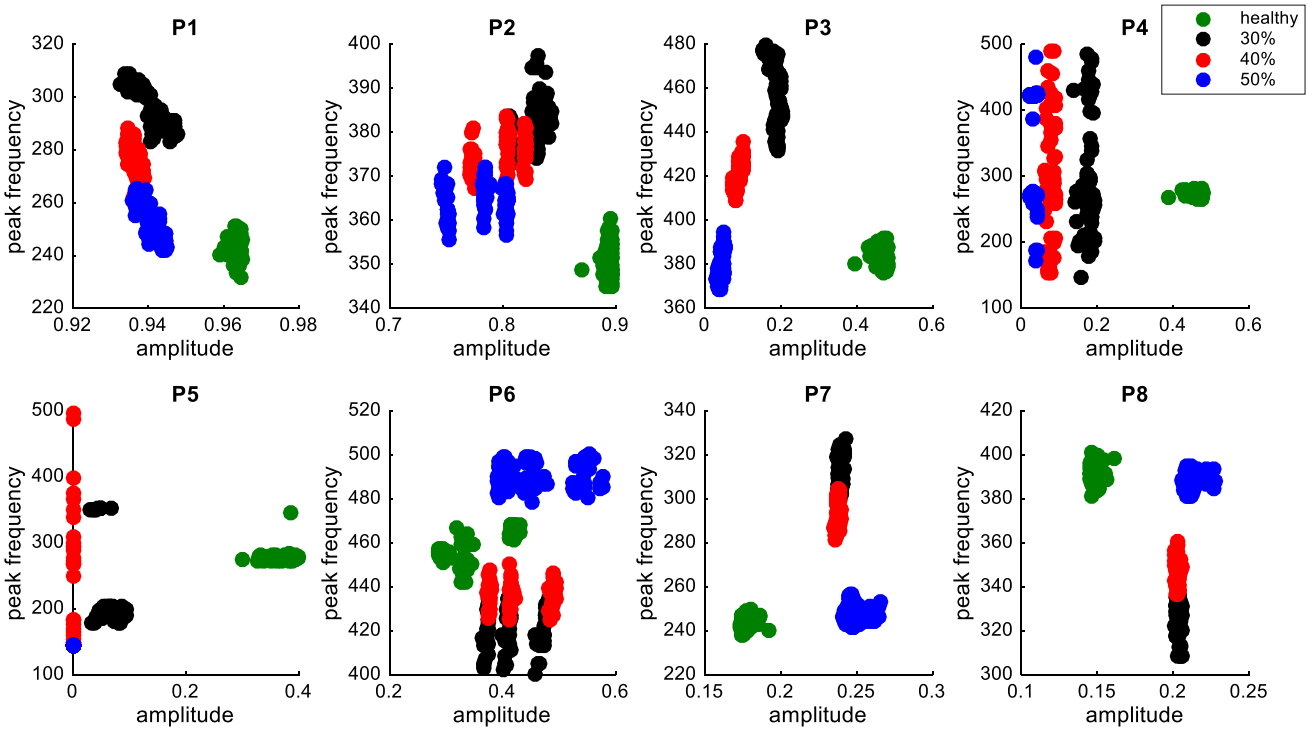

Fig. 6 Clustering visualisation for all 8 considered channels, where each case with a fault has 81 combinations of degradation

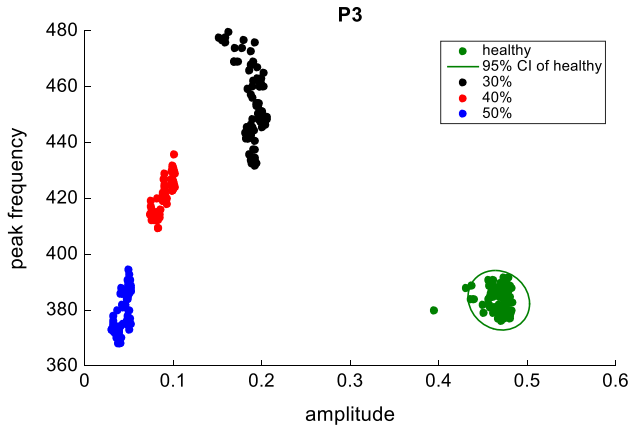

Fig. 7 An example of fault detection based on the channels P3, where the ellipse represents the 95% confidence level of healthy behaviour

Table 5 Performance quantification of fault detection

Channel	False positive number	ACC (%)
P1	4	97.86
P2	2	99.38
P3	2	99.38
P4	2	99.38
P5	2	99.38
P6	3	99.07
P7	3	99.07
P8	4	97.86

$$f_l = \frac{f_s}{2N} \arg \max_{\tau \le k \le N-1} |X_l(k)| \tag{3}$$

where τ the left bound for the searching and f_s denotes the sample rate.

2.3.2 Fault detection and classification

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). The most commonly used method is k-means clustering. It is an unsupervised categorisation of samples (features) into separate classes in a way that the similarity inside the class is maximised while the similarity outside the class is minimised. Given a set of observations $(x_1, x_2, \ldots, x_n)$, where each observation is a d-dimensional real vector, the k-means clustering aims to partition the n observations into $k(\le n)$ sets $S = \{S_1, S_2, \ldots, S_k\}$ to minimise the within-cluster sum of squares. The objective can be written as

$$\arg\min_S \sum_{i=1}^{k} \sum_{x \in S_i} \|x - u_i\| = \arg\min_S \sum_{i=1}^{k} |S_i| \mathrm{Var}(S_i) \tag{4}$$

where u_i is the mean of points in S_i. In this study, the observation is a two-dimensional vector $\{\overline{P}_l, f_l\}$ and sampled number (n) is $4 \times 81 = 324$, which include four-fault severity levels (0%, 30%, 40%, and 50%), and each level includes 81 combinations of degradation.

For the task of fault detection, assuming a Gaussian distribution of features, a 95% confidence ellipse in the two-dimensional feature space is calculated based on the healthy data. An ellipse can be written as

$$\frac{(x - c_x)^2}{a^2} + \frac{(y - c_y)^2}{b^2} = 1 \tag{5}$$

where (c_x, c_y) is the centre, and a and b are two radiuses of horizontal and vertical directions. To extend the equation to a more general case where the ellipses could be rotated, the Principle Component Analysis is applied to calculate the 2 by 2 coefficients PC. If a feature vector (v_1, v_2) is located inside the ellipse or determined as healthy, the following condition must be satisfied in Eq. (6). Otherwise, this vector is located outside the ellipse or determined as faulty.

$$\frac{((v_1 - c_x) \cdot \mathrm{PC}')^2}{a^2} + \frac{((v_2 - c_y) \cdot \mathrm{PC}')^2}{b^2} < 1 \tag{6}$$

For the task of fault classification, the number of group k is chosen as 4 and the unsupervised k-means method is applied. Given the knowledge of the ground truth of classes, the performance of classification can be evaluated by the Adjusted Rand Index (ARI) [19], Mutual Information (MI) [25], Silhouette Coefficient (SC) [26]. Adjusted Rand Index is a function that measures the similarity of the two assignments, ignoring permutations and with chance normalisation. Given a set S of n elements, and two groupings or clustering of these elements, namely $X = \{X_1, X_2, \ldots, X_r\}$ and $Y = \{Y_1, Y, \ldots, Y_s\}$, the overlap between X and Y can be summarised in a contingency table $[n_{ij}]$ where each entry n_{ij} denotes the number of objects in common between X_i and Y_j: $n_{ij} = |X_i \cap Y_j|$, the Adjusted Rand Index is described by

$$\mathrm{ARI} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}} \tag{7}$$

where $a_i = \sum_{q=1}^{r} n_{iq}$, $b_i = \sum_{q=1}^{r} n_{qj}$ and $\binom{n}{2}$ is calculated as $n(n-1)/2$.

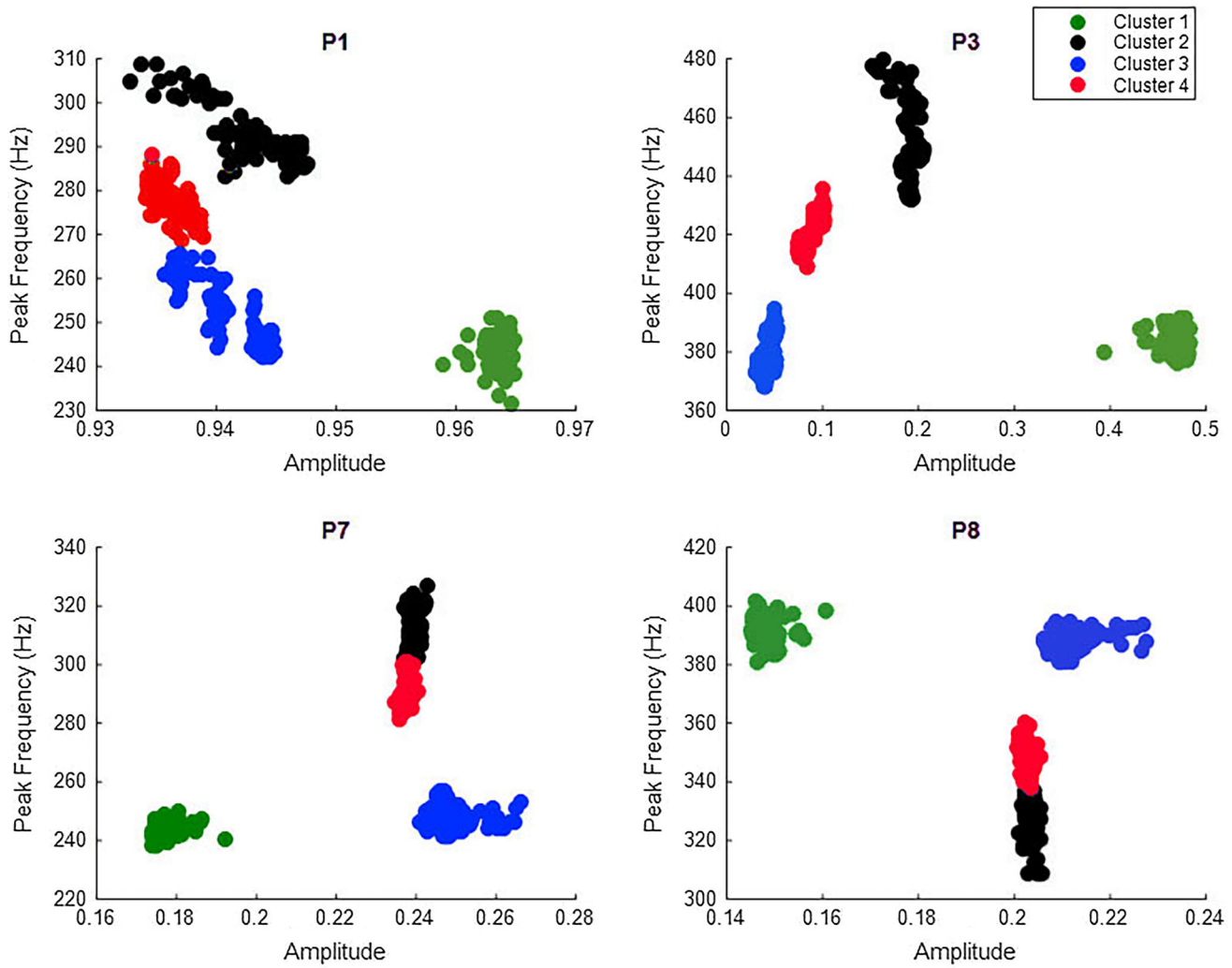

Fig. 8 Classification results of fault severity level using the k-means approach (k = 4)

Table 6 Performance quantification of fault classification

Channel	Amplitude + PSD			Amplitude			PSD		
	ARI	MI	SC	ARI	MI	SC	ARI	MI	SC
P1	0.960	1.892	0.765	0.482	1.086	0.823	0.680	1.471	0.770
P3	1.000	2.000	0.955	1.000	2.000	0.959	0.603	1.407	0.868
P7	0.960	1.915	0.870	0.573	1.318	0.892	0.686	1.485	0.701
P8	0.984	1.958	0.893	0.511	1.271	0.804	0.594	1.411	0.822

Mutual information is another function to measure the agreement of the two clusterings, and it can be written as

$$I(X,Y) = \sum_i \sum_j p(X_i, Y_j) \log\left(\frac{p(X_i, Y_j)}{p(X_i)p(Y_j)}\right) \tag{8}$$

where $p(X_i, Y_j)$ is the joint probability function of X and Y, and $p(X_i)$ and $p(Y_j)$ are the marginal probability distribution functions of X and Y respectively. The Silhouette Coefficient can be used when the ground truth labels are not known and it represents the degree of isolation. A higher Silhouette Coefficient score relates to a model with better-defined clusters. Let $a(i)$ be the average distance between i and all other data within the same cluster, $b(i)$ be the lowest average distance of i to all points in any other cluster, the Silhouette Coefficient is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{9}$$

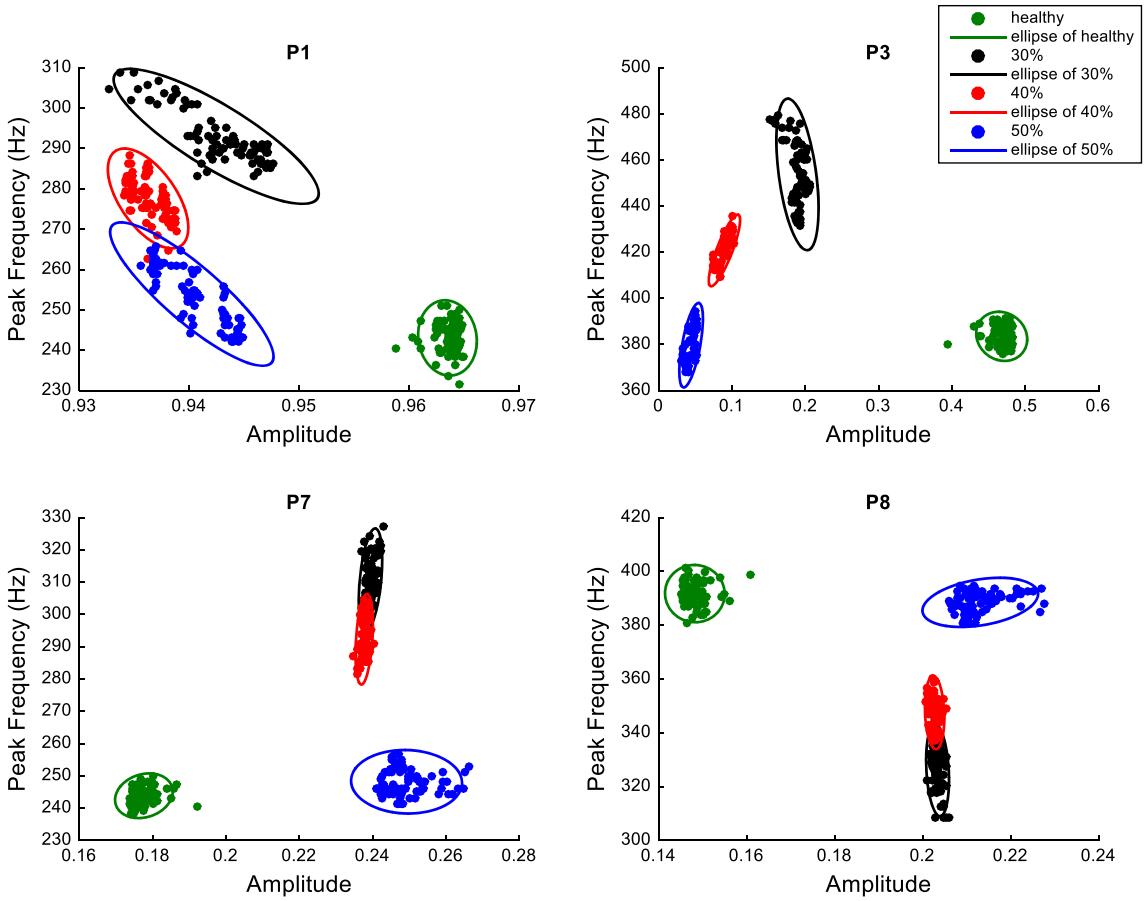

Fig. 9 The result of fault identification using the 95% confidence ellipse

From the above definition, it is clear that $s(i) \in [-1, 1]$. An $s(i)$ close to 1 means that the data is appropriately clustered. For the task of fault identification, one approach is to establish the 95% confidence ellipse for each level of fault (healthy, 30%, 40% and 50%). Another approach is to use the supervised learning approaches, such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Complex Tree and Boosted Tree. The confusion matrix is used to measure the identification results and help select the optimal channels for monitoring regarding the experiments conducted in this paper. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa). Typical derivations from a confusion matrix include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). To describe the percentage of the correctly classified samples, the accuracy of classification is given by

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{10}$$

2.3.3 Degradation identification

Considering the large number of the possible state of combination index representing the degradation location and level, this paper proposes to use a parametric modelling approach to establish the relationship between the extracted features and the combination index. Given a set of observations $(x_1, x_2, \ldots, x_{81})$ under different combination indexes, denoted by $y(y \in [1, 81])$, for each severity level of fault, a model can be identified and written as

$$\hat{y} = f(x) \tag{11}$$

with the least square errors. If there is only one feature selected, the linear representation of Eq. (11) can be written as

$$\hat{y} = \alpha x + \beta \tag{12}$$

This model can then be used to predict the combination index based on the observed feature. The model

Table 7 Identified parameters of 95% confidence ellipses for the fault identification

Level	c_x	c_y	a	b	PC
P1					
Healthy	0.9635	243.06	0.0027	9.33	$\begin{pmatrix} -2.34e-5 & 1 \\ 1 & 2.34e-5 \end{pmatrix}$
30%	0.9424	292.99	0.0051	16.74	$\begin{pmatrix} -4.68e-4 & 1 \\ 1 & 4.68e-4 \end{pmatrix}$
40%	0.9363	277.62	0.0030	12.37	$\begin{pmatrix} -1.74e-4 & 1 \\ 1 & 1.74e-4 \end{pmatrix}$
50%	0.9402	253.91	0.0037	17.78	$\begin{pmatrix} -3.61e-4 & 1 \\ 1 & 3.61e-4 \end{pmatrix}$
P2					
Healthy	0.4677	383.53	0.0345	10.66	$\begin{pmatrix} -3.81e-4 & 1 \\ 1 & 3.81e-4 \end{pmatrix}$
30%	0.1893	453.75	0.0257	32.86	$\begin{pmatrix} -3.96e-4 & 1 \\ 1 & 3.96e-4 \end{pmatrix}$
40%	0.0896	420.79	0.0114	15.54	$\begin{pmatrix} 1.20e-3 & 1 \\ 1 & -1.20e-3 \end{pmatrix}$
50%	0.0442	379.68	0.0125	18.31	$\begin{pmatrix} 5.53e-4 & 1 \\ 1 & -5.53e-4 \end{pmatrix}$
P7					
Healthy	0.1776	243.83	0.0077	6.92	$\begin{pmatrix} 2.11e-4 & 1 \\ 1 & -2.11e-4 \end{pmatrix}$
30%	0.2394	310.72	0.0029	15.90	$\begin{pmatrix} 8.04e-5 & 1 \\ 1 & -8.04e-5 \end{pmatrix}$
40%	0.2377	292.28	0.0022	14.01	$\begin{pmatrix} 5.51e-5 & 1 \\ 1 & -5.51e-5 \end{pmatrix}$
50%	0.2492	248.16	0.0151	9.83	$\begin{pmatrix} -5.28e-5 & 1 \\ 1 & 5.28e-5 \end{pmatrix}$
P8					
Healthy	0.1482	391.72	0.0068	10.65	$\begin{pmatrix} -8.92e-6 & 1 \\ 1 & 8.92e-6 \end{pmatrix}$
30%	0.2034	326.04	0.0025	17.08	$\begin{pmatrix} -3.23e-5 & 1 \\ 1 & 3.23e-5 \end{pmatrix}$
40%	0.2027	345.51	0.0022	13.61	$\begin{pmatrix} -2.28e-5 & 1 \\ 1 & 2.28e-5 \end{pmatrix}$
50%	0.2131	388.47	0.0124	9.14	$\begin{pmatrix} 4.79e-4 & 1 \\ 1 & 4.79e-4 \end{pmatrix}$

performance can then be evaluated by calculating the Root Mean Square Error and R-Squared between the ground truth y and the predicted $\hat{y}$.

3 Results

To demonstrate the effectiveness of the proposed framework, the case study where the leaking pipe is faulty has been tested.

3.1 Fault detection

For each channel, a total of 4 s of data were sampled with a sample rate of 1000 Hz. An example of raw healthy data and data with a fault is shown in Fig. 4, where no degradation is introduced (combination index = 0). It can be observed that the amplitudes for most of the channels are different between the healthy case and the cause with a fault, particularly for P2, P3, P4 and P5. For P1, P6, P7 and P8, although the amplitude difference can be observed if the means are taken into account, they are not suitable for real-time fault detection due to the overlap of amplitude. It

Table 8 The confusion matrix of the fault severity level identification using the 95% confidence ellipse

	Predicted					Total
	Healthy	30%	40%	50%	Unknown	
P1 actual						
Healthy	77	0	0	0	4	81
30%	0	78	0	0	3	81
40%	0	0	79	2	0	81
50%	0	0	0	80	1	81
Total	77	78	79	82	8	
P3 actual						
Healthy	79	0	0	0	2	81
30%	0	76	0	0	5	81
40%	0	0	79	0	2	81
50%	0	0	0	81	0	81
Total	79	76	79	81	9	
P7 actual						
Healthy	78	0	0	0	3	81
30%	0	78	18	0	3	99
40%	0	25	79	0	2	106
50%	0	0	0	78	3	81
Total	78	103	97	78	11	
P8 actual						
Healthy	77	0	0	0	4	81
30%	0	76	5	0	5	86
40%	0	16	78	0	3	97
50%	0	0	0	78	3	81
Total	77	92	83	78	15	

Table 9 The accuracy of the fault severity level identification using 95% CI ellipse

Channel	ACC (%)
P1	96.9
P3	97.2
P7	85.3
P8	89.6

can be seen that the discrepancy between the amplitude for P1, P2, P6, P7 and P8 to distinguish healthy and fault scenario is no more than 6% (1.7%, 4.2%, 3.0%, 3.6%, 5.1% respectively), but the discrepancy between the amplitude for P3, P4 and P5 is more than 40% (42.1%, 44.8%, 66.7% respectively). So, moving away from the fault location will decrease the discrepancy of the pressure amplitude.

Using the same data, Fig. 5 shows the results of PSD for the 8 channels. It can be observed that there are significant peaks for both healthy and faulty data for all channels except P4. The peak frequency is different for different channels, as highlited by the blue circles. It can be observed from all channels, except P4, that there is a shift of the location of peak frequency for the same channel between the healthy case and the case with a fault (e.g., P3), which is used as the second feature for clustering.

The clustering results using the mean amplitude and peak frequency are shown in Fig. 6, where different colours indicate different severity levels of fault, and each level of fault includes 81 cases. It is suggested from visual inspection that P1, P3, P7 and P8 have good performance to separate these four groups while the features in the other 4 channels are partially overlapping. Considering P3 for example, the mean amplitude can separate them into 4 groups but the groups of 30%, 40% and 50% are not very well separated. The peak frequency can separate 30%, 40% and 50%, but healthy and 50% are overlapped. The degree of isolation is significantly improved if both features are considered.

Through the visual inspection, P3 was selected for fault detection. The 95% confidence ellipse for the healthy group was calculated and illustrated in Fig. 7. The estimated parameters in Eq. (6) are described by

$$c_x = 0.4677, \quad c_y = 383.53, \quad a = 0.0345, \quad b = 10.66$$
$$\mathrm{PC} = \begin{pmatrix} -3.81e-4 & 1 \\ 1 & 3.81e-4 \end{pmatrix} \tag{13}$$

If a tested vector locates inside this ellipse, the system is determined as healthy, otherwise the system is faulty. A similar approach can be applied to the other 7 channels and the fault detection results are shown in Table 5. All 8 channels produce a sound performance with the false positive number less than 5 of 324 (1.54%), although P2, P4, P5 and P5 are not appropriate to distinguish the four groups.

Table 10 The accuracy of the fault severity level identification using selected supervised learning-based approaches

Channel	KNN (%)	Linear SVM (%)	Complex tree (%)	Boosted tree (%)
P1	99.1	99.1	96.0	87.3
P3	100	100	100	100
P7	97.5	98.1	97.8	69.4
P8	98.8	99.1	98.5	74.7

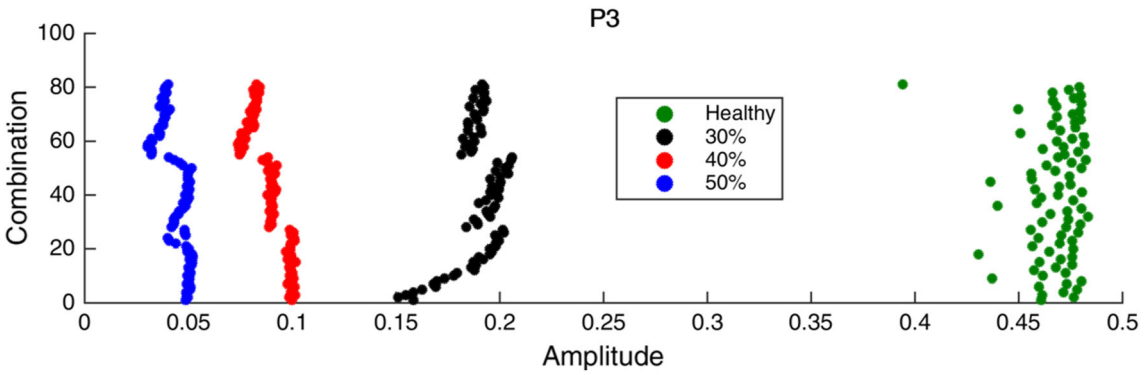

Fig. 10 The relationship between the mean amplitude and combination index for different groups using P3

3.2 Fault classification

Base on the above results, the channels P1, P3, P7 and P8 were selected for the fault classification and identification. The k-mean approach was applied to the data shown in Fig. 6, where k was chosen as 4. The classification results are shown in Fig. 8. Since it is an unsupervised approach, the correspondence between the colour and severity level is still unknown. Comparison of Figs. 6 and 8 suggests that P3 produces a 100% classification result while the other three channels produce slightly lower accuracy. To quantify the performance, Table 6 shows the results of ARI, MI and SC between the true tags and the produced tags from the k-means method for these four channels. It can be observed that P3 has the best performance for all three criteria. P8 has the second-best performance, and P1 and P7 have slightly worse performance. Table 6 also compares the performance between a single feature and two features. For P3, the amplitude itself produces the perfect results. While for the other three channels, the performance is significantly improved when two features are used than a single feature. This observation suggests that multiple features can increase the robustness of the proposed method in terms of channel selection.

3.3 Fault identification

To identify the severity level of fault, the 95% confidence ellipse was estimated for each group in P1, P3, P7 and P8, and the results are shown in Fig. 9. It can be observed that P3 produces the best result where there is no overlap between ellipses. For P1, there is a small region of overlap between 40 and 50%; for P7 and P8, there is a small region of overlap between 30 and 40%. The estimated parameters of 95% confidence ellipses for the fault identification are shown in Table 7, which can be used to determine which severity of fault from the given features. The detailed performance of this approach can be described by the confusion matrix shown in Table 8. The total accuracy of fault identification for each considered channel is shown in Table 9, which again proves that P3 produces the best performance of fault identification.

Other supervised learning-based approaches, namely KNN, SVM, Complex Tree and Boosted Tree, have also been applied to identify the fault level and results are shown in Table 10. The channels P3 still has the best performance. It should be noted that although the performance of some of these methods is better than the 95% confidence ellipse, the models lack transparency and cannot be written down.

3.4 Degradation identification

Figure 10 shows the relationship between the mean amplitude and the combination index for the 4 groups using the best channel P3. For the healthy group, there is no significant association observed. For the severity level 30%, 40% and 50%, although the correlation looks promising, it is difficult to use a continuous function to describe their relationship. Therefore, although the mean amplitude is a good feature for fault detection and identification, it is not an appropriate feature for degradation identification. Figure 11 plots the relationships between the peak frequency and combination index, and the results are more promising. A strong linear relationship for each group and each considered channel has been observed. The clusters can be well fitted using a linear model, as shown in Eq. (9). The linear fitting for each group is also plotted in Fig. 11. It can be observed that the channels P3, P7 and P8 have the best performance. The identified parameters of the linear fitting models for the group of healthy, 30%, 40% and 50%, respectively, for each considered channel are shown in Table 11, where the corresponding model performance based on RMSE and R-squared is also shown. The model of channel P7 produces the best degradation identification for the healthy and 40% group with an R-squared value of 0.657 and 0.888, respectively. The model of channel P3 produces the best degradation identification

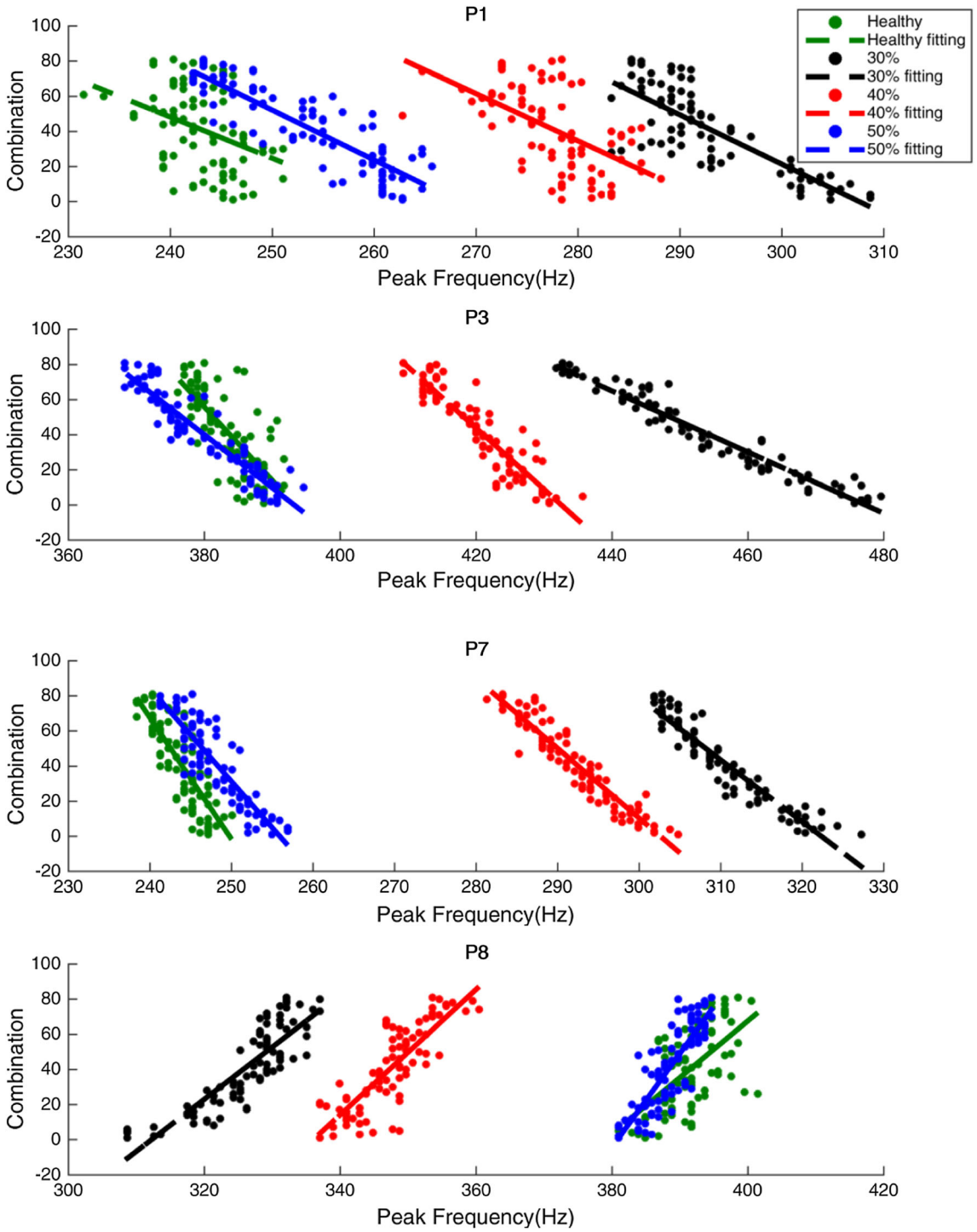

Fig. 11 The relationship between the combination and peak frequency for different levels of fault

for the 30% and 50% group with an *R*-squared value of 0.938 and 0.878, respectively.

4 Conclusions

To address the challenge of faulty and degradation diagnosis of a complex engineering system where the multi-component degradation is presented, this paper has presented a test rig and a corresponding data analytical

Table 11 Identified linear fitting models to describe the relationship between the combination index and the peak frequency with corresponding model performance

Level	α	β	RMSE	R^2
P1				
Healthy	− 2.4	612.3	8.16	0.138
30%	− 2.8	864.9	11.04	0.635
40%	− 2.7	787.5	11.00	0.317
50%	− 2.8	763.3	10.50	0.733
P3				
Healthy	− 4.2	1643.4	11.70	0.568
30%	− 1.7	831.0	5.69	0.938
40%	− 3.5	1506.5	8.69	0.839
50%	− 3.0	1189.1	7.74	0.878
P7				
Healthy	− 6.9	1727.2	11.20	0.657
30%	− 3.5	1136.6	7.09	0.901
40%	− 4.0	1201.7	7.48	0.888
50%	− 5.3	1351.0	9.93	0.772
P8				
Healthy	3.2	− 1220.1	11.20	0.337
30%	3.0	− 928.3	10.4	0.739
40%	3.6	− 1199.1	11.10	0.676
50%	5.3	− 2037.1	11.00	0.685

approach for four tasks: fault detection, fault classification, fault identification and degradation identification. The results suggest the following conclusions:

- The discrepancy of the pressure amplitude between healthy and faulty scenario depends on the fault location. P3–5 have more than 40% amplitude discrepancy while other sensors have less than 6% amplitude discrepancy when the fault is emulated between P3 and P4. Observing the change of amplitude of certain channels is sufficient to detect, classify and identify the fault.
- For the cases where the severity level of fault is the same while the degradation level increases, there is no regular pattern of amplitude change, which suggests the amplitude cannot be used for degradation identification. The shift of the frequency peak, showing a linear trend of decrease following the increment degradation level, is an effective feature to identify the degradation level.
- By clustering two features including amplitude and peak frequency, the fault can be detected with an accuracy $> 97\%$; the severity of fault can be identified with an accuracy $> 99\%$; the degradation can be successfully identified with the R-square value > 0.9.

A limitation of this approach is that it only works on stationary systems where the fault or degradation severity level does not change when shifted in time. To widen its applications, future study will focus on its extension on dynamical systems. The time when the system behaviour changes will then be resolved.

Acknowledgements We would like to appreciate the Integrated Vehicle Health Management (IVHM) Centre at Cranfield University for their contribution of data collection.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Heisey R (2002) 717–200: low maintenance costs and high dispatch reliability. AERO Mag 19:18–29
2. Oster CV, Strong J, Zorn CK (2006) Why airplanes crash: aviation safety in a changing world. Oxford University Press, Oxford
3. Wittmer A, Bieger T, Müller R (2013) Aviation systems: management of the integrated aviation value chain. Springer, Berlin
4. Aaseng GB (2004) Blueprint for an integrated vehicle health management system. Adv Astronaut Sci 118:373–386
5. National Aeronautics and Space Adm Nasa (Author) (NASA) (1992) Research and technology goals and objectives for Integrated Vehicle Health Management (IVHM)
6. Benedettini O, Baines TS, Lightfoot HW, Greenough RM (2009) State-of-the-art in integrated vehicle health management. Proc Inst Mech Eng Part G J Aerosp Eng 223(2):157–170
7. Aivaliotis P, Georgoulias K, Chryssolouris G (2019) The use of digital twin for predictive maintenance in manufacturing. Int J Comput Integr Manuf 32(11):1067–1080
8. Papacharalampopoulos A, Tzimanis K, Sabatakakis K, Stavropoulos P (2020) Deep quality assessment of a solar reflector based on synthetic data: detecting surficial defects from manufacturing and use phase. Sensors 20(19):5481
9. Chien CF, Chen SL, Lin YS (2002) Using Bayesian network for fault location on distribution feeder. IEEE Trans Power Deliv 17(3):785–793
10. Zhang A, Cui L, Zhang P (2013) Advanced military aircraft of study on condition-based maintenance. In: Proceedings—2013 international conference on information technology and applications, pp 462–465

11. Nor NM, Hussain MA, Hassan CRC (2020) Multi-scale kernel Fisher discriminant analysis with adaptive neuro-fuzzy inference system (ANFIS) in fault detection and diagnosis framework for chemical process systems. Neural Comput Appl 32(13):9283–9297
12. Gao XZ, Wang X, Zenger K (2014) Motor fault diagnosis using negative selection algorithm. Neural Comput Appl 25(1):55–65
13. Shah D (2014) Millions of data points flying in tight formation. http://www.aerospacemanufacturinganddesign.com/article/millions-ofdata-points-part2-121914/
14. Lv Z, Wang J, Zhang G, Jiayang H (2015) Prognostics health management of condition-based maintenance for aircraft engine systems. In: 2015 IEEE conference on prognostics and health management (PHM), pp 1–6
15. Jakovljevic M, Artner M (2006) Protocol-level system health monitoring and redundancy management for integrated vehicle health management. In: AIAA/IEEE digital avionics systems conference proceedings, pp 1–7
16. Madhikermi M, Buda A, Dave B, Framling K (2017) Key data quality pitfalls for condition-based maintenance. In: 2nd international conference on system reliability and safety (ICSRS), pp 474–480
17. Klingelschmidt T, Weber P, Simon C, Theilliol D, Peysson F (2017) Fault diagnosis and prognosis by using Input–Output Hidden Markov Models applied to a diesel generator. In: 25th Mediterranean conference on control and automation (MED), pp 1326–1331
18. Prakash G, Narasimhan S, Pandey MD (2019) A probabilistic approach to remaining useful life prediction of rolling element bearings. Struct Health Monit 18(2):466–485
19. Gates A, Ahn Y (2017) The impact of random models on clustering similarity. J Mach Learn Res 18(87):1–28
20. Lee H, Lim HJ, Chattopadhyay A (2021) Data-driven system health monitoring technique using autoencoder for the safety management of commercial aircraft. Neural Comput Appl 33(8):3235–3250
21. Sadiq A, Ahmad F, Khan SA (2014) Modeling and analysis of departure routine in air traffic control based on Petri nets. Neural Comput Appl 25:1099–1109
22. Lin Y, Zakwan S, Jennions I (2020) A Bayesian approach to fault identification in the presence of multi-component degradation. Int J Progn Health Manag 8(1). https://doi.org/10.36001/ijphm.2017.v8i1.2530
23. Yu J, Zhang C, Wang S (2021) Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes. Neural Comput Appl 33:3085–3104
24. Subramanian N, He H, Jennions I (2020) Fault detection for aircraft fuel system with neural network. SSRN Electron J. https://doi.org/10.2139/ssrn.3718044
25. Baudot P, Tapia M, Bennequin D, Goaillard JM (2019) Topological information data analysis. Entropy 21(9):869
26. de Amorim RC, Hennig C (2015) Recovering the number of clusters in data sets with noise features using feature rescaling factors. Inf Sci 324:126–145

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://dspace.org/

A machine learning-based clustering approach to diagnose multi-component degradation of aircraft fuel systems

Liu, Haochen; Zhao, Yifan; Zaporowska, Anna; Skaf, Zakwan

2021-10-07
Springer
https://doi.org/10.1007/s00521-021-06531-4
Liu H, Zhao Y, Zaporowska A, Skaf Z. (2021) A machine learning-based clustering approach to diagnose multi-component degradation of aircraft fuel systems. Neural Computing and Applications, Available online 7 October 2021

Downloaded from a DSpace repository. https://dspace.org/