IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 21, 2020, accepted January 5, 2021, date of publication January 11, 2021, date of current version January 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3050792

State-of-Health Assessment for Aero-Engine Based on Density-Distance Clustering and Fuzzy Bayesian Risk

NING MA[1], FAN YANG[2,3], LAIFA TAO[2,3], AND MINGLIANG SUO[2,3], (Member, IEEE)
[1]Aviation Maintenance NCO Academy, Air Force Engineering University, Xinyang 464000, China
[2]School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[3]Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China

Corresponding author: Mingliang Suo (suozi@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61903015 and Grant 61973011, in part by the Fundamental Research Funds for the Central Universities under Grant YWF-20-BJ-J-723, in part by the National key Laboratory of Science and Technology on Reliability and Environmental Engineering under Grant 6142004180501, and in part by the China Postdoctoral Science Foundation under Grant 2019M650438.

ABSTRACT State-of-health (SOH) assessment for aero-engine can effectively reduce maintenance cost and operational risk, and is also a significant part of the prognostics and health management (PHM) system. However, the current SOH assessment is usually closely coordinated with other parts of PHM to achieve specific functions. This is not conducive to generalizing the function of SOH assessment. Therefore, this paper proposes a data-driven framework of SOH assessment that mainly includes data preprocessing, pseudo label generation, weight assignment and feature selection, and assessment, which enhances the systematicness of SOH assessment. A combination model based on density-distance clustering and fuzzy Bayesian risk models is designed to generate a pseudo label, select optimal parameter subset, and assign weight. Then, two assessment indicators including state membership degree and health degree are produced based on two fuzzy models for horizontal and vertical comparisons. These two indicators expand the dimensions and perspectives of SOH measurement, which can more comprehensively characterize the health state of the engine. Finally, the correctness and effectiveness of the proposed methodology are verified by the widely used Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset.

INDEX TERMS State-of-health assessment, aero-engine, state membership degree, health degree, pseudo label generation.

I. INTRODUCTION

Aero-engine is one of the typical representatives of complex industrial equipment, who is also called ''the heart of airplane'' [1], [2]. For complex equipment, condition-based maintenance is the main task of its health management. An important prerequisite for condition-based maintenance is accurate health assessment that directly affects the mission planning of the airplane. Accurate and effective SOH assessment of the engine will be conducive to performing condition-based maintenance on the engine and making an appropriate decision on the flight task [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

Prognostics and health management (PHM) is a significant guarantee for reducing the cost of aero-engine maintenance and improving the flight quality [4], [5], which has aroused wide interest in relevant scholars and engineers [6], [7]. Data-driven PHM technology occupies a dominant position that is more suitable for the complex aero-engine than model one [8]. SOH assessment is one of the key steps in PHM [9]. Generally, after receiving the monitoring data, the SOH assessment performs on the concerned object, and then the assessed result is sent to diagnose some faults, predict the remaining life or decision-making [10], [11].

In view of the above analysis, SOH assessment and aero-engine are the two focuses of our work, which are the main research item of PHM and the main object of complex equipment, respectively. Therefore, we will introduce the

existing research results of SOH assessment and the SOH assessment of aero-engine successively in the following part.

With respect to the SOH assessment, a generalized and systematically theoretical model is a practical requirement for engineering applications. Such a theoretical model is easier to transplant and expand, and can eventually be developed into an industry standard and promote the PHM system to become generalized and universal. For a department with multiple assets of the same type, it usually hopes to receive recommendations for equipment optimal selection from the results of SOH assessment. This is also the most common issue we encounter in engineering applications, such as aircraft scheduling and combat decision-making.

For the research results of SOH assessment, scholars usually divide it and state classification into the same issue. Literature [12] classified the SOH of the piston pump using an LSTM neural network. Soft failure and hard failure are proposed in literature [13] to improve the effectiveness of health assessment for electronic components. Some health indicators are extracted for diagnosing different fault types of the tools in smart manufacturing [10]. In addition to the above forms, some scholars combined SOH assessment with other key parts in PHM as an entire model to achieve some concerned functions. Literature [14] performed decision-making after acquiring the result of the health assessment for a wind turbine blade. Literature [11] proposed a quantum assimilation based method for SOH assessment and remaining useful life (RUL) prediction for electronic systems, in which the assessment result was fed into the part of RUL. Literature [15] proposed a framework of health assessment and RUL based on multi-dimensional performance and multi-failure mode for a wind turbine. Some other related achievements can refer to [16]–[18]. Therefore, it can be seen from the above literature that the research on SOH assessment has not yet been independently and systematically developed, which is the main deficiency in the study of SOH assessment. The main motivation of this paper is to make up for this shortcoming, a detailed and systematic study of SOH assessment is the main focus of our work.

Regarding the SOH assessment of aero-engine, some scholars consider that SOH assessment is an auxiliary link to help other functions of PHM [19]–[21]. Among them, the combination of health assessment and RUL for an engine is the most common research issue, and researchers prefer to use the publicly available Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset as a typical representative to carry out relevant study [22]–[24]. In the C-MAPSS dataset, each engine starts with different degrees of initial wear and manufacturing variation and begins to degrade at some time, all these details are unknown to the public. In other words, these public datasets are not accompanied by corresponding state labels, which makes it difficult to assess the SOH of these engines. In this regard, scholars have proposed some methods including equal frequency method [25], clustering method [26], piece-wise method [20], average state level [27], manually segmented [28], and other methods [21]. For more research results on the health state dividing of the C-MAPSS dataset, interested readers can refer to literature [22], [29]. It can be seen that the states produced by the aforementioned methods are too rigid, which is a problem that we try to solve by introducing a clustering model in the study of SOH assessment.

For the SOH assessment of the C-MAPSS dataset, scholars have carried out some successful research work. A data-driven health indicator constructor and health state division framework for the degradation model was proposed based on relative entropy Weibull-SAX [29]. State estimation and prediction based on Belief functions and hidden Markov models were discussed and applied to the C-MAPSS dataset [28]. In the literature [30], state classification for structural health diagnosis was carried out based on Deep Belief Network and demonstrated with the C-MAPSS dataset. Bayesian methods were employed to assess the operational safety of multi-mode engineering systems including the C-MAPSS gas turbine [31]. Multiple deterioration level assessment was discussed in the literature [21].

So far, the research work related to SOH assessment can be summarized as shown in Table 1. From these results in Table 1, the following conclusions can be drawn: 1) None of these studies provide a framework model with generalization ability for health state assessment, 2) The health state division and expression of these research results usually take their own lifespan as the yardstick, e.g., statistical distribution [29] and equal frequency health state [30], and do not have the ability to consider the comparative analysis of similar equipment operation state, which is to provide users with some recommendations of equipment optimal selection. Filling these two gaps is the main motivation of our work.

To make up for the deficiencies mentioned above, we employ the C-MAPSS dataset as the object to systematically research the SOH assessment of aero-engine. After an in-depth analysis of the characteristics of the C-MAPSS dataset, we propose a theoretical framework based on labeled-multiple attribute decision-making (LMADM) for data-driven SOH assessment, which mainly includes data preprocessing, pseudo label generation, weight assignment and feature selection, and assessment. For label generation, we propose a pseudo label generation strategy assisted by a density-distance based clustering (DDC) method proposed in [32]. The DDC model can yield some ideal results that is why it is employed in our work. The SOH assessment of aero-engine is directly related to flight safety that requires the SOH assessment to take fully into account the decision-making risk. Thus, we introduce a fuzzy Bayesian risk (FBR) model that is specifically designed for decision system [33] to extract an optimal parameter subset, and assign some appropriate weights to these selected parameters. The pseudo label generated by DDC makes the C-MAPSS dataset from an information system to a decision system. This operation helps FBR to be applicable to the information system. To obtain a more comprehensive health assessment result, we put forward two aspects of health assessment including

TABLE 1. The summary of related work.

Ref. No.	Object	Method of SOH assessment	Method of state generation	Health score	Target
Ref. [10]	Robot cutting tool	N/A	Feature extraction	Vertical score	Diagnosis
Ref. [11]	Electronic systems	Quantum assimilation	Scaled	Vertical score	Prediction
Ref. [12]	Piston pump	Long Short-Term Memory neural network	Manually segmented	Vertical score	Classification
Ref. [13]	Electronic components	Conditional monitoring	Discretization technique	Vertical score	Prediction
Ref. [14]	Wind turbine blade	Grey relation	Three indexes	Vertical score	Decision-making
Ref. [15]	Wind turbine	Combination method	Multi-performance	Vertical score	Prediction
Ref. [16]	Aircraft	Grey clustering and fuzzy comprehensive evaluation	Expert knowledge	Vertical and horizontal scores	Management
Ref. [17]	Circuit	Dynamic Bayesian network	Expert knowledge	Vertical score	Prediction
Ref. [18]	Lithium-ion battery	Relevance vector machine	Feature fusion	Vertical score	Mission planning
Ref. [19]	Aero-engine	Case-based reasoning	Equal frequency	Vertical score	Prediction
Ref. [20]	Aero-engine	Long Short-Term Memory neural network	Piece-wise	Vertical score	Prediction
Ref. [21]	Aero-engine	Hyper-parameter optimization	Multi-state deterioration	Vertical score	Prediction
Ref. [22]	Aero-engine	Unsupervised reconstruction	Data-driven labeling	Vertical score	Prediction
Ref. [23]	Aero-engine	Deep belief network	Self-defined	Vertical score	Prediction/Diagnosis
Ref. [24]	Aero-engine	Weighted fusion	Linear regression	Vertical score	Prediction
Ref. [25]	Aero-engine	Hidden Markov model	Equal frequency	Vertical score	Prediction
Ref. [26]	Aero-engine	Kernel principal component analysis and Weibull failure rate function	Clustering	Vertical score	Prediction
Ref. [27]	Aero-engine	Multistate deterioration process	Average state level	Vertical score	Prediction/Diagnosis
Ref. [28]	Aero-engine	Belief functions and hidden Markov model	Manually segmented	Vertical score	SOH assessment
Ref. [29]	Aero-engine	Relative entropy Weibull-SAX	Relative entropy	Vertical score	SOH assessment
Ref. [30]	Aero-engine	Deep belief network	Manually segmented	Vertical score	SOH assessment
Ref. [31]	Aero-engine	Bayesian methods	Gaussian mixture model	Vertical score	SOH assessment

horizontal comparison and vertical comparison using two fuzzy models, which generate the membership degree of engine state and the health degree of the engine relative to its historical state, respectively. Therefore, the main contributions of this paper are concluded as follows:

1) A framework of data-driven state-of-health assessment is proposed that mainly includes data preprocessing, pseudo label generation, weight assignment and feature selection, and assessment. This highlight tries to make up for the shortcoming that there is no independently and systematically research on the SOH assessment, which can also expand the theoretical scope of the PHM system and enhance its application value.
2) A combination model based on the density-distance clustering and fuzzy Bayesian risk models is designed to select a relative optimal parameter subset and assign weights for the parameters in the selected subset, which approaches the final result with the minimum risk and is more suitable for the proposed SOH assessment framework. This combined model can expand the application ability of the fuzzy Bayesian risk model in the information system.
3) Two indicators characterizing the state-of-health of aero-engine, state membership degree and health degree, are developed by horizontal and vertical comparison fuzzy models, respectively. These two indicators can reveal the health state of equipment more comprehensively to assist users to make more appropriate decisions.
4) A detailed process of state-of-health assessment is demonstrated through the C-MAPSS dataset, in which the difference between some theoretical results and practical application is explained. The processing techniques can help the proposed theory be easily applied to practical engineering.

The remainder of this paper is organized as follows. The detailed methodology of state-of-health assessment is introduced in Section II, some comparison experiments are carried out in Section III. Then, the validation of the proposed methodology is elaborated in Section IV. Finally, Section V concludes our work and provides some future work.

II. METHODOLOGY

A. BASIC FRAMEWORK OF STATE-OF-HEALTH ASSESSMENT

Multiple attribute decision-making (MADM) theory and its extension model labeled-multiple attribute decision-making are two of the important theories in management science [33]. We transplant the idea of LMADM into the health state assessment of aero-engine and design a general framework of health state assessment that is special for complex equipment. The basic steps of this framework are as follows:

1) **Data preprocessing:** noise reduction, sensor preselection, normalization;
2) **Pseudo label generation:** dimensionality reduction, state center generation, label calibration;
3) **Weight assignment and feature selection:** decision system building, weight producing, sensor secondary selection;
4) **Assessment:** aggregation of health score, horizontal state membership degree and vertical health degree calculation.

The relationship between the above-mentioned steps is shown in Figure 1. Regarding the object data, there are two categories including historical data and current data, which are also considered as training data and test data. The training

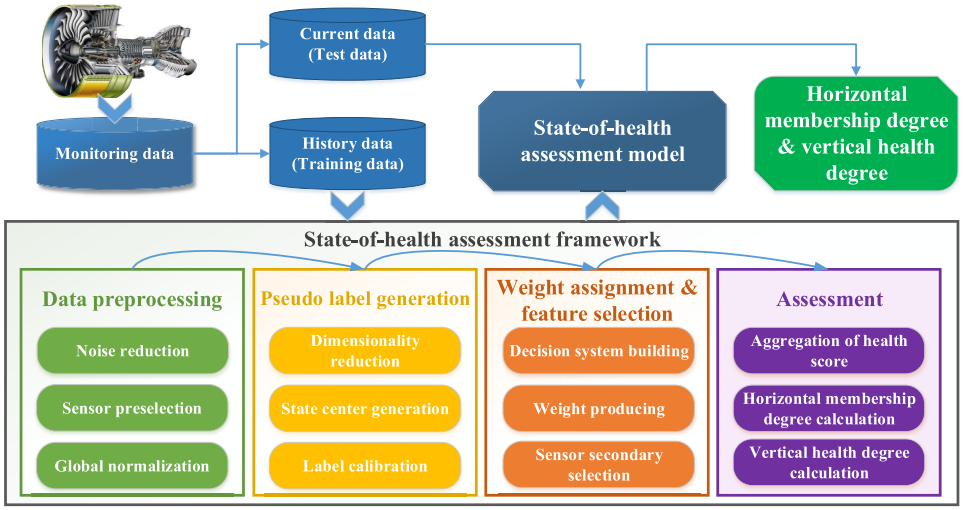


FIGURE 1. The schematic of our strategy for SOH assessment.

data are fed into the proposed framework to build an assessment model, two health degrees are subsequently obtained after the current data being input the trained model.

B. DATA ANALYSIS AND PREPROCESSING

Regarding the dataset for SOH assessment, there are some selection criteria to follow:

1) The dataset can characterize the operation state of the engine.
2) The dataset contains some potential multiple states of the engine, which are limited to the same operating condition.
3) The used dataset can characterize the full lifespan of the engine.

Affected by regular inspections and overhauls, the actual dataset of the engine usually contains fluctuating characteristics. We can obtain the initial state information and end-of-life information of the engine through the user manual and maintenance manual. This information can be equivalently converted into the above three criteria.

The first criterion is a basic principle. The second one is an important criterion to obtain more accurate results of SOH assessment, and it is also a guarantee that the horizontal health state comparison mentioned in the following sections is feasible. The third one can ensure that we can achieve a longitudinal health state comparison of a single-engine.

Following the above criteria, the dataset used in this paper is generated by a simulation platform of C-MAPSS [34]. A schematic diagram of the upper half of the simulated engine is shown in Figure 2. The corresponding relationship of the abbreviations in Figure 2 is shown in Table 2.

In this dataset, there are 21 sensory data collected by sensors that effectively reflect the running state of the engine, and three others denote the setting of the operating conditions. In our work, we focus on the most widely used dataset named FD001 in C-MAPSS. The details of this dataset are shown in Table 3, and the descriptions of these 21 sensor attributes are shown in Table 4.

We take the first engine named FD001-01 as an example to illustrate the data trends of these 21 sensors, which is shown

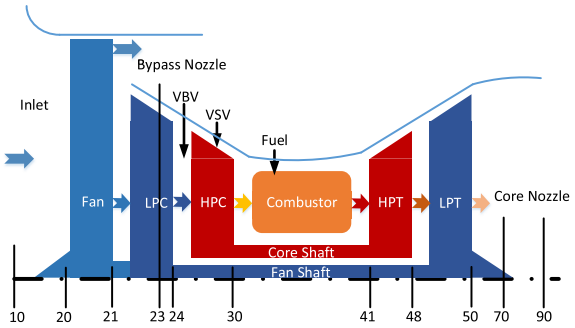


FIGURE 2. The schematic diagram of the upper half of the simulated engine.

TABLE 2. The details of the abbreviations in Figure 2.

Abbreviation	Full name
LPC	Low pressure compressor
HPC	High pressure compressor
HPT	High pressure turbine
LPT	Low pressure turbine
VBV	Variable bleed valve
VSV	Variable stator vanes

TABLE 3. The details of FD001 dataset.

Information	Value
Engines in training set	100
Samples in training set	20,631
Engines in testing set	100
Samples in testing set	13096
Operating conditions	Sea level
Fault modes	HPC degradation

TABLE 4. The descriptions of the 21 sensor attributes.

ID	Symbol	Description	Units
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	epr	Engine pressure ratio (P50/P2)	–
11	Ps30	Static pressure at HPC outlet	psia
12	phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass Ratio	–
16	farB	Burner fuel-air ratio	–
17	htBleed	Bleed Enthalpy	–
18	Nf-dmd	Demanded fan speed	rpm
19	PCNfR-dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

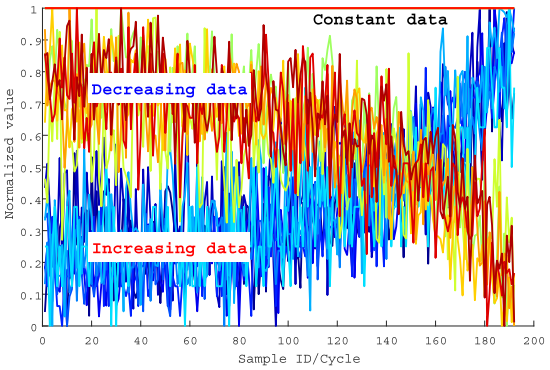


FIGURE 3. The data trends of the 21 sensors of FD001-01 engine. There are three types of data, including decreasing data, increasing data and constant data.

in Figure 3 after scaling into the range of [0, 1]. The engine runs for 192 cycles in its entire lifespan, so 192 samples representing the engine state are recorded. Obviously, there are three typical trend characteristics in this dataset:

1) The data with Sensor IDs 7, 9, 12, 14, 20, 21 (see the lines marked with warm color in Figure 3) show a downward trend over time, this kind of data is the decreasing type.
2) The data with Sensor IDs 2, 3, 4, 8, 11, 13, 15, 17 (see the lines highlighted by cold color in Figure 3) present an upward trend over time, this kind of data is the increasing type.
3) The data with Sensor IDs 1, 5, 6, 10, 16, 18, 19 (see the red straight line in Figure 3) show a time-independent characteristic, this kind of data is the constant type.

Regarding the FD001 engine, it has experienced a high pressure compressor degradation failure at some time that results in the different trend characteristics of these data [34].

As can be seen from Figure 3, there exist three problems that need to be solved in the data preprocessing stage, which are the problems of noise, unhelpful constant value, and different development trend characteristics. The following ways can solve these three problems.

1) THE WAY FOR REDUCING NOISE

It is difficult to find the nature of data when the data is contaminated by noise. We employ a windowed average method to alleviate the influence of noise in the time series data of C-MAPSS.

Given a finite time series $X = \{x_1, x_2, \cdots, x_n\}$ and an integer window length L, $1 < L < n$, then the new element is denoted by

$$\widetilde{x}_i = \frac{\sum_{i-L+1}^{i} x_i}{L}, \tag{1}$$

where $L \leq i \leq n$. Then, the new time series is $\widetilde{X} = \{\widetilde{x}_L, \widetilde{x}_{L+1}, \cdots, \widetilde{x}_n\}$.

2) THE WAY FOR DELETING THE UNHELPFUL CONSTANT VALUE

According to the selection criteria of the SOH assessment dataset, a parameter that cannot reflect the changes of the engine state will not assist in assessing health state, and such a parameter with constant value data is this type of parameter. A group of data whose maximum and minimum values are equal can be considered as a constant one.

The constant type of data presents a constant characteristic because they are not related to the only degradation mode of FD001, i.e., HPC degradation, but more related to fan performance, which can be clearly concluded by the information in Table 4. From the perspective of engine structure, the fan and HPC do not share the same shaft, thus the performance degradation of HPC will not directly interfere with the performance of the fan. These constant data are not helpful in the SOH assessment for FD001 because they cannot reveal the performance change of the engine, and should be deleted.

3) THE WAY FOR NORMALIZING THE DATA WITH TWO DEVELOPMENT TRENDS

To better represent the degradation trend of the engine and lay a foundation for the subsequent health assessment, i.e., all the samples need to be normalized to a descending interval with the range of [0,1], we thus use different normalization models for the data with an upward trend and downward trend. The cost and income normalized models are the most commonly used models that can meet the above requirements, which are shown as follows [35], [36]:

$$\overline{x}_i = \frac{\max(X) - x_i}{\max(X) - \min(X)}, \tag{2}$$

$$\overline{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}, \tag{3}$$

where $x_i \in X = \{x_1, x_2, \cdots, x_n\}$, $\overline{x}_i$ is the normalized element, $\max(\cdot)$ and $\min(\cdot)$ are respectively the maximum and minimum operators, i.e., $\max(X)$ denotes the maximum value of X, and $\min(X)$ is the minimum one for X.

The income normalized model (Eq. (3)) is utilized for the decreasing type of data, and the cost one (Eq. (2)) is used for the increasing type of data. It is worth noting that the normalization operation is performed for all the training samples rather than for the samples of a single-engine.

Consequently, based on the above three solutions, for the data of FD001-01 engine, we can obtain the reconstructed data in the range of [0, 1] as shown in Figure 4, where $L = 20$ that is set in terms of the experimental results of literature [37] and our engineering experience.

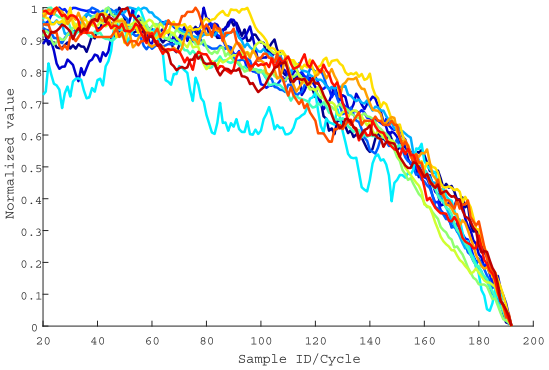


FIGURE 4. The reconstructed data of the remaining 14 sensors of FD001-01 engine.

C. PSEUDO LABEL GENERATION BASED ON DENSITY-DISTANCE CLUSTERING

The labeling of the C-MAPSS data has always been a hot issue for scholars. To solve this problem, we propose a label generation method based on clustering idea. The clustering categories can be regarded as the pseudo labels of the C-MAPSS dataset.

It is worth noting that the label produced by the clustering method is not used for final decision-making, but is only an intermediate product to assist in SOH assessment. Thus, the clustered label is named a pseudo label.

Regarding the clustering method, literature [32] has proposed an idea that a cluster center is characterized by higher local density than most other points and large relative distance to other points with higher local densities, based on which a new cluster model named density-distance clustering (DDC) is proposed. The basic concepts are as follows.

In DDC, the local density ρ_i of sample x_i is denoted as

$$\rho_i = \sum_{x_j} D(d_{ij} - d_c), \tag{4}$$

where i and j are the serial number of the sample in the current set, d_{ij} is the distance between samples x_i and x_j, d_c is a distance threshold, we set d_c as recommended in [32] that the average number of neighbors is around 2% of the total number of points in the discussed dataset, $D(\cdot)$ is a cut-off function, $D(\cdot) = 1$ when $d_{ij} - d_c > 0$ and $D(\cdot) = 0$ otherwise.

The relative distance of sample x_i is defined as

$$\delta_i = \begin{cases} \min(\cup \text{dist}(x_i, x_j)), & \text{sort}(\rho_i, \rho_j) = [\rho_j, \rho_i], \\ \max(\delta), & \rho_i = \max(\rho), \end{cases} \tag{5}$$

where $\text{dist}(\cdot, \cdot)$ denotes some distance function, as which the Euclidean distance is generally used. $\text{sort}(\cdot, \cdot)$ is a descending order function. For the first part, we sort the local density of samples, and then obtain the minimum relative distance according to the sorting results, where the relative distance of sample x_i is only related to the samples with larger local densities than x_i's. From the second part, it can be seen that if the local density of x_i is the maximum one, its relative distance is also the maximum one. Some more detailed explanations for this equation can be found in [32].

The center of each cluster is generated based on the combination of local density ρ_i and relative distance δ_i, the combination is called a decision function that is defined by

$$\theta_i = \bar{\rho}_i \bar{\delta}_i, \tag{6}$$

where $\bar{\rho}_i$ and $\bar{\delta}_i$ are the new local density and relative distance that are scaled into the range of [0, 1], this scaled operation could guarantee identical weights for both during choosing cluster centers. The cluster centers are those with relatively greater decision function values than other points after setting the number of clusters. Then, the remaining elements could be labeled according to the neighborhood principle that their neighbor has a higher density. The basic steps can be illustrated as shown in Figure 5.

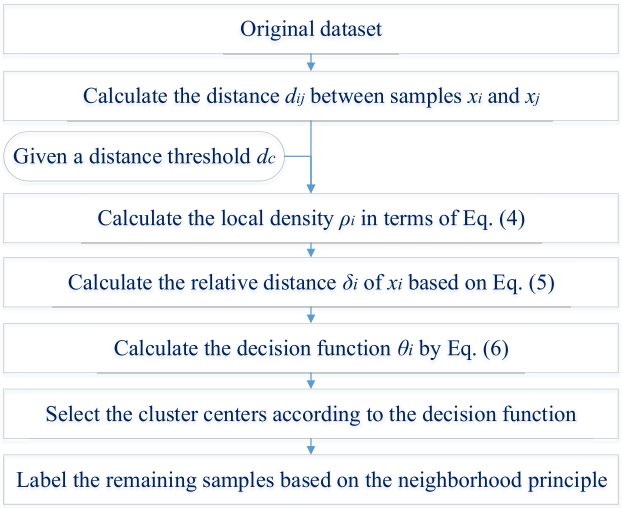


FIGURE 5. The flow chart of pseudo label generation.

Based on the above steps, the data of the FD001-01 engine can be reconstructed as shown in Figure 4. The number of clusters is three according to the three basic states, i.e., health state, degradation state, and failure state.

When utilizing the clustering method in our work, we put forward the following hypothesis: If the interference samples in the original dataset are removed and more representative samples are extracted, the clustering effect may be improved. Therefore, to improve the quality of clustering and generate more reliable labels, Principal Component Analysis (PCA) [38] is employed that is a quantitatively rigorous method for achieving this result.

First, the original data are compressed by PCA to extract more representative data. Then, the extracted data are clustered to generate a pseudo label. Therefore, based on PCA and DDC, the decision function and state classification result of FD001 data are shown in Figures 6 and 7, respectively.

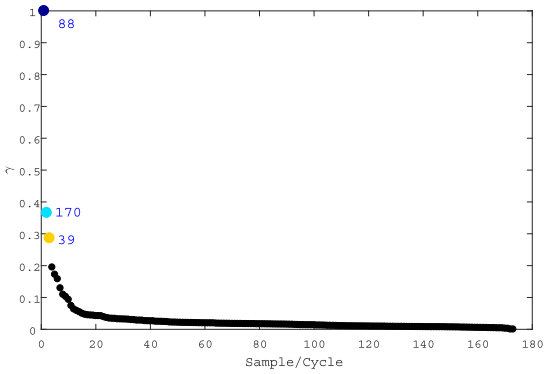


FIGURE 6. The decision function for the data of FD001-01 engine. The selected clustering centers could be the samples with ID 88, 170, and 39.

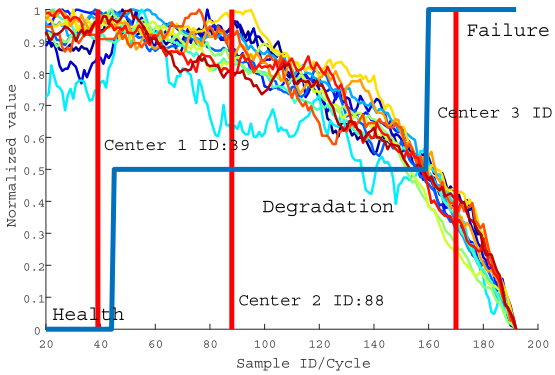


FIGURE 7. The classified states of FD001-01 engine. There are three states including health, degradation, and failure after clustering.

In Figure 6, the samples with IDs 88, 170, 39 have greater decision values than other samples, which are selected as the cluster centers. The pseudo labels as shown in Figure 7 can fully match the trend of engine data, namely, the clustering results are consistent with the real states of the engine.

D. WEIGHT ASSIGNMENT AND FEATURE SELECTION BASED ON FUZZY BAYESIAN RISK MODEL

After obtaining the pseudo label of engine data, the C-MAPSS dataset can be transformed into a decision system that includes conditional attributes (the sensors of the engine) and decision attribute (the pseudo label). Therefore, the weight of each sensor and optimal feature selection can be yielded based on FBR.

In the PCA method, the original data are mapped into the feature space through spatial transformation, then the principal components are found to achieve the purpose of dimensionality reduction. The reduced data is actually a principal component representation of the original data rather than a subset of the original set. Therefore, the feature selection in our work is based on the original data rather than the data generated by PCA.

The decision system associated with C-MAPSS is combined with the sensors selected from the step of data preprocessing and the array of pseudo labels generated by the DDC model, which can be denoted by $DS = (U, C \cup D)$, where U is the concerned data space, $U = \{x_1, x_2, \cdots, x_m\}$, $C = \{c_1, c_2, \cdots, c_n\}$, $D = \{d_1, d_2, \cdots, d_K\}$ [39]. The relationship of the elements in DS is demonstrated in Figure 8. Regarding the aero-engine dataset, $x_i \in U$ is the sample at a given cycle, $c_j \in C$ is the sensor, $d_k \in D$ is a state of the aero-engine, v_{ij} is a specific collection value corresponding to a certain sensor c_j in a certain cycle i.

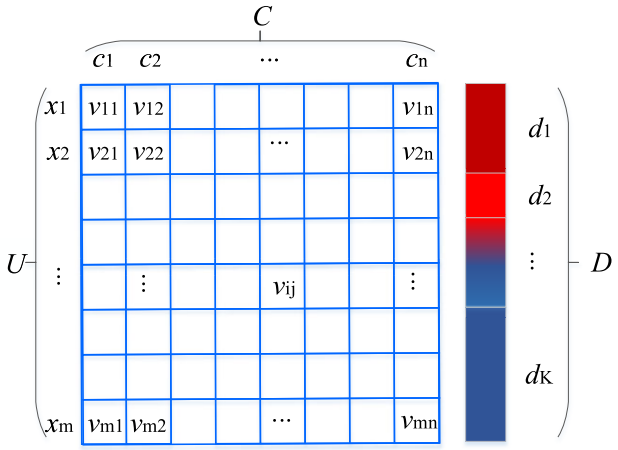


FIGURE 8. The demonstration of decision system.

The basic definitions of FBR are described as follows.

Given a decision system $DS = \{U, C \cup D\}$, $U = \{x_1, x_2, \cdots, x_m\}$, $D = \{d_1, d_2, \cdots, d_K\}$, for an arbitrary $x_i \in U$ induced by an attribute subset c ($c \in C$), it may be classified into any decision class of D using some metrics, but it belongs to a certain class $d_k (d_k \in D)$ according to the corresponding relationship in DS. Therefore, the Bayesian risk of x_i vesting in d_k with respect to c is defined as

$$R_c(d_k|x_i) = \sum_{j=1}^{K} \lambda_j^k(c, x_i) P(d_j|x_i), \tag{7}$$

where $\lambda_j^k(c, x_i)$ is a loss function that measures the loss relative to its attributed class d_k when classifying x_i into the possible class d_j, and $P(d_j|x_i)$ is the probability of x_i belonging to d_j.

With respect to the loss function in Eq. (7), we have proposed a data-driven model named Gaussian kernel loss function that is described as follows [33]:

$$\lambda_j^k(c, x_i) = \begin{cases} \exp(-\mathrm{D}(x_i, \mu_k)) & k \neq j \\ 0 & k = j, \end{cases} \tag{8}$$

where $d_k \in D$ is the inherent class of x_i that comes from DDC, μ_k is the expectation of the samples belonged to class d_k induced by c, $k, j \in \{1, 2, \cdots, K\}$, $\mathrm{D}(\cdot, \cdot)$ is a distance

function, and Euclidean distance is employed in this paper. Usually, we take the loss function as λ_j^k for short.

For the Gaussian kernel loss function, a) if the sample is divided into its inherent class, i.e., $k = j$, then $\lambda_j^k = 0$; b) if the sample is nearer to the expectation, then λ_j^k will be greater; c) if the corresponding standard deviation $\sigma_k = 0$, then $\lambda_j^k = 1$.

With respect to the probability in Eq. (7), we have proposed a fuzzy probability function to flexibly characterize the membership of sample. The fuzzy probability of x_i classifying to d_k is defined as follows:

$$P_c(d_k|x_i) = \frac{\sum\{f(x_i, x_j)|x_j \in d_k\}}{\sum\{f(x_i, x_j)|x_j \in Neigh(x_i)\}}, \tag{9}$$

where d_k is the inherent class of x_i in DS, $Neigh(x_i)$ is a fuzzy neighborhood set of x_i, $f(\cdot, \cdot)$ is the fuzzy similarity relation that is denoted by

$$f(x_i, x_j) = \exp(-\mathrm{D}(x_i, x_j)), \tag{10}$$

where $\mathrm{D}(\cdot, \cdot)$ is the distance function, and we also select the Euclidean distance in our work.

Thus, the fuzzy neighborhood set of x_i can be obtained according to

$$Neigh(x_i) = \{x_j|x_j \in U, f(x_i, x_j) \geq \eta\}, \tag{11}$$

where η is a threshold.

Obviously, there exists $0 < P_c(d_k|x_i) \leq 1$.

Theorem 1: The fuzzy Bayesian risk defined as $R_c(d_k|x_i) = \sum_{j=1}^{K} (\lambda_j^k P(d_j|x_i))$ *is equivalent to* $R_c(d_k|x_i) = \lambda_{\sim k}^k(1 - P(d_k|x_i))$, *where* $d_{\sim k}$ *is the decision class set except* d_k, *and the inherent decision class of* x_i *is* d_k.

Proof: It follows from Eq. (8) that the loss function can be rewritten as $\lambda_{\sim k}^k = \exp(-\mathrm{D}(x_i, \mu_k))$ if $k \neq j$, and $\lambda_k^k = 0$ $(k = j)$. Therefore, the risk function can be written as

$$\begin{aligned}
R_c(d_k|x_i) &= \sum_{j=1}^{K} \lambda_j^k P(d_j|x_i) \\
&= \lambda_1^k P(d_1|x_i) + \lambda_2^k P(d_2|x_i) + \cdots + \lambda_k^k P(d_k|x_i) \\
&\quad + \cdots + \lambda_K^k P(d_K|x_i) \\
&= \lambda_{\sim k}^k P(d_1|x_i) + \lambda_{\sim k}^k P(d_2|x_i) + \cdots + \lambda_k^k P(d_k|x_i) \\
&\quad + \cdots + \lambda_{\sim k}^k P(d_K|x_i) \\
&= \lambda_{\sim k}^k (P(d_1|x_i) + P(d_2|x_i) + \cdots + P(d_K|x_i)) \\
&\quad + \lambda_k^k P(d_k|x_i) \\
&= \lambda_{\sim k}^k (1 - P(d_k|x_i)) + \lambda_k^k P(d_k|x_i) \\
&= \lambda_{\sim k}^k (1 - P(d_k|x_i)).
\end{aligned}$$

□

This theorem shows that we do not need to calculate the loss function and probability of the sample relative to each decision class in the actual calculation. On the contrary, we only need to calculate these two functions of the sample associated with its inherent decision class. This conclusion will greatly reduce the complexity of FBR.

Theorem 2: The fuzzy Bayesian risk satisfies that $0 \leq R_c(d_k|x_i) < 1$.

Proof: According to Theorem 1, the Bayesian risk could be written as $R_c(d_k|x_i) = \lambda_{\sim k}^k(1 - P(d_k|x_i))$, and $\lambda_{\sim k}^k = \exp(-\mathrm{D}(x_i, \mu_k))$, and $0 < \lambda_{\sim k}^k \leq 1$ holds. On the other hand, the probability satisfies $0 < P(d_k|x_i) \leq 1$ according to Eq. (9). Therefore, the above theorem holds. □

Regarding the SOH assessment for aero-engine, the risk in this model refers to the decision-making risk caused by deducing a possible state by the samples generated from sensors. Therefore, given a sensor whose test data brings a less risk when deducing the possible engine state, we will assign it a greater weight. This means that the weight of the sensor is inversely proportional to the risk. The weight of the sensor (i.e., the conditional attribute c in DS) induced by FBR is defined as

$$w_c = \frac{1}{\overline{R}_c + \epsilon}, \tag{12}$$

where $\overline{R}_c = \frac{1}{m}\sum_{i=1}^{m} R_c(d_k|x_i), x_i \in U, d_k \in D, \epsilon$ is a positive number approaching 0 to ensure that the denominator is not 0. This weight assignment model is different from the original one in literature [33], which is denoted by

$$w_c = 1 - \overline{R}_c. \tag{13}$$

For more detailed information about this weight assignment scheme can be found in our previous work [33].

By comparison, it is easy to know that the nonlinear variation relationship in Eq. (12) is more conducive to distinguishing the importance of attributes than the linear variation relationship in Eq. (13). Therefore, instead of Eq. (13), we employ Eq. (12) to assign weight in this work.

Thus, the inequality relationship $0 < w_c \leq 1$ holds according to Theorem 2. Then, the weight vector of DS is $\overline{W} = (\overline{w}_1, \overline{w}_2, \cdots, \overline{w}_n)$, where $\overline{w}_c = w_c / \sum_{c=1}^{n} w_c$, and $0 < \overline{w}_c \leq 1$ holds, $\overline{w}_c = 1$ holds if and only if $n = 1$.

To avoid the problem of too many candidate attributes leading to small differences between the attribute weights, we select those attributes with relatively lower risks as representatives based on the obtained Bayesian risks.

E. STATE ASSESSMENT AND RESULT OUTPUT

Given a set of selected sensors $C = \{c_1, c_2, \cdots, c_n\}$ and its corresponding weight vector $\overline{W} = (\overline{w}_1, \overline{w}_2, \cdots, \overline{w}_n)$, a set of test samples $U = \{x_1, x_2, \cdots, x_m\}$, the test matrix $\mathbf{V}$ can be therefore denoted by

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}, \tag{14}$$

where v_{ij} $(i = 1, 2, \cdots, m, j = 1, 2, \cdots, n)$ is scaled into the range of $[0, 1]$.

The most commonly used operator, i.e., weighted average (WA) operator [40], is selected for the aggregation.

The health score of each sample in U can be calculated as follows:

$$\begin{aligned}\mathbf{S} &= \mathbf{V} \cdot \overline{W}^{\mathrm{T}} \\ &= \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}_{m\times n} \cdot \begin{pmatrix} \overline{w}_1 \\ \overline{w}_2 \\ \vdots \\ \overline{w}_n \end{pmatrix}_{n\times 1} \\ &= (s_1, s_2, \cdots, s_m)^{\mathrm{T}}, \end{aligned} \tag{15}$$

where $\mathbf{S} \in R^{m\times 1}$ is the health score vector of all the samples, the superscript T stands for matrix transposition. Obviously, $0 \le s_i \le 1$ holds.

Based on the health score, we can obtain two outputs of SOH assessment denoted by state membership degree and health degree, one is the horizontal comparison state relative to the same type of engine, the other is the vertical comparison trend state relative to its own historical state. The two outputs are generated by two fuzzy models that are described as follows.

Output 1 (State of Horizontal Comparison): Given N training engines with their health scores, we can obtain three fuzzy rules based on these health scores, the generalized fuzzy rules can be represented by Gaussian-shaped fuzzy functions as shown in Figure 9.

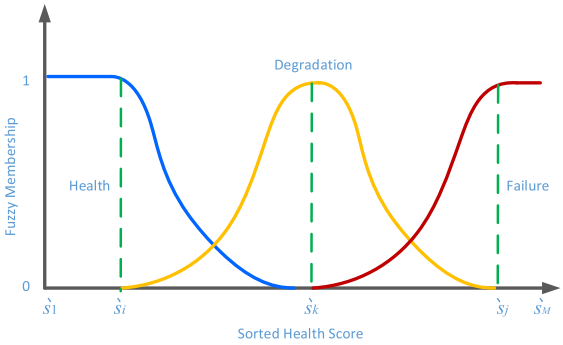


FIGURE 9. The schematic of the fuzzy rules of horizontal comparison.

Specifically, given N training engines with their M lifespan health scores, and the descending sequence is $S = \{\grave{s}_1, \grave{s}_2, \cdots, \grave{s}_M\}$. Then, given three thresholds α, γ and β, $0 < \alpha < \gamma < \beta < 100\%$, the sequence S will be divided into three segments, i.e., $S_1 = \{\grave{s}_1, \grave{s}_2, \cdots, \grave{s}_i\}$, $S_2 = \{\grave{s}_{i+1}, \grave{s}_{i+2}, \cdots, \grave{s}_k, \cdots, \grave{s}_{j-1}\}$, $S_3 = \{\grave{s}_j, \grave{s}_{j+1}, \cdots, \grave{s}_M\}$, where $i = \mathrm{I}(M \cdot \alpha)$, $k = \mathrm{I}(M \cdot \gamma)$, $j = \mathrm{I}(M \cdot \beta)$, $\mathrm{I}(\cdot)$ is a rounding function, $i + 1 < k < j - 1$.

Therefore, the health degree is represented as a set of fuzzy membership degrees indicating the degree of subordination to the three states, i.e., health state, degradation state, and failure state. Given a test engine with its current parameters, then we can calculate its health score using Eq. (15). Thus, the set of fuzzy membership degrees can be easily obtained according to the fuzzy rules as shown in Figure 9.

Output 2 (State of Vertical Comparison): In the data of C-MAPSS, the initial states of these engines are healthy, and then they go into a degraded state at some time during their operation. Therefore, in the longitudinal comparison, the initial state of each engine can be calibrated as its relative health state, and the termination state of each engine can be considered as its complete failure state. In the practical engineer, if there is not enough full life data to support, the termination health state is usually given by the engine manufacturer.

Based on the above idea, the initial health and the terminal health scores of each engine can be used to build a mapping relationship and assess the vertical health of a single-engine.

Specifically, given N training engines with their lifespan health scores, the initial health scores are $H_{in} = \{s_{in1}, s_{in2}, \cdots, s_{inN}\}$, the terminal health scores are $H_{ter} = \{s_{ter1}, s_{ter2}, \cdots, s_{terN}\}$. For the initial health scores, their ascending sequence is $\acute{H}_{in} = \{\acute{s}_{in1}, \acute{s}_{in2}, \cdots, \acute{s}_{inN}\}$, its corresponding sequence of terminal health scores is $\acute{H}_{ter} = \{\acute{s}_{ter1}, \acute{s}_{ter2}, \cdots, \acute{s}_{terN}\}$, where the terminal sequence $\acute{H}_{ter}$ does not guarantee strict ascending or descending order.

After getting the sequences of initial and terminal health scores, we can use some fitting methods to obtain the initial and terminal health baselines. Given two fitted functions $\Theta(x, \acute{s}_{in})$ and $\Phi(x, \acute{s}_{ter})$, where x is a dimensionless sequence with no physical meaning used to ensure that the two functions are in the same coordinate system, $\acute{s}_{in}$ is an initial health score, $\acute{s}_{ter}$ is a terminal health score. In fact, in the actual calculation, the above two ranking health scores can be generalized into general health scores, namely, we have $\Theta(x, s_{in})$ and $\Phi(x, s_{ter})$. Therefore, given a test engine with its initial and current parameters, we can calculate its initial and current health scores s_{in} and s_{curr} based on Eq. (15). Then, we have a certain x induced by the initial fitted function $\Theta(x, s_{in})$. Subsequently, we have the corresponding terminal health score s_{ter} of this test engine based on the terminal fitted function $\Phi(x, s_{ter})$. Finally, the health degree Dg of the test engine can be denoted by

$$Dg = \frac{s_{curr} - s_{ter}}{s_{in} - s_{ter}}, \tag{16}$$

where $s_{in} > s_{ter}$. If $Dg > 1$ holds, we set Dg as 1. If $Dg < 0$ holds, we let Dg be 0.

The above calculation process can be expressed as the training part and the testing part as shown in Figure 10.

III. COMPARISON OF THE COMBINATION OF DDC AND FBR

In this experiment, we will use the public datasets to verify the effectiveness of the combination of DDC and FBR that is the core of our proposed model. The validity of FBR has been verified and analyzed in detail in literature [33]. Therefore, this paper focuses on the combined ability, i.e., we only compare the effectiveness of different clustering methods with FBR on the weight assignment. This comparison experiment is also the foundation of algorithm validation for the

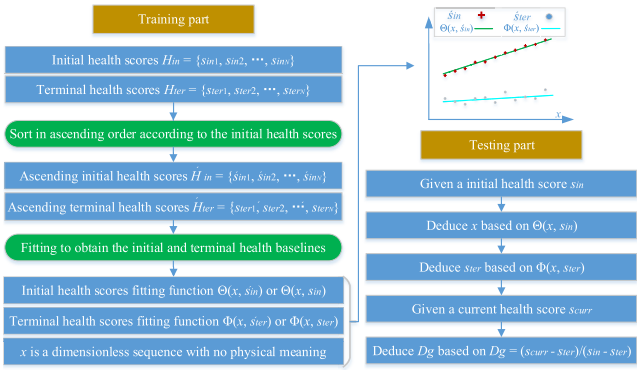


FIGURE 10. The flow chart of vertical comparison.

subsequent SOH assessment. In other words, the main goal of the comparison experiment is to find a suitable combination approach for our SOH assessment framework.

A. EXPERIMENT SETTING

In this experiment, some University of California Irvine (UCI)[1] datasets are used to be the objects, and the details of the UCI datasets are shown in Table 5. Therein, the feature is the number of conditional attributes in DS and class is the number of decision categories. It is worth noting that these datasets are for classification, we utilize these datasets to verify our method that is also to cater to the statement in Section I: many scholars consider that the SOH assessment and classification problem are equivalent. This is the first reason why we choose these classification datasets rather than regression datasets, e.g., the C-MAPSS dataset.

TABLE 5. The details of the used UCI datasets.

ID	Data	Samples	Feature	Class
1	Banknote	1372	4	2
2	Blood	748	4	2
3	Ecoli	336	7	8
4	Iris	150	4	3
5	Seeds	210	7	3
6	Vote	435	16	2
7	Wdbc	569	30	2
8	Wine	178	13	3

This paper continues to use the evaluation criterion of attribute weight assignment given in literature [33]. The classification accuracy of the decision system can effectively measure the importance of conditional attributes. Pearson Correlation Coefficient (PCC) [41] is an effective means to measure the similarity of spatial vectors. Therefore, we employ some classification accuracy evaluation methods as the references and PCC as the tool to measure the pros and cons of the attribute weight assignment strategy. This is the second reason why we choose these classification datasets instead of regression datasets in this experiment.

With the aid of Weka,[2] we employ as many as ten classifiers and 10-fold cross-validation to guarantee that the results are highly credible. Therein, the used classifiers in Weka are C4.5(J48), REPTree, NaiveBayes, SVM(SMO), IBk, Bagging, LogitBoost, FilteredClassifier, JRip, and PART, and the default parameters in Weka are selected. The weights produced by classification accuracies are the average values of the ten classifiers. All of the following experiments are run on the same platform.

B. COMPARISON EXPERIMENT

Regarding the clustering methods combined with FBR, some mature and reliable clustering methods, such as K-means [42], K-medoids [43], and Fuzzy C-means (FCM) [44] are selected to compare to our used DDC. In this case, only the conditional attributes in the employed UCI datasets (as shown in Table 5) are used, which could be considered as information systems without labels. The number of clusters in these information systems is the same as those of decision classes. This is the third reason why we choose these classification datasets rather than regression datasets.

From the data in Table 5, it can be seen that the minimum number of conditional attributes is 4, thus, the maximum reduction dimensions produced by PCA is set as 4. The weight assignment results produced by the four clustering methods and classification accuracy are shown in Figure 11. Therein, the five color bars indicate five kinds of dimensions, i.e., 1-dimension, 2-dimension, 3-dimension, 4-dimension, and the original dimensions, respectively. The red lines are the weights determined only by FBR from the original decision systems. The parameter η in all the approaches is 0.8 as recommended by literature [33].

On the basis of the same test conditions, datasets, and evaluation criterion, from the results in Figure 11, the following analyses and conclusions can be drawn:

1) In most cases, the results of the eight datasets produced by DDC are the best ones, and those are better than others for DDC if the processed datasets have been reduced to 3-dimension by PCA. These results show that in the compared clustering methods, DDC with the aid of its skillful clustering idea is more suitable for the combination with FBR to assign better weight.
2) Usually, the dimensionality reduction by PCA can improve the PCCs between the weights assigned by the combination methods and classification accuracies. This shows that our hypothesis in Subsection II-C is reasonable.
3) Obviously, the PCCs generated by the combination methods are less than those from the original decision systems, namely, the weights determined by the clustering method and FBR from the information system are not better than those produced by FBR from the decision system. The reason of this conclusion is that we use both the inherent label as the comparison scheme

[1] http://archive.ics.uci.edu/ml/datasets.php

[2] https://www.cs.waikato.ac.nz/ml/weka/, v3.6.13

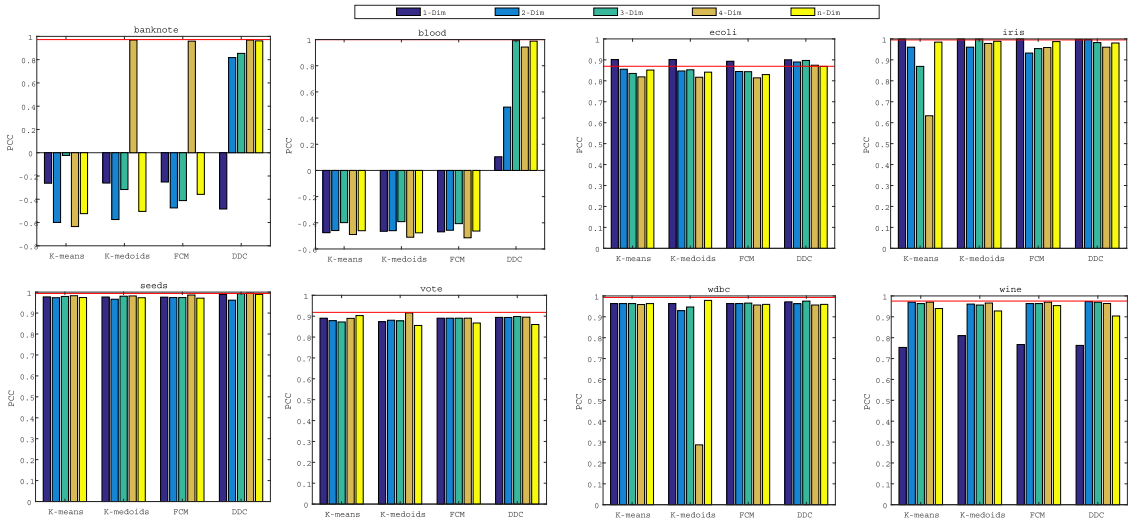


FIGURE 11. The PCCs between weight assignment results produced by the four clustering methods and classification accuracy.

and the evaluation standard. Therefore, this conclusion has no effect on the weight assignment method for the information system that uses the combination of clustering method and FBR.

IV. VALIDATION OF THE PROPOSED METHODOLOGY BY C-MAPSS DATASET

In this experiment, we will utilize the most commonly used FD001 dataset in the C-MAPSS datasets to verify the effectiveness of the proposed methodology. There are 100 engines in the FD001 dataset, we select the monitoring series of the first 80 engines as the training data, and the series of the last 20 engines as the test data.

In the following part, we will gradually illustrate how to utilize the SOH assessment model.

Step 1 (Data Preprocessing): As mentioned in Subsection II-B, we first reduce the noise in all the data and then delete the attributes with constant values to achieve data preselection, and then the attributes with IDs 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, 21 are fed into the following experiment.

For the normalization of training data, we first select the maximum and minimum values of the training data that are as shown in Table 6.

TABLE 6. The maximum and minimum values of the training data.

ID	2	3	4	7	8	9	11
Max	644.5	1,614.9	1,441.5	556.1	2,388.6	9,244.6	48.5
Min	641.2	1,571.0	1,382.3	549.9	2,387.9	9,021.7	46.9
ID	12	13	14	15	17	20	21
Max	523.4	2,388.6	8,293.7	8.6	399.0	39.4	23.6
Min	518.7	2,387.9	8,099.9	8.3	388.0	38.2	22.9

Then, by using the two normalization models (Eqs. (2) - (3)), all the training values are scaled into the range of [0, 1]. It is worth noting that such trained normalization models do not consider the range of test data. Therefore, during the processing of test data by using the trained normalization boundary, if the normalized result $\bar{x}$ of test data x satisfied that $\bar{x} > 1$ or $\bar{x} < 0$, the normalized boundary value in Table 6 should be updated according to the test data x.

Step 2 (Pseudo Label Generation: Based on the combination of PCA and DDC, the pseudo clustering centers of the training dataset will be easily produced. The statistical result of the clustering centers associated with the three given states, i.e., health, degradation, failure, is shown in Figure 12, from where it can be seen that the distribution of clustering centers of the three states generally presents a regular order although there are some overlapping regions. Furthermore, the following statistical results can be obtained:

1) In the region of health state, the clustering centers in the statistical region of samples with IDs between 30 and 45 are as many as 27, and this region has the most clustering centers.
2) The clustering centers of degradation state are concentrated in the region of samples with IDs between 75 and 90.
3) The region of samples with IDs between 120 and 135 has the most clustering centers regarding the failure state.
4) The clustering centers of degradation state has the largest span that covers the samples with IDs from 30 to 210.

Step 3 (Weight Assignment and Feature Selection): With the help of the generated pseudo labels for the training dataset, we can calculate the risk of each attribute/sensor by FBR. The cumulative risk and average risk of each attribute are shown in Table 7 and Figure 13, respectively.

To improve the accuracy of health assessment, we carry out a secondary screening on the selected attributes and select those attributes with smaller decision risks as the final

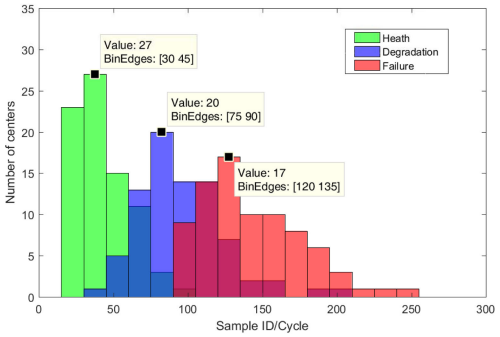


FIGURE 12. The statistical result of clustering centers associated with three given states. The marked text corresponds to the cluster centers.

TABLE 7. The cumulative risk produced by all the training data for each attribute.

ID	2	3	4	7	8	9	11
Risk	6.07	7.18	4.88	5.82	7.59	8.40	4.72
ID	12	13	14	15	17	20	21
Risk	5.23	7.56	8.52	5.95	6.79	5.83	5.83

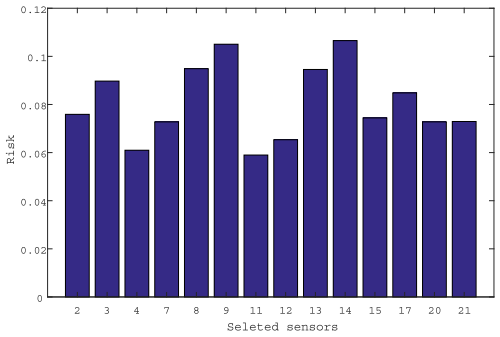


FIGURE 13. The average risk of each attribute.

selected attributes. Thus, according to the above risk results, we select four attributes with the smallest risks, i.e., Sensor IDs 11, 4, 12, and 7. Then, the corresponding weights are 0.2647, 0.2642, 0.2373, and 0.2337.

The selected sensors are static pressure at HPC outlet (Ps30), total temperature at LPT outlet (T50), ratio of fuel flow to Ps30 (phi), and total pressure at HPC outlet (P30), respectively. It can be known from the given information of FD001 dataset that this type of engine has experienced a high pressure compressor degradation failure. This failure will directly cause the outlet pressure of HPC to change, including the outlet static pressure (Ps30) and the outlet total pressure (P30). Besides, the parameters phi and Ps30 are directly related and have a proportional relationship, thus the degradation state of HPC can also be characterized by the ratio phi. Affected by the degradation of HPC, the performance of the entire engine also shows a degradation trend, and the main performance can be obtained by observing the exhaust temperature, that is, the outlet temperature at LPT outlet T50. Consequently, the selected four parameters Ps30, T50, phi, and P30 are closely related to the degradation failure of HPC, which illustrates that the risk calculation and weight assignment strategy of our work are effective.

Step 4 (Assessment): In this step, we can obtain the health scores of the 80 training datasets by using Eq. (15), and the results are shown in Figure 14. Based on these health scores, we can fit the horizontal comparison rules and the vertical comparison rules.

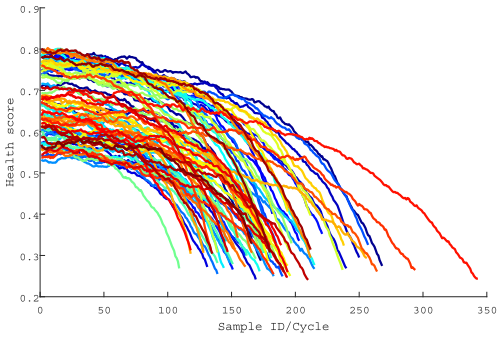


FIGURE 14. The health scores of the 80 training datasets. On the one hand, considering all the training engines, the health scores show the characteristics of interval in the longitudinal direction under the premise of a fixed horizontal axis. Thus, it can realize the horizontal health degree expression. On the other hand, considering a single-engine, its health score presents a monotonous decreasing trend along the horizontal axis. Thus, the vertical health degree can be expressed.

Regarding the horizontal comparison, the number of training engines N is 80, and the number of lifespan health scores M is 14618. If we set the three thresholds α, γ and β as 20%, 50% and 80%, the three segments will be $S_1 = \{\grave{s}_1, \grave{s}_2, \cdots, \grave{s}_{2924}\}$, $S_2 = \{\grave{s}_{2925}, \grave{s}_{2926}, \cdots, \grave{s}_{7309}, \cdots, \grave{s}_{11693}\}$, $S_3 = \{\grave{s}_{11694}, \grave{s}_{11695}, \cdots, \grave{s}_{14618}\}$. Therein, $\grave{s}_1 = 0.8026$, $\grave{s}_{2924} = 0.7007$, $\grave{s}_{7309} = 0.5837$, $\grave{s}_{11694} = 0.4752$.

Based on the above basic information, the fuzzy rule functions for horizontal comparison can be induced by the way of function fitting, which are listed as follows:

$$R_{hea}(x) = \begin{cases} 0, & x \le 0.584, \\ \exp\left(-\left(\dfrac{x-0.693}{0.03715}\right)^2\right), & 0.584 < x < 0.693, \\ 1, & x \ge 0.693, \end{cases} \tag{17}$$

$$R_{degr}(x) = \begin{cases} 0, & x \le 0.475, \\ \exp\left(-\left(\dfrac{x-0.584}{0.03715}\right)^2\right), & 0.475 < x < 0.693, \\ 0, & x \ge 0.693, \end{cases} \tag{18}$$

$$R_{fail}(x) = \begin{cases} 1, & x \le 0.475, \\ \exp\left(-\left(\dfrac{x-0.475}{0.03715}\right)^2\right), & 0.475 < x < 0.584, \\ 0, & x \ge 0.584, \end{cases} \tag{19}$$

where R_{hea} is the fuzzy rule of health state, R_{degr} is the fuzzy rule of degradation state, and R_{fail} is the fuzzy rule of failure state. The three fuzzy rules are shown in Figure 15.

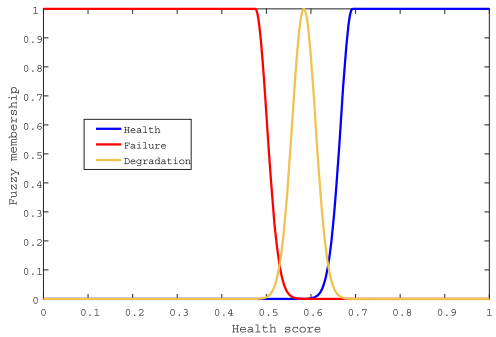


FIGURE 15. The fuzzy rules of horizontal comparison.

Regarding the vertical comparison, we can obtain the initial and terminal health scores H_{in} and H_{ter}, and their corresponding ascending sequences $\acute{H}_{in}$ and $\acute{H}_{ter}$. Then, two fitted functions can be yielded with the aid of the least square fitting method. In the process of fitting, we try to fit the health score data using the 1st to 15th order functions and measure the fitting result with relative error. The fluctuation of data fitted by higher-order function can not be used to describe the general trend of terminal health score, which is not conducive to calculation of the vertical health score. The final results show that the result of the first-order function is the most ideal. Therefore, the two fitting functions are shown as follows:

$$\Theta(x) = 0.0036x + 0.5279, \tag{20}$$

$$\Phi(x) = 0.0007x + 0.2748, \tag{21}$$

where Θ and Φ are the initial fitted function and terminal fitted function. The fitted functions are shown in Figure 16.

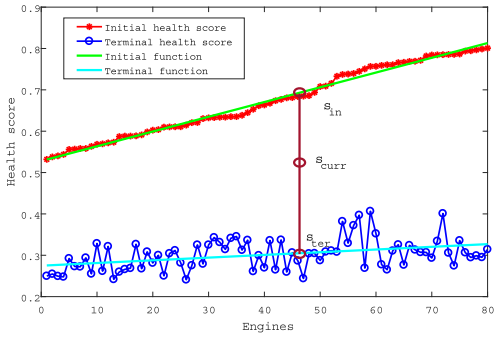


FIGURE 16. The fuzzy rules of vertical comparison. Thus, the vertical health degree can be calculated based on the function of brown line.

Therefore, given an engine with its health score s_{curr} and initial health score s_{in}, we can first calculate the corresponding x as follows according to Eq. (20):

$$x = \frac{s_{in} - 0.5279}{0.0036}. \tag{22}$$

Then, we have the terminal health score by using Eq. (21):

$$s_{ter} = 0.0007 \times \frac{s_{in} - 0.5279}{0.0036} + 0.2748. \tag{23}$$

Finally, we can calculate the health degree Dg of this engine by Eq. (16) based on the given current health score s_{curr}, the given initial health score s_{in}, and the calculated terminal health score s_{ter}.

From the results in Figure 16, it can be seen that the fitting function of terminal health score is almost a horizontal line, which is in line with the form of engine retirement condition, that is, for the same type of engine, the given retirement condition for each engine is consistent, but the actual performance is slightly different. However, the statistical result shows that the terminal health score presents a typical normal distribution as shown in Figure 17. Therefore, to avoid the uncertainty caused by the actual utilization, we use the fitting function (Eq. 21), which can effectively avoid overfitting and sensitivity to the initial health score, and can reveal the main characteristic of the terminal health score.

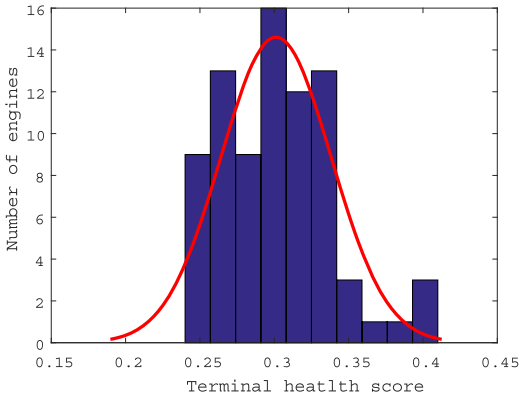


FIGURE 17. The statistical result of terminal health score with the average value 0.3012, which presents the characteristic of a normal distribution.

So far, we have two different approaches to assessing the SOH of the engine. The first one is to obtain the three states of the given engine through horizontal comparison, and the second one is to obtain the health of the given engine through vertical comparison.

Unfortunately, there is no uniform and appropriate method for evaluating the correctness and accuracy of the results of health assessment [11], and we have not been able to find or design such a reasonable method or indicator for this verification. As we all know, the purpose of equipment health assessment is to clarify the current performance state of equipment, so as to make a reasonable decision for the utilization and maintenance of equipment. Therefore, we assess the running state of the engine from two aspects of horizontal and vertical comparison. In this regard, we visually express the assessment results and screen the validity and correctness of the results through ablation analysis, which comes from the idea in literature [45]. The assessment results are shown in Figure 18. Therein, the green line indicates the health score induced by horizontal comparison, the line composed of

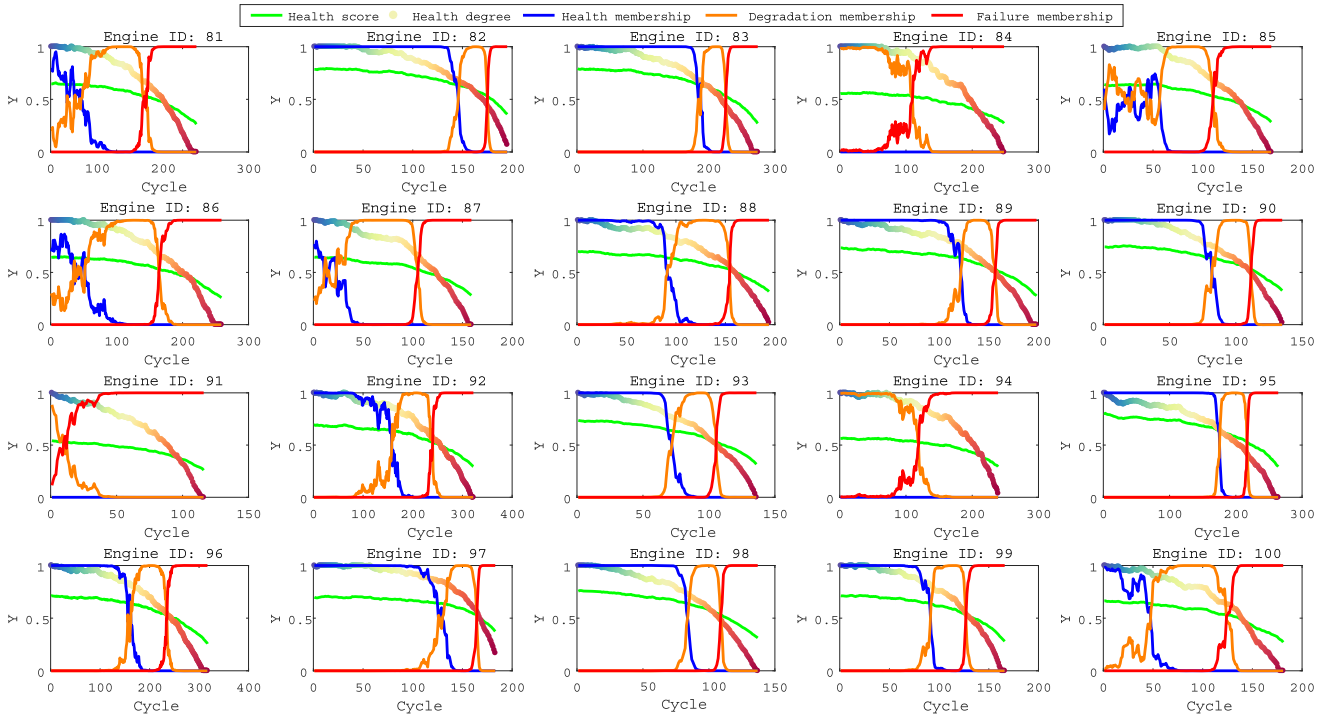


FIGURE 18. The assessment results of test engines.

colored dots stands for the health degree obtained by vertical comparison.

In Figure 18, we draw the health score, the health degree, and the membership degrees for the three states. The vertical axis Y refers to the aforementioned three results, the horizontal axis represents the cycle number of the engine. From the results in Figure 18, we can provide the following ablation analysis:

1) All the health scores (the green lines in Figure 18) calculated by our proposed method show good monotonous downward trends. These smooth and monotonous downward trends are very helpful for assessing the SOH of the engine and the subsequent trend prediction. If we assess the SOH of engine only using the raw dataset including 21 sensors (as shown in Figure 3), it is difficult to intuitively derive such a smooth monotonous trend.
2) In the vertical comparison considering the historical state, the calculated health degree (marked with color dots) shows a smooth and monotonous downward trend in the range of [0, 1], which is a good representation of the SOH variation during the engine's entire lifespan. The health degree presents a relative state, which is a measurement index relative to one's own historical state.
3) In the horizontal comparison considering the other engines' state, according to the fuzzy rules established by 80 training datasets, the possibility that the test engine belongs to the three states at each cycle is obtained. From the results shown in Figure 18, it can be seen that the boundary of the three states is very clear, which is helpful for judging the state of the engine and performing fault diagnosis.
4) Through two aspects of health assessment, we can obtain the current state of the engine may belong to a certain state, and the engine's variation relative to its initial state. Both of the two results fulfill the initial purpose of the health assessment, namely, clarifying the current performance state of the engine to make a reasonable decision for the utilization and maintenance of the engine. This also proves that the method proposed in this paper is effective and correct.

V. CONCLUSION

In order to improve the generalization ability of state-of-health assessment, we proposed a data-driven framework that mainly includes data preprocessing, pseudo label generation, weight assignment and feature selection, and assessment. To realize the function of label generation and reduce the decision risk of assessment, we designed a combination model based on DDC and FBR, which could extract a relatively optimal parameter subset and assign weights for these parameters. To evaluate the state of the engine more comprehensively, we raised up two fuzzy models from two aspects including the horizontal comparison of engine group and vertical comparison of single-engine, and two indicators,

i.e., state membership degree and health degree, are designed. The results of the comparison experiment show that the combined model of DDC and FBR has an advantage over other compared models. The results of the verification experiment based on the C-MAPSS dataset illustrate that the proposed methodology is correct and effective.

Based on the proposed methodology and experimental results, some more conclusions can be drawn as follows:

1) The systematic design of the SOH assessment system for equipment is very meaningful to fully grasp the equipment operation status and also helps to expand the theoretical scope of the PHM system and enhance its application value.
2) Accurately obtaining the intermediate calculation results and approaching the final result with the minimum risk, e.g., our combination method, will generate a clearer and more accurate evaluation result, which is more conducive to decision-making.
3) Multi-dimensional SOH assessment of equipment, e.g., the horizontal and vertical assessments in our work, can more fully reveal the operating status of the equipment, thereby providing a more comprehensive reference for equipment support.

In our future work, the proposed framework can be used to assess the SOH of other types of equipment, and it can also incorporate other models to improve the adaptability of the framework. In addition, it is worth noting that there is no comparison and index evaluation experiments in our case verification part in Section IV, because we have not been able to find or design some reasonable and recognized evaluation methods or indexes. Therefore, the design and development of evaluation methods and indexes for SOH assessment have been the focus of our research, we will also carry out relevant research under the background of this paper in our future work.

REFERENCES

[1] F. Lu, J. Wu, J. Huang, and X. Qiu, "Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm," *Aerosp. Sci. Technol.*, vol. 84, pp. 661–671, Jan. 2019.

[2] J. Yu, "Aircraft engine health prognostics based on logistic regression with penalization regularization and state-space-based degradation framework," *Aerosp. Sci. Technol.*, vol. 68, pp. 345–361, Sep. 2017.

[3] C. M. Ezhilarasu, Z. Skaf, and I. K. Jennions, "The application of reasoning to aerospace integrated vehicle health management (IVHM): Challenges and opportunities," *Prog. Aerosp. Sci.*, vol. 105, pp. 60–73, Feb. 2019.

[4] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.

[5] M. Tahan, E. Tsoutsanis, M. Muhammad, and Z. A. A. Karim, "Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review," *Appl. Energy*, vol. 198, pp. 122–144, Jul. 2017.

[6] Z. Liu, Z. Jia, C.-M. Vong, J. Han, C. Yan, and M. Pecht, "A patent analysis of prognostics and health management (PHM) innovations for electrical systems," *IEEE Access*, vol. 6, pp. 18088–18107, 2018.

[7] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2849–2858, May 2019.

[8] Y. Jiang, S. Yin, and O. Kaynak, "Data-driven monitoring and safety control of industrial cyber-physical systems: Basics and beyond," *IEEE Access*, vol. 6, pp. 47374–47384, 2018.

[9] H. Hanachi, C. Mechefske, J. Liu, A. Banerjee, and Y. Chen, "Performance-based gas turbine health monitoring, diagnostics, and prognostics: A survey," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 1340–1363, Sep. 2018.

[10] M. Soualhi, K. Nguyen, K. Medjaher, D. Lebel, and D. Cazaban, "Health indicator construction for system health assessment in smart manufacturing," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 45–50.

[11] Y. Cui, J. Shi, and Z. Wang, "Quantum assimilation-based State-of-Health assessment and remaining useful life estimation for electronic systems," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2379–2390, Apr. 2016.

[12] D. Xiao, Y. Huang, H. Wang, H. Shi, and C. Liu, "Health assessment for piston pump using LSTM neural network," in *Proc. Int. Conf. Sensing,Diagnostics, Prognostics, Control (SDPC)*, Aug. 2018, pp. 131–137.

[13] S. Zhao, V. Makis, S. Chen, and Y. Li, "Health assessment method for electronic components subject to condition monitoring and hard failure," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 138–150, Jan. 2019.

[14] X. Bai, Z. An, Y. Hou, and Q. Ma, "Health assessment and management of wind turbine blade based on the fatigue test data," *Microelectron. Rel.*, vol. 75, pp. 205–214, Aug. 2017.

[15] C. Deng, C. Duan, P. Liang, and J. Wu, "Study on health assessment and residual useful life prediction of wind turbine," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jul. 2018, pp. 1112–1117.

[16] G. Zhang, S. Li, N. Yang, H. Su, J. Wang, L. Huang, and H. Wang, "Research on general aircraft cluster health assessment method," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–6.

[17] J. Shi, Y. Zhao, and X. Duan, "Study about health assessment of circuit based on three-state health DBN algorithm," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chengdu)*, Oct. 2016, pp. 1–6.

[18] Y. Song, D. Liu, and Y. Peng, "Data-driven on-line health assessment for lithium-ion battery with uncertainty presentation," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2018, pp. 1–7.

[19] E. Ramasso, M. Rombaut, and N. Zerhouni, "Joint prediction of continuous and discrete states in time-series based on belief functions," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 37–50, Feb. 2013.

[20] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Long short-term memory for machine remaining life prediction," *J. Manuf. Syst.*, vol. 48, pp. 78–86, Jul. 2018.

[21] G. Chen, J. Chen, Y. Zi, and H. Miao, "Hyper-parameter optimization based nonlinear multistate deterioration modeling for deterioration level assessment and remaining useful life prognostics," *Rel. Eng. Syst. Saf.*, vol. 167, pp. 517–526, Nov. 2017.

[22] A. L. Ellefsen, S. Ushakov, V. Æsøy, and H. Zhang, "Validation of data-driven labeling approaches using a novel deep network structure for remaining useful life predictions," *IEEE Access*, vol. 7, pp. 71563–71575, 2019.

[23] C. Che, H. Wang, Q. Fu, and X. Ni, "Combining multiple deep learning algorithms for prognostic and health management of aircraft," *Aerosp. Sci. Technol.*, vol. 94, Nov. 2019, Art. no. 105423.

[24] Z. Zhao, B. Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Rel. Eng. Syst. Saf.*, vol. 164, pp. 74–83, Aug. 2017.

[25] H. Zhou, J. Huang, and F. Lu, "Reduced kernel recursive least squares algorithm for aero-engine degradation prediction," *Mech. Syst. Signal Process.*, vol. 95, pp. 446–467, Oct. 2017.

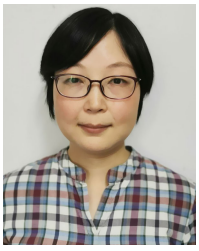
[26] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.

[27] R. Moghaddass and M. J. Zuo, "An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process," *Rel. Eng. Syst. Saf.*, vol. 124, pp. 92–104, Apr. 2014.

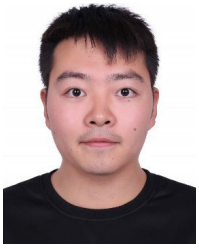
[28] E. Ramasso and T. Denoeux, "Making use of partial knowledge about hidden states in HMMs: An approach based on belief functions," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 2, pp. 395–405, Apr. 2014.

[29] O. O. Aremu, D. Hyland-Wood, and P. R. McAree, "A relative entropy weibull-sax framework for health indices construction and health stage division in degradation modeling of multivariate time series asset data," *Adv. Eng. Inform.*, vol. 40, pp. 121–134, 2019.

[30] P. Tamilselvan, Y. Wang, and P. Wang, "Deep belief network based state classification for structural health diagnosis," in *Proc. IEEE Aerosp. Conf.*, Mar. 2012, pp. 1–11.
[31] Y. Lin, M. Chen, and D. Zhou, "Online probabilistic operational safety assessment of multi-mode engineering systems using Bayesian methods," *Rel. Eng. Syst. Saf.*, vol. 119, pp. 150–157, Nov. 2013.
[32] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, pp. 196–1492, Jun. 2014.
[33] M. Suo, B. Zhu, Y. Zhang, R. An, and S. Li, "Fuzzy bayes risk based on mahalanobis distance and Gaussian kernel for weight assignment in labeled multiple attribute decision making," *Knowl.-Based Syst.*, vol. 152, pp. 26–39, Jul. 2018.
[34] D. K. Frederick, J. A. DeCastro, and J. S. Litt, "User's guide for the commercial modular aeropropulsion system simulation (C-MAPSS)," NASA, Washington, DC, USA, Tech. Memorandum TM-2007-215026, 2007.
[35] Y. Xu, W. Zhang, and H. Wang, "A conflict-eliminating approach for emergency group decision of unconventional incidents," *Knowl.-Based Syst.*, vol. 83, pp. 92–104, Jul. 2015.
[36] M. Suo, Y. Cheng, C. Zhuang, Y. Ding, C. Lu, and L. Tao, "Extension of labeled multiple attribute decision making based on fuzzy neighborhood three-way decision," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17731–17758, Dec. 2020.
[37] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni, "Direct remaining useful life estimation based on support vector regression," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2276–2285, Mar. 2017.
[38] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. 26, no. 1, pp. 17–32, Feb. 1981.
[39] M. Suo, L. Tao, B. Zhu, X. Miao, Z. Liang, Y. Ding, X. Zhang, and T. Zhang, "Single-parameter decision-theoretic rough set," *Inf. Sci.*, vol. 539, pp. 49–80, Oct. 2020.
[40] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, no. 1, pp. 183–190, Jan. 1988.
[41] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Beijing, China: Posts & Telecom Press, 2011.
[42] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Berkeley Symp. Math. Statist. Probab.*, 1967, pp. 281–297.
[43] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.
[44] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, Jan. 1984.
[45] M. Suo, R. An, D. Zhou, and S. Li, "Grid-clustered rough set model for self-learning and fast reduction," *Pattern Recognit. Lett.*, vol. 106, pp. 61–68, Apr. 2018.



NING MA received the M.S. degree in systems engineering from the School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, in 2011. She is currently a Lecturer with the Department of Aviation Maintenance NCO Academy, Air Force Engineering University, Xinyang, China. Her current research interests include health management, decision support, fault diagnosis, artificial intelligence, and maintenance of aero-engine.



FAN YANG received the B.S. degree in systems engineering from the School of Reliability and Systems Engineering, Beihang University, in 2018, where he is currently pursuing the master's degree with the School of Reliability and Systems Engineering. His research interests include artificial intelligence, big data, deep learning, fault diagnosis, prognostics, and health management.

LAIFA TAO received the B.S. and Ph.D. degrees from the School of Reliability and Systems Engineering, Beihang University, in 2010 and 2014, respectively. He is currently an Associate Professor with the School of Reliability and Systems Engineering, Beihang University. His research interests include fault diagnosis, prognostics, health state assessment, optimization and determination, and health management for electromechanical systems.

MINGLIANG SUO (Member, IEEE) received the B.S. degree from the School of Energy and Power Engineering, Beihang University, China, in 2009, and the M.S. and Ph.D. degrees from the School of Astronautics, Harbin Institute of Technology, in 2013 and 2018, respectively. Since 2018, he has been with the School of Reliability and Systems Engineering, Beihang University, where he is currently a Postdoctoral Fellow. He is also a member of the Science and Technology on Reliability and Environmental Engineering Laboratory. His current research interests include artificial intelligence, state-of-health assessment, decision-making, data mining, granular computing, prognostics and health management, and maintenance of aero-engine and aircraft.