

Article

# A Hybrid Degradation Modeling and Prognostic Method for the Multi-Modal System

Jun Peng <sup>1,†</sup>, Shengnan Wang <sup>1,†</sup>, Dianzhu Gao <sup>2</sup>, Xiaoyong Zhang <sup>1,\*,†</sup>, Bin Chen <sup>2,†</sup>,  
Yijun Cheng <sup>2,†</sup>, Yingze Yang <sup>1,†</sup>, Wentao Yu <sup>3,†</sup> and Zhiwu Huang <sup>2,†</sup>

<sup>1</sup> School of Computer Science and Engineering, Central South University, Changsha 410075, China; pengj@csu.edu.cn (J.P.); snwang@csu.edu.cn (S.W.); yangyingze@csu.edu.cn (Y.Y.)

<sup>2</sup> School of Automation, Central South University, Changsha 410075, China; gaodianzhu@csu.edu.cn (D.G.); chenbinrxrg@163.com (B.C.); yijuncheng@csu.edu.cn (Y.C.); hzw@csu.edu.cn (Z.H.)

<sup>3</sup> College of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China; wtyu\_csuft@126.com

\* Correspondence: zhangxy@csu.edu.cn

† This paper is an extended version of our paper published in 2018 21st International Conference on Intelligent Transportation Systems.

Received: 31 December 2019; Accepted: 14 February 2020; Published: 18 February 2020



**Abstract:** Engineering systems typically go through complicated degradation processes, partially due to their multiple operating modes. Therefore, how to accurately estimate their remaining useful life is a critical issue. To address this challenge, a hybrid degradation modeling and prognostic method for the multi-modal system is proposed. Firstly, the cumulative dynamic differential health indicator is constructed for the multi-modal switching system using a multi-objective optimization approach. The long-term cumulative degradation assessment model is constructed based on the gated recurrent unit. Then, considering that the damage in the latest stage has a significant impact on the remaining useful life, the time window is used to extract local features of the sequence, including energy features and statistical features. The latest stage degradation is predicted based on the light gradient boosting machine. Finally, model averaging is used to integrate the two predicted results, which is expected to improve the prognostic robustness. The proposed model is evaluated with synthetic analysis and NASA turbofan aero-engine datasets. Extensive experimental results demonstrate the proposed method provides a better characterization of the degradation status of the system and provides a higher estimation accuracy than existing methods.

**Keywords:** remaining useful life; multi-modal system; prognostic; health indicator; local feature

## 1. Introduction

Prognostic health management is a critical piece of technology for providing early warnings of system failure [1], aiming to minimize economic loss and potentially catastrophic accidents. As a vital issue of prognostic health management, the remaining useful life (RUL), which is defined as the duration from the current time to the failure time, should be predicted accurately. Therefore, remaining useful life prediction has drawn much attention from both industry and academia.

An engineering system can work under a single operating condition or multiple operating conditions depending the application scenario. The sensor data often have stationary aging trends or features when the system works under a simple condition. Most researchers concentrate on the prognostic health management of system with the single operation condition [2–10]. However, in practice, the engineering system often works under multiple operating conditions [11] and switches operation modes according to the changes in the operating environment and its states. For example,

the battery charger typically works either on the constant-current (CC) mode or constant-voltage (CV) mode, depending on the battery voltage [12]. When the system works under multiple operating conditions, the stationary feature is difficult to select and cannot be utilized merely to master the underlying physical processes [13]. Therefore, it is necessary to extract and utilize the features of the collected information to the maximum extent. In conclusion, there are two problems for the prognostic analysis of multi-modal systems.

On the one hand, the different number and order of mode switching makes the aging trend of systems more complicated. The aging data of the single-mode system is often used to predict remaining useful life directly. But for multi-modal systems, inter-frame dynamic information hides in aging data [14]. Thus, there is an urgent requirement to study the physical description of mode switching in the life and the cumulative effect of mode switching on the aging degree of the system. [15] used sensor sequence information to reveal the hidden characteristics of the data with multiple operating conditions to some extent. However, the work ignores the time interval between signals in the same operating mode to make the degradation characteristics of signals under varying operating modes.

Health indicator (HI) has shown superiority in the problem of systematic health management [16]. The suitable health indicator could simplify prognostic modeling and calculate accurate prediction results [17]. Some work has constructed physics health indicators related to the physics of degradation from sensor signals based on statistical methods or signal processing methods, such as root mean square [18] and kurtosis [19]. Each physical feature can be considered as a physical health indicator [20] which contains partial information about the aging process. For multi-modal systems, many physical indicators will be proposed based on the mode switching rule. Remaining useful life prediction with only one physical feature is one-sided [21], but information overlaps between a large number of physical features. Besides, the collected sensor data usually includes noise signals, which may lead to mutual exclusion or invalid information between features. Using all physical health indicators of sensors to predict consumes a lot of computing resources, especially for large-scale datasets. Therefore, the physics health indicators are fused to construct the virtual health indicator. Virtual health indicators usually present the degradation trends without physics meanings; see [22] and [23]. Therefore, the establishment of an explanatory health indicator applicable to the complex working state of multi-modal systems is a challenge for the accurate prediction of remaining useful life.

The system aging is a long-term cumulative process. [24] selected a recurrent neural network for remaining useful life prediction from the perspective of long-term prediction performance and obtained an excellent prediction effect. To get better results for long sequences, the long short-term memory neural network was used to estimate remaining useful life in [25]. Less computation was required in [26] based on the gated recurrent unit. However, these works did not consider the impact of the switching of the working state on the machine. The prediction model should be established more suitably for cumulative aging characteristics of a multi-modal system.

On the other hand, the system's ability to withstand the damage is different for the system in different aging states. The ability decreases gradually with the increase of action times. The same shock will cause more severe damage to the system in a relatively late aging stage than in an earlier aging stage. Thus, it is significant to study the latest physical degradation description of mode switching in the life and the effect of mode switching on the aging degree of the system. The authors in [27] used the time window to extract the data features in segments, but did not consider the change of the system's ability to withstand damage. The statistical features of short sequence sensor data were extracted in [28], but the frequency-domain features of wave data remained unexplored. It is necessary to study a comprehensive local feature extraction method to analyze the latest-stage aging predictions specifically.

The multi-dimensional local features lead to a large amount of computation for the prediction model. Thus, the algorithm with less computational overhead is selected as the prediction model that could improve the prediction speed. The decision tree ensemble methods are popular in solving practical prediction problem based on the interpretability and effectiveness, such as gradient boosting decision tree [29] and extreme gradient boosting [30]. However, these methods calculate

the information gain and find the optimal partition point by scanning all samples. Efficiency and scalability are difficult to satisfy when calculating large amounts of data. The light gradient boosting machine (LightGBM) has proven to have a faster and more accurate performance in [31]. In addition, the light gradient boosting machine has great potential in remaining useful life estimation when the sensor data contains a lot of noise, which is very common in industrial applications.

After solving the above problems, a decision-level fusion model is needed to get the final aging prediction result. Most previous researchers intended to remedy the lack of performance of one of the algorithms, such as convergence and computational complexity [32,33], rather than focusing on the physical characteristics of the system, which are important for remaining useful life prognostic and are analyzed in this paper. The averaging method has the advantages of low computational complexity and the ability to adjust parameters online. In [34], the strategy, which uses different methods to construct the degradation models and combines them for remaining useful life prediction, is defined as the hybrid prognostics approach.

In this paper, a hybrid degradation modeling and prognostic method for multi-modal systems is proposed to address shortcomings mentioned above. First, a cumulative dynamic differential health indicator (CDD-HI) is proposed based on the physical characteristics with the existence of mode switching. The proposed health indicator is used to predict long-term cumulative degradation by the gated recurrent unit network. Then, local features are constructed to predict latest-stage degradation by light gradient boosting machine. Finally, based on averaging method, the results of the gated recurrent unit and light gradient boosting machine are integrated to obtain remaining useful life. In conclusion, this paper aims to construct a hybrid degradation modeling and prognostic method for multi-modal systems, which combines the long-term cumulative degradation assessment and latest-stage degradation assessment, and takes into account changes in system capabilities to withstand damage under different aging conditions. The main contributions of this paper are as follows:

- 1) The cumulative dynamic physical difference feature is constructed to deal with the superposition effect of different operation modes, and gains the physical characteristics with the existence of mode switching, which is hidden in the inter-frame data. The health indicators are constructed based on the physical properties to improve the interpretability and robustness of a prediction model. The combination of a cumulative dynamic differential health indicator and the gated recurrent unit will further enhance the accuracy of cumulative degradation prediction.

- 2) During the latest stage of system operation, the system's ability to resist damage is usually the worst in its life time. The time window is used to extract local features, including statistical and energy features, which could mine aging information more fully to predict latest-stage degradation with the light gradient boosting machine.

- 3) The experiments on theoretical examples and the accessible turbofan aero-engine degradation dataset from NASA, and comparisons with the related model validate the effectiveness and superiority of the proposed method.

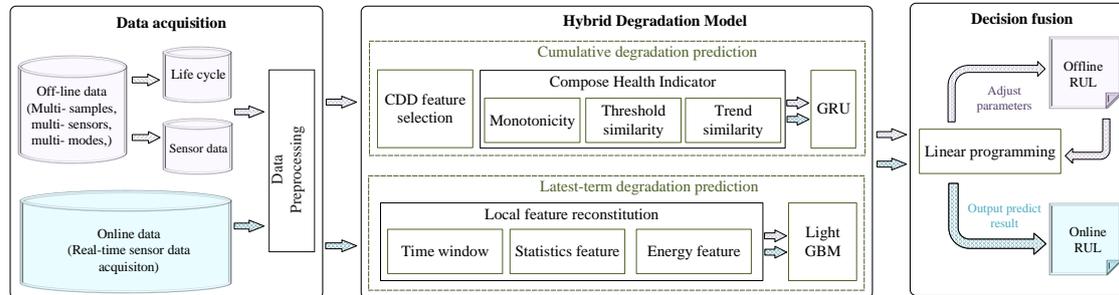
The rest of this paper is organized as follows. The proposed model is detailed in Section 2. Section 3 evaluates the proposed model on the NASA degradation dataset and presents an analysis of the results. The conclusion and prospects are given in Section 4.

## 2. Degradation Modeling and Prognostic Hybrid

This paper develops a hybrid degradation modeling and prognostic method for the multi-modal system to characterize the underlying degradation process accurately. Considering that different combinations of modes have different effects on the degradation process, the method predicts remaining useful life from two aspects: the long-term cumulative degradation assessment based on a cumulative dynamic differential health indicator, and the latest stage degradation assessment based on local features.

As shown in Figure 1, the proposed method works in an offline stage and an online stage. In the offline stage, the degradation model is learned from history monitoring data. The learned model is

used to predict the remaining useful life in real-time in the online stage.  $S$  is defined as the input dataset.  $s_{i,k}(t)$  is the data sequence of  $k$ th sensor data of the  $i$ th sample (system) to represent the change in data as the number of actions in time  $t$  changes. Then, the remaining useful life estimation process is as follows.



**Figure 1.** Structure of the proposed hybrid degradation model. The offline data are used to adjust parameters of prognostic model. The learned prognostic model is used to predict remaining useful life based on the online data.

### 2.1. Long-Term Cumulative Degradation Assessment with Cdd-Hi

#### 2.1.1. Cumulative Dynamic Difference Feature Extraction

For the single mode system, static features are extracted from the raw sensor data to represent the degradation trend. For multi-modal systems, different combinations of switching order have different effects on the degradation. So for multi-modal systems, the whole degradation cannot be built only based on stationary features. To solve it, the proposed cumulative dynamic difference features consist of four parts, which are sthe mooth feature ( $F_1$ ), cumulative modes ( $F_2$ ), difference times ( $F_3$ ) and difference data ( $F_4$ ).

In order to leave out the noise in the original signal  $S$ , smoothing is used to obtain the smoothed feature  $F_1$ . The size of  $F_1$  is the number of sensor  $K$ .  $F_2$  represents the cumulative action times of each operation mode. The size of  $F_2$  equals the number of modes  $M$ . To present the information of mode switching,  $F_3$  and  $F_4$  are constructed.  $F_3$  represents the forward differences of action times of each operation mode, each of which is gained by calculating the time interval between the current action time and the last action time of each mode.  $F_4$  represents the forward differences of the sensor data of each operation mode ( $d$ ). The differences of sensor data between the current action time and the last action time of each mode are calculated to characterize data changes between different modes.

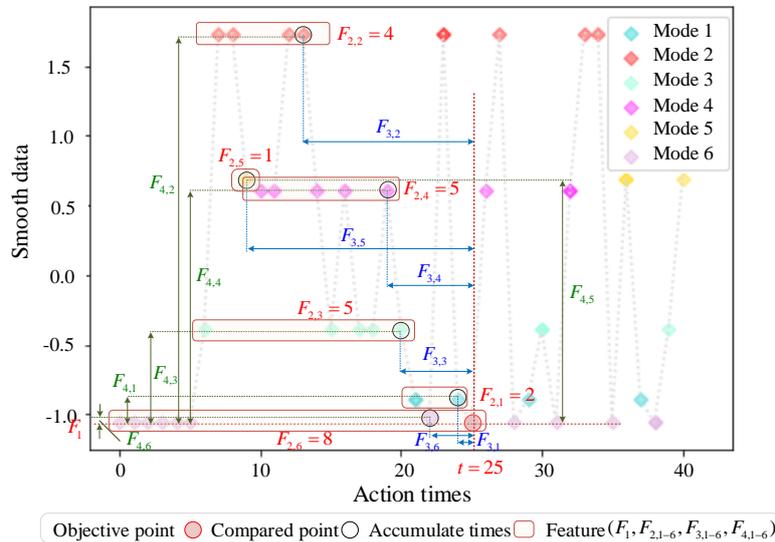
The cumulative dynamic physical difference feature consists of the above four parts for the multi-modal system to extract the switching order of modes. Figure 2 illustrates an example of the cumulative dynamic difference feature extraction process for a sensor. The number of modes is  $m = 6$ ; the length of data sequence is 40. The cumulative dynamic difference feature is the  $F_t = [F_1, F_{2,1}, F_{2,2}, F_{2,3}, F_{2,4}, F_{2,5}, F_{2,6}, F_{3,1}, F_{3,2}, F_{3,3}, F_{3,4}, F_{3,5}, F_{3,6}, F_{4,1}, F_{4,2}, F_{4,3}, F_{4,4}, F_{4,5}, F_{4,6}]$  for the system in  $t$ th action time.

#### 2.1.2. Composite Health Indicator Construction

To avoid the redundancy among different features, this paper develops a feature-level fusion method to combine cumulative dynamic physical difference features to get a composite health indicator based on the multi-objective optimization model. The health indicator  $H(t)$  is regarded as a linear combination of each cumulative dynamic physical difference feature  $f_o$ , as Equation (1),

$$H(t) = w_1f_1(t) + w_2f_2(t) + \dots + w_of_o(t), \tag{1}$$

where  $o$  is the dimension of cumulative dynamic physical difference feature, and  $w$  is the weight coefficient of each cumulative dynamic difference feature,  $-1 < w < 1$ . To represent the actual degradation trend of system, the interpretable health indicator is constructed based on three properties.



**Figure 2.** The example of cumulative dynamic physical difference feature construction. Calculate the proposed features of  $t = 25$ . Compared points are the last action times of all modes.  $F_1$  is the value of smooth data.  $F_{2,1}, F_{2,2} \dots F_{2,6}$  is the cumulative time of each mode.  $F_{3,1}, F_{3,2} \dots F_{3,6}$  are the difference times to compared points.  $F_{4,1}, F_{4,2} \dots F_{4,6}$  are the difference data to compared points.

*Property 1:* The degradation of the system should be monotonic in general without external maintenance.

To establish the monotonic health indicator, the slack variable  $\varepsilon_{i,t}$  is introduced to measure the violation of monotonicity of health indicator. The  $\varepsilon_{i,t}$  of system  $i$  and action time  $t$  is  $\varepsilon_{i,t} = \max(H_{i,t} - H_{i,t+1}, 0)$ . Assuming that the indicator is monotonically increasing, Equation (2) tries to minimize the total weighted sum of violations to ensure the monotonicity of the health indicator.

$$\begin{aligned} \min_{w, \varepsilon_{i,t}} & \sum_{i=1}^p \sum_{t=1}^{q_i-1} \varepsilon_{i,t} \\ \text{s.t. } & w' M' \mathbf{1} = 1, \quad M w \geq 0, \quad \varepsilon_{i,t} \geq 0, \end{aligned} \tag{2}$$

where  $p$  is number of system samples, and  $q$  is the maximum observation epochs (action times) of each sample.  $M$  is a diagonal matrix to express the degradation trend information. For example, if the feature shows an increasing (decreasing) trend, the diagonal entry of  $M$  is 1 (−1). The different system  $i$  varies in the life cycle length  $q_i$ .

*Property 2:* The first hitting time of failure is regarded as the failure threshold  $\theta_i$  of each system. The systems with the same functions and components should have similar fault thresholds under the same environmental conditions.

The predictability is expressed by minimizing the variance of  $\theta_i$ , as  $\sum_{i=1}^p (\theta_i - \bar{\theta})^2 / (p - 1)$ . Assuming  $A \in R^{p \times o}$  is the matrix recording the feature of the last observation epoch in full life-cycle of each system, the variance can be translated as the quadratic term  $w' A' B A w$  based on Equation (3):

$$\begin{aligned} & ((Aw)'(Aw) - p((1'Aw)/(p))^2)/(p - 1) \\ &= (w' A' A w - (w' A' 11' A w) / p) / (p - 1) \\ &= w' A' ((I - (11') / p) / (p - 1)) A w \\ &= w' A' B A w, \end{aligned} \tag{3}$$

where  $B$  is a symmetric matrix;  $B = (I - O/p) / (p - 1)$ .  $I$  is the identity matrix, and  $O$  is the matrix which each entry equals 1. Above all, Property 2 could be formulated with an optimization problem, as Equation (4):

$$\begin{aligned} & \min_w w' A' B A w \\ & s.t. w' M' 1 = 1, M w \geq 0. \end{aligned} \tag{4}$$

*Property 3:* For multi-modal systems, although different mode combinations would result in different short-term fluctuations of each sample, there are still similar degradation trends for those samples from the long-term perspective.

Samples of multi-modal systems with identical structures usually have different lengths of life cycle. Therefore, it is necessary to calculate the similarities between different lengths of degradation sequences. In this paper, the dynamic time warping (DTW) method is used to measure the similarities of different health indicator series.

The similarity between  $H_\alpha(t_\alpha)$ ,  $t_\alpha \in [0, q_\alpha]$  and  $H_\beta(t_\beta)$ ,  $t_\beta \in [0, q_\beta]$  is calculated with the dynamic programming method based on finding the min-distance warp path.  $D = D_1, D_2, \dots, D_d, D_{N_{path}}, N_{path} \in [\max(q_\alpha, q_\beta), q_\alpha + q_\beta]$ . The obtained warp path  $D$  indicates the correspondence between  $H_\alpha(t_\alpha)$  and  $H_\beta(t_\beta)$ , permissible point-to-point, and many-to-one point correspondences.

The obtained warp path  $D$  should satisfy three constraints: 1) The constraint of endpoint ensures that the starting point and the endpoint are consistent when the sequences are aligned, as  $D_1 = (1, 1)$ ,  $D_{N_{path}} = (q_\alpha, q_\beta)$ . 2) The continuity constraint is used to limit the excessive expansion of the dynamic time warping and limit the jump correspondence, as  $t_{\beta_{d+1}} - t_{\beta_d} \leq \zeta$ ,  $t_{\alpha_{d+1}} - t_{\alpha_d} \leq \zeta$ , where  $\zeta$  is a positive integer. 3) The monotonicity constraint requires  $t_{\beta_{d+1}} - t_{\beta_d} \geq 0$ ,  $t_{\alpha_{d+1}} - t_{\alpha_d} \geq 0$ . Then, the minimum distance of dynamic time warping is revised as Equation (5),

$$DTW(H_\alpha, H_\beta) = \min_p \sum_{d=1}^{N_{path}} d(H_\alpha(t_{\alpha_d}), H_\beta(t_{\beta_d})), \tag{5}$$

where  $d(H_\alpha(t_{\alpha_d}), H_\beta(t_{\beta_d}))$  is the euclidean distance between two points. To reduce the amount of calculation, the upper triangular matrix of the DTW matrix which consists of  $DTW(H_\alpha, H_\beta)$ ,  $1 \leq \alpha \leq p$ ,  $1 \leq \beta \leq p$  is used to calculate the distance sum of samples. Then, the Property 3 could be formulated as Equation (6),

$$\begin{aligned} & \min_w \sum_{\alpha=1}^{p-1} \sum_{\beta=\alpha}^p \frac{2DTW(H_\alpha, H_\beta)}{(p-1)^2} \\ & s.t. w' M' 1 = 1, M w \geq 0. \end{aligned} \tag{6}$$

Above all, the optimization problem of composite health indicator construction model could be formulated as Equation (7),

$$\begin{aligned} & \min_w \lambda_1 \sum_{i=1}^p \sum_{t=1}^{q_i-1} \varepsilon_{i,t} + \lambda_2 w' A' B A w \\ & + \lambda_3 \sum_{\alpha=1}^{p-1} \sum_{\beta=\alpha}^p \frac{2DTW(H_\alpha, H_\beta)}{(p-1)^2} \\ & s.t. w' M' 1 = 1, M w \geq 0, \varepsilon_{i,j} \geq 0, \lambda' 1 = 1, \\ & -1 < w < 1, \end{aligned} \tag{7}$$

where  $\lambda$  is the parameter adjusting the relative importance of the three properties. Based on solving the multi-objective programming equation to calculate  $(w_1, w_2, \dots, w_o)$ , the composite health indicator  $H(t)$  with good monotonicity and generalization will be calculated from cumulative dynamic physical difference feature with Equation (1).

### 2.1.3. Cumulative Degradation Assessment with Gated Recurrent Unit

The recurrent neural network has good performance when processing arbitrary time-sequences, based on the network structure in which neurons could connect to themselves across time. Because the problem of long-term dependencies, it is hard for the recurrent neural network to learn to store information for very long [35]. To solve the above problems, this paper adopts the gated recurrent unit, which is another recurrent neural network variant, to estimate remaining useful life with a long-term cumulative degradation health indicator.

In the gated recurrent unit, the gradient disappearance or explosion problem is alleviated by introducing an additive structure of the gate. There are two gates to process data information. The reset gate  $r(t)$  is used to adjust the incorporation of current input with the previous memory. The preservation of the previous memory is controlled by the update gate  $u(t)$ . The transition function is formulated as Equations (8) and (9),

$$u(t) = \sigma(W_u H(t) + V_u \hat{y}(t-1) + b_u), \tag{8}$$

$$r(t) = \sigma(W_r H(t) + V_r \hat{y}(t-1) + b_r), \tag{9}$$

where  $W$  and  $V$  represent the weight matrices of  $H(t)$  and  $\hat{y}(t-1)$ , respectively, and  $b$  is the bias vector for the update gate and reset gate. Since the reset gate controls the degree of ignoring the information and the update gate controls the impact of the previous information, the output of current status can be represented as Equation (10):

$$\begin{aligned} \tilde{y}(t) &= \tanh(W_h H(t) + V_h (r(t) \odot \hat{y}(t-1))) \\ \hat{y}(t) &= (1 - u(t)) \odot \hat{y}(t-1) + U(t) \odot \tilde{y}(t). \end{aligned} \tag{10}$$

Regarding the cumulative dynamic differential health indicator  $H(t)$  as input, the corresponding forecast result  $\hat{y}(t)$  could be obtained, which is the long-term cumulative degradation assessment  $\hat{y}(t)_c$ .

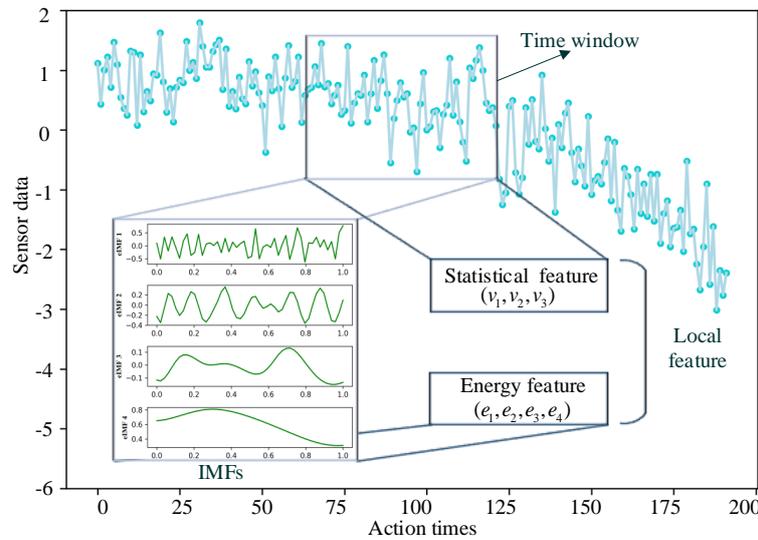
## 2.2. Latest-Term Degradation Assessment Based on Local Features

The aging of an engineering system is usually slow in the early stage and rapid in the later stage. Therefore, it is necessary to study the latest-stage aging data of the system in detail.

### 2.2.1. Construction of Local Aging Features

Constructing local aging features of these complex signal series is significant for obtaining useful aging information. In order to predict latest stage degradation, the overlapped time window is used to decompose the time series. The size of the time window is  $L$ . Two kinds of features, statistical features and energy features, are extracted, as shown in Figure 3.

The statistical features are used to represent the mathematical rules of sequence, which are the average value  $v_1$ , variance  $v_2$  and skewness  $v_3$ . For multi-modal systems, the ensemble empirical mode decomposition (EEMD) algorithm is used to decompose the original signal to intrinsic mode function (IMF) components based on the local characteristic time scale of the raw signal. Among them, the white Gaussian noise with uniform distribution is added in the time-frequency space to reduce the mixing degree of IMF component model and achieve signal continuity in different frequency regions. The algorithm's steps are as follows.



**Figure 3.** The process of local feature extraction based on a sample from NASA’s Prognostics Repository. The local features include statistical features (average value  $v_1$ , variance  $v_2$  and skewness  $v_3$ ) and the energy features ( $e_0, e_1 \dots e_z$ ) which are calculated based on intrinsic mode function (IMF) components.

Firstly, the number of empirical mode decomposition executions is  $N$  and the amplitude coefficient of white noise. In execution epoch  $n$ , the white Gaussian noise series  $g(t)$  is added to the signal data  $s(t)$ , as the  $s_n^{noise} = s(t) + g_n(t)$ . In order to obtain IMF components of  $s_n^{noise}$ , the mean of the upper and lower envelope  $m(t)$  is obtained between the local maximum and minimum of  $s_n^{noise}$ . Let  $l(t) = s_n^{noise} - m(t)$ ; if  $l_n(t)$  satisfies the condition of intrinsic mode function component, it means that it is the first IMF. When it does not, replace  $s_n^{noise}$  with  $l_n(t)$  and repeat calculation  $l_n(t)$  until the IMF conditions are satisfied. The  $l_n(t)$  from  $s_n^{noise}$  is subtracted to obtain the remainder. The process of extracting an IMF component is repeated until all components are separated from the  $s_n^{noise}$ . The remainder becomes a monotonic function  $r_n(t)$ , Equation (11):

$$s_n^{noise} = \sum_{z=1}^Z s_{nz}^{imf}(t) + r_n(t), \tag{11}$$

where  $Z$  is the number of IMFs which contains the local characteristics of original signal. In each empirical mode’s decomposition execution time, a different white noise series is added to  $s(t)$ . After that, we calculate the mean value of all IMFs in  $N$  times as in Equation (12):

$$s_z^{imf}(t) = \frac{1}{N} \sum_{n=1}^N s_{nz}^{imf}(t). \tag{12}$$

$s_z^{imf}(t)$  is the  $z$ th IMF based on signal decomposition. After the signal is decomposed, each IMF component represents a stable signal of different frequency. The frequency contains important aging information too. The energy variation of each frequency band could show the aging state of the system. Therefore, the energy entropies of different IMF components are extracted as the local aging features, which can be calculated as  $e_{s^{imf}(t)} = \sum_{t=1}^L |s^{imf}(t)|^2$ .

### 2.2.2. Latest-Term Degradation Assessment Based on Local Features and LightGBM

In present research, the gradient boosting decision tree has good performance when predicting short sequences, but the computational complexity needs to be improved. In this paper, the improved gradient boosting decision tree, the light gradient boosting machine, is used to predict the remaining useful life of latest stage series based on local features. LightGBM used the leaf-wise growth with depth-limiting strategy instead of the traditional level-wise strategy to train decision trees to reduce errors and improve accuracy. The leaf-wise strategy is used to find the leaf with the largest split gain from all current leaves, and then split each time to reduce more errors and obtain higher accuracy at the same split time. To ensure high efficiency and prevent over-fitting, the maximum depth limit is added to the leaf-wise strategy.

For the local features  $(v_1, v_2, v_3, e_0, e_1 \dots e_z)$ ,  $J$  decision trees are trained to predict the residuals of the prior models. If  $\varphi_j$  is the learning function of the  $j$ th decision tree and the  $\hat{y}_i^{(j)}$  is the prediction of for  $i$ th sample at the  $j$ th iteration; the process can be expressed as Equation (13):

$$\begin{aligned} \hat{y}_i^{(0)} &= 0 \\ \hat{y}_i^{(1)} &= \varphi_1(x_i) = \hat{y}_i^{(0)} + \varphi_1(x_i) \\ \hat{y}_i^{(2)} &= \varphi_1(x_i) + \varphi_2(x_i) = \hat{y}_i^{(1)} + \varphi_2(x_i) \\ &\vdots \\ \hat{y}_i^{(J)} &= \sum_{j=1}^J \varphi_j(x_i) = \hat{y}_i^{(J-1)} + \varphi_j(x_i) \end{aligned} \tag{13}$$

In each iteration, the current model  $\hat{y}_i^{(j)}$  is reserved and the learning residual function  $\varphi$  is learned by minimizing the objective function, as in Equation (14):

$$\ell^{(J)} = \sum_{i=1}^p \text{loss}(y_i, \hat{y}_i^{(J)}) + \sum_{j=1}^J \Omega(\varphi_j), \tag{14}$$

The function consists of two parts. One is a loss function regarding the difference between the prediction  $\hat{y}_i^{(J)}$  and the actual  $y_i$ . The other is the regular term  $\sum_{j=1}^J \Omega(\varphi_j)$  that penalizes the complexity of model, as in Equation (15):

$$\Omega(\varphi_j) = \mu Z + \frac{1}{2} \tau \sum_{z=1}^Z \omega_z^2, \tag{15}$$

where  $\mu$  is the penalty parameter for the number of leaves  $Z$ ,  $\omega$  is the weight of a leaf and  $\tau$  is the penalty parameter. Considering that the samples with larger gradients play a more important role in calculating information gain, the gradient-based one-side sampling (GOSS) strategy [31] is used to segment the data. Firstly, the local feature dataset is sorted in descending order based on absolute values of gradients. Then, data instances with large gradients are retained, and samples with small gradients are randomly sampled while splitting out the set according to the estimated variance gain.

$$\hat{y}_i = \sum_{j=1}^J \varphi_j(x_i), \varphi_j \in S_{tree}. \tag{16}$$

As Equation (16) shows, predictions of all decision trees are added to obtain the latest stage prediction.  $S_{tree}$  is a space that contains all possible structures of trees.

### 2.3. Decision-Level Fusion Based on Model Averaging

The system's ability to withstand the damage is different for the system in different aging states. In order to integrate the long-term cumulative degradation and the latest stage degradation, the fusion

model is used in this paper. The fusion model includes two kinds of methods, one is based on a bagging, boosting or stacking method, which consumes a lot of training time. Therefore, the averaging method is used to make a final decision in this paper. The long-term cumulative prediction  $\hat{y}_c$  is obtained as in Section 2.2. The latest stage prediction  $\hat{y}_l$  is obtained as in Section 2.2.2. Finally, the final result  $\hat{y}$  is calculated as in Equation (17) using the averaging method.

$$\hat{y} = \gamma\hat{y}_c + (1 - \gamma)\hat{y}_l. \tag{17}$$

where coefficient  $\gamma$  is set as in [36].

### 3. Experiment and Discussion

In this section, we investigate the performance of proposed method using a synthetic dataset and an aero-engine turbofan dataset from NASA. The synthetic dataset is used to theoretical analysis. The turbofan aero-engine degradation dataset was published publicly in NASA’s Prognostics Repository [37]; it simulates the degradation of engines in four working environments, so should verify the generality of the proposed model.

#### 3.1. Evaluation Metrics

Samples working in different environments have different life cycle lengths. In this paper, therefore, the remaining useful life is set as the percentage of useful times. A linear function represents the degradation trajectory of system, and the remaining useful life  $y_t$  is expressed by Equation (18).

$$y(t) = 1 - \frac{t}{q}. \tag{18}$$

$\hat{y}$  and  $y$  represent the predicted result and the real result. If  $\hat{y} < y$ , the prediction result is considered as an early prediction. If  $\hat{y} > y$ , the prediction result is considered as a late prediction. Considering that late predictions have more serious consequences, two evaluation metrics, the root mean squared error (RMSE) and scores are used to evaluate the performance of the proposed method. Root mean squared error gives an equal penalty to late and early predictions, as in Equation (19):

$$RMSE = \sqrt{\frac{1}{q} \sum_{t=1}^q y'(t)^2}, \tag{19}$$

which is  $y'(t) = \hat{y}(t) - y(t)$ . The score metric was proposed in the PHM 2008 Data Challenge [38] and defined as an asymmetric function that penalizes late predictions more than the early predictions. The expression of score function is a weighted sum of remaining useful life errors, as in Equation (20):

$$\begin{aligned} Scores &= \sum_{t=1}^q score(t), \\ score(t) &= \begin{cases} e^{-\frac{y'(t)}{13}} - 1, & if\ y'(t) < 0 \\ e^{\frac{y'(t)}{10}} - 1, & if\ y'(t) \geq 0. \end{cases} \end{aligned} \tag{20}$$

#### 3.2. Theoretical Analysis

This section is used to discuss the performance of the proposed method based on different trend synthetic data. The unchanged data has constant outputs throughout the life of the system, which cannot provide useful information to facilitate prediction [1]. Most of the systems show a slow aging

trend in the early stage and a fast aging trend in the later stage. Therefore, the synthetic data is constructed with exponential functional form (Equation (21)).

$$S_{syn}(t) = \exp(t) + Gaussian(t), \tag{21}$$

where,  $\exp(t) = e^{a_{exp}t+b_{exp}}$  is the exponential function to generate exponential data for different trends by adjusting parameter  $a_{exp}$  and  $b_{exp}$ .  $Gaussian(t)$  is the Gaussian random variables with the mean  $\mu_g$  and the variance  $\sigma^2$  to simulate noise. The probability density of Gaussian noise follows the standard normal distribution (Equation (22)).

$$N_{exp}(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu_g)}{2\sigma^2}}. \tag{22}$$

The normal distribution is symmetrical about  $t = \mu_g$ .  $\mu_g$  and  $\sigma$  are used to adjust the data fluctuation. Therefore, different parameter settings,  $(a_{exp}, b_{exp}, \mu_g, \sigma^2)$ , are used to generate synthetic data for different trends. In the theoretical analysis, small scale samples are used for simulation analysis. The parameter settings of synthetic data samples are shown in Table 1.

**Table 1.** Parameter settings of synthetic data samples.

$a_{exp}, b_{exp}$	Parameter	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
(0.01, 0.25)	$\mu_g$	1.58	1.93	2.94	3.45	3.62	3.88	4.22	4.60	4.80	4.84
	$\sigma^2$	0.04	0.06	0.08	0.10	0.10	0.11	0.12	0.13	0.14	0.14
(0.02, 0.5)	$\mu_g$	4.91	4.99	5.36	6.86	12.21	12.52	17.20	17.84	18.44	19.98
	$\sigma^2$	0.14	0.15	0.16	0.20	0.37	0.379	0.52	0.54	0.55	0.60
(-0.04, 1.3)	$\mu_g$	20.02	22.43	23.27	24.67	25.81	25.92	28.11	28.45	29.07	33.13
	$\sigma^2$	0.60	0.67	0.70	0.74	0.78	0.79	0.85	0.86	0.88	1.00
(-0.06, 1.6)	$\mu_g$	33.80	33.84	33.91	34.04	34.08	35.59	35.86	37.51	37.68	39.32
	$\sigma^2$	1.02	1.02	1.02	1.03	1.03	1.07	1.08	1.13	1.14	1.19

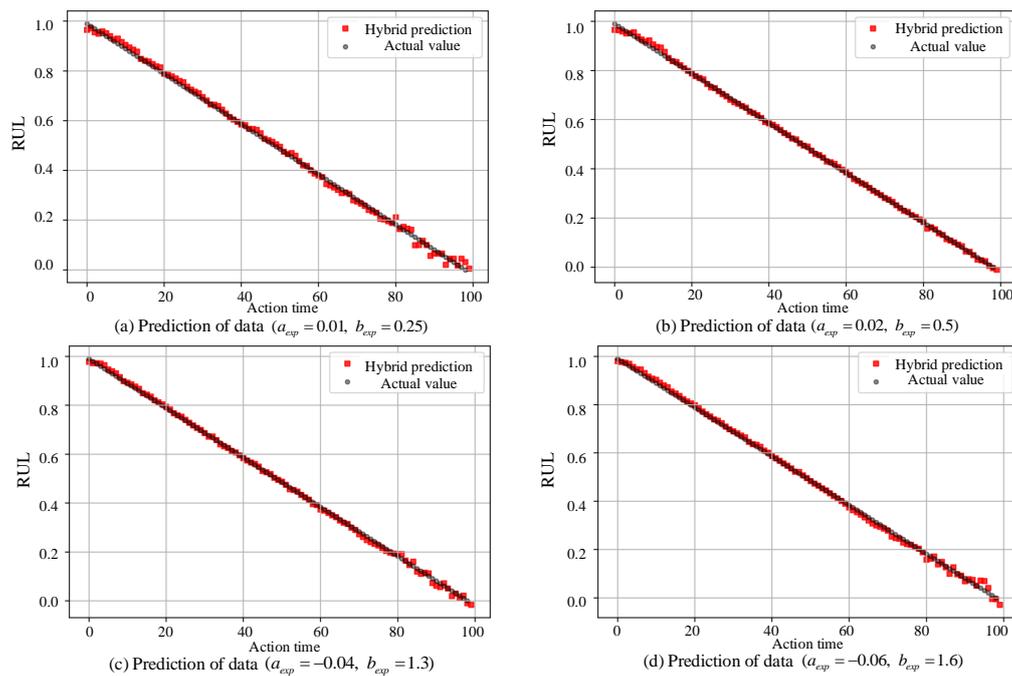
The dataset has 40 synthetic data samples, which have four different exponential forms and forty different Gaussian random variables. When  $a_{exp} > 0$ , the trend is ascending and the trend is descending for  $a_{exp} < 0$ . Then, the synthetic dataset is used to study the sensitivity of the proposed method to different types of input data. The results are shown in Figure 4.

RMSE are (0.100, 0.101, 0.102, 0.103) and scores are (0.174, 0.120, 0.131, 0.149) for different exponential forms  $(a_{exp} = 0.01, b_{exp} = 0.25)$ ,  $(a_{exp} = 0.02, b_{exp} = 0.5)$ ,  $(a_{exp} = -0.04, b_{exp} = 1.3)$ ,  $(a_{exp} = -0.06, b_{exp} = 1.6)$ . Experimental results show that the method has good prediction performance on samples with different trends. The conclusions in the section are used as the basis for sensor selection during the experimental phase. More discussion on the method performance will be given in the experimental simulation section based on NASA’s dataset.

### 3.3. Experimental Simulation

#### 3.3.1. Benchmark Data Description

The commercial modular aero-propulsion system simulation (C-MAPSS) turbofan aero-engine dataset [39] is widely used in prognostic studies. This dataset simulated the degradation of aero-engine in four working environments to get the corresponding sub-datasets FD001, FD002, FD003 and FD004. Each subset consists of a training set, a test set and corresponding remaining useful life values. The training samples degrade until the system failure, and the last time is regarded as the failure time of the engine unit. For test samples, the sensor data recording is stopped before the system fails and the failure period is recorded for verification. The details and statistical data of each subset are listed in Table 2. The monitoring data in each sample consist of 21 sensors (e.g., total temperature, pressure at fan inlet, physical core speed, etc.) and three operational settings which jointly determine the system’s working mode [40].



**Figure 4.** The prediction results of the proposed model for synthetic data. (a) The sample of exponential function with ( $a_{exp} = 0.01, b_{exp} = 0.25$ ). (b) The sample of exponential function with ( $a_{exp} = 0.02, b_{exp} = 0.5$ ). (c) The sample of exponential function with ( $a_{exp} = -0.04, b_{exp} = 1.3$ ). (d) The sample of exponential function with ( $a_{exp} = -0.06, b_{exp} = 1.6$ )

**Table 2.** Detailed information of C-MAPSS dataset.

Subset	Fault Type	Operation Mode	Training Scale	Testing Scale	Max-Lifespan	Selected Sensor
FD001	1	1	100	100	362	13
FD002	1	6	260	259	378	-
FD003	2	1	100	100	525	11
FD004	2	6	249	248	543	-

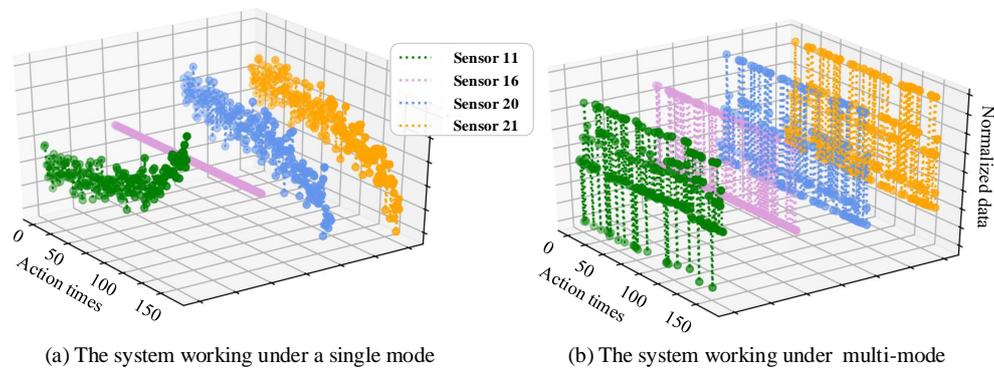
The sample of FD001 suffered high-pressure compressor failure under a single operating condition. For FD002, the sample suffered the high pressure compressor failure under six operating conditions. In FD003, the sample suffered high pressure compressor and fan failure under a single operating condition, while in FD004 the sample suffered high pressure compressor and fan failure under six operating conditions. The proposed method was applied to four sub-datasets which have different working environments to verify the generality of the proposed model.

### 3.3.2. Data Preprocessing

Different sensors have different physical meanings and numerical ranges. In order to eliminate the influence of ranges of value on the degree of contribution, a standardized method is used to adjust the range of each sensor. The standardized rules are learned in the training set. Then, the sensors are selected based on the trends of the raw data.

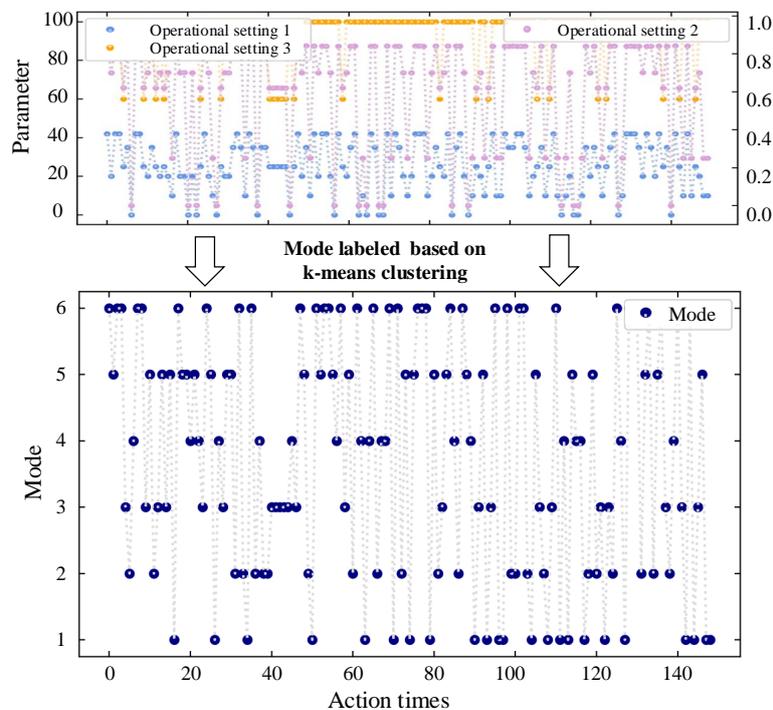
To the system working under the single operating condition, sensors can be divided into three categories, which are ascending, descending and unchanged, according to the trend of the data, as shown in Figure 5a. Data without obvious changes throughout the life time of the system do not contribute to predicting the remaining useful life of system. Thus, 13 kinds of sensors are selected as the raw data of the aging status, including sensor data with ascending trends (2, 3, 4, 8, 9, 11, 13, 15, 17)

and descending trends (7, 12, 20, 21) for FD001, and 11 kinds of sensors (the 3, 4, 8, 13, 15, 17 have an ascending trend and the 7, 11, 12, 20, 21 have a descending trend) are selected for FD003.



**Figure 5.** The trends of four sensors in different working conditions. (a) System working under a single mode in FD001. (b) System working under multi-modes in FD002.

For the system working under multiple operating conditions, the trend of sensor data depends mainly on the switching of operating modes, so the degradation trend does not show a noticeable monotonic trend. As shown in Figure 5b, the same sensors 11, 16, 20 and 21 show irregular trends in FD002 (working in multi-modal). Therefore, principal component analysis method is used to reconstruct features based on the principle of maximum difference. The threshold value of total contribution is set to 95% to filter cumulative dynamic difference features of irregular sensors. More details about principal component analysis may be found in [41]. The working mode of the sample in the dataset is represented by three operational settings. To determine the mode switching sequence, k-means clustering method is used to label six working modes. The results are as shown in Figure 6.



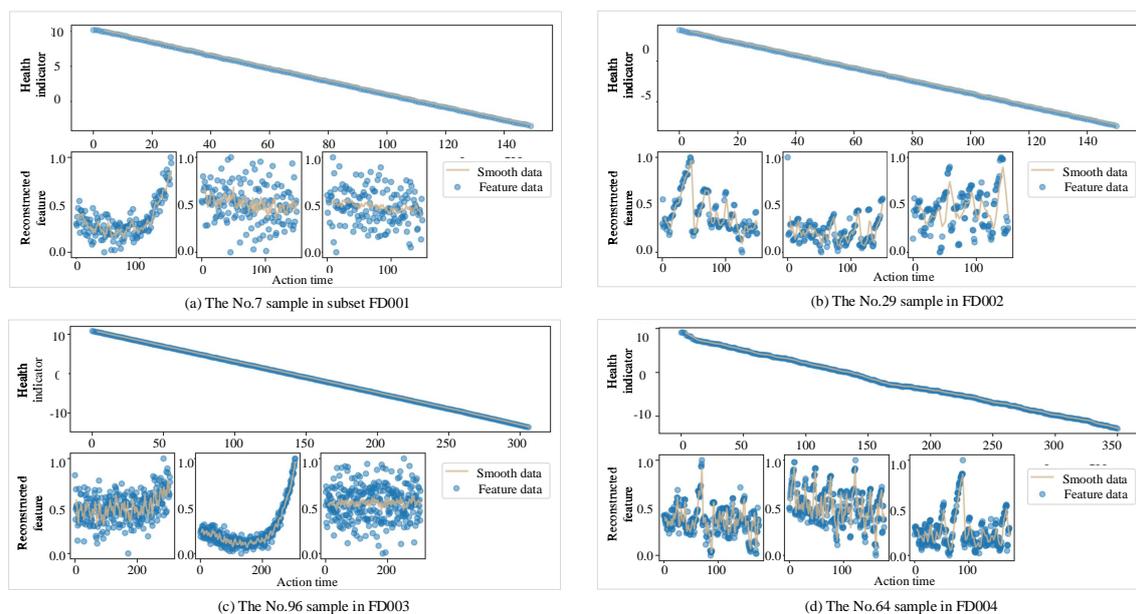
**Figure 6.** Mode labeled from three operational settings based on k-means clustering algorithm, where the data are from the first sample of FD002.

### 3.3.3. Discussion of Results

The detailed description of the proposed method is given in the Section 2. Here, the processed data features are input into the framework. Then, the long-term cumulative degradation prediction and the latest stage degradation prediction are performed separately and the results are fused. After that, the model performance is evaluated and analyzed according to the evaluation metrics.

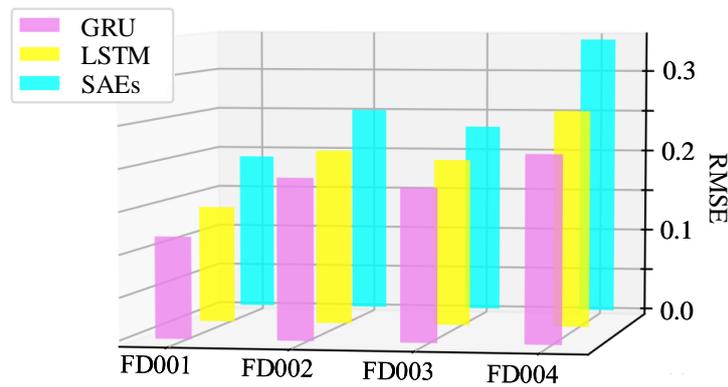
Since each sub-dataset has larger feature dimensions, considering that there may be a linear correlation between multi-dimensional features, the principal component analysis is used to reconstruct the data to reduce unnecessary calculations. Based that, FD001 retains 27 dimensional features, FD002 retains 14 dimensional features, FD003 retains 23 dimensional features and FD004 retains 14 dimensional features. Then, the multi-objective plan is used to establish health indicators based on monotonicity, threshold similarity and trend similarity.

Figure 7 illustrates the comparison of a health indicator and three representative reconstructed features of random samples from different sub-datasets. The points show the data and the solid line is the smoothing data based on moving average smooth method to show the trend of the data more clearly. The experimental results show that the proposed health indicator is more monotonic and better model fitting.



**Figure 7.** The comparison of health indicator and three representative reconstructed features of four samples from each sub dataset. (a) The comparison of the health indicator and three representative reconstructed features of the sample number 7 in subset FD001. (b) The same comparison with the sample number 29 in subset FD002. (c) The same comparison with the sample number 96 in subset FD003. (d) The same comparison with the sample number 64 in subset FD004. The trend of data is shown more clearly by smoothing data.

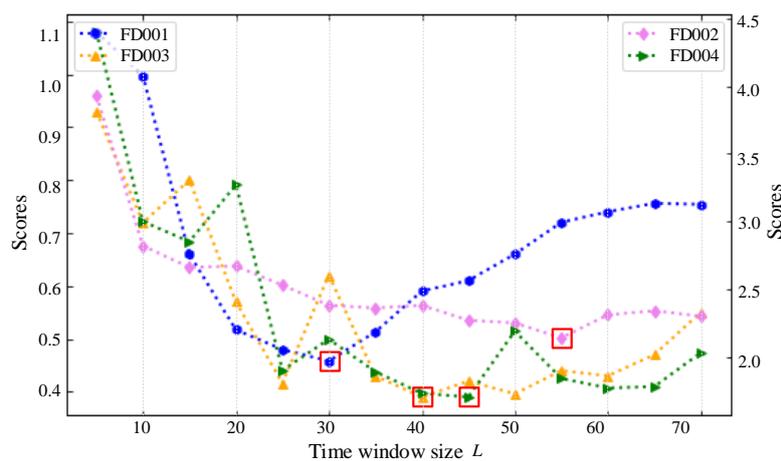
Then, the gated recurrent unit model is used to predict remaining useful life for long-term cumulative degradation. Long short term memory network (LSTM) and stacked autoencoders (SAEs) are used to compare predictions. The results are shown in Figure 8, which indicates the proposed CDDHI-GRU mode has a lower RMSE and score values than other algorithms. In addition, it was found that the values of evaluation metrics of FD002 and FD004 are higher than those of FD001 and FD003, indicating that the samples working in multi-modal switching are more challenging for obtaining accurate prediction results than those working in single mode.



**Figure 8.** The values of RMSE and scores of stacked autoencoders (SAEs), long short term memory network (LSTM) and GRU from four sub datasets are on the chart. The differences are plotted in the figure.

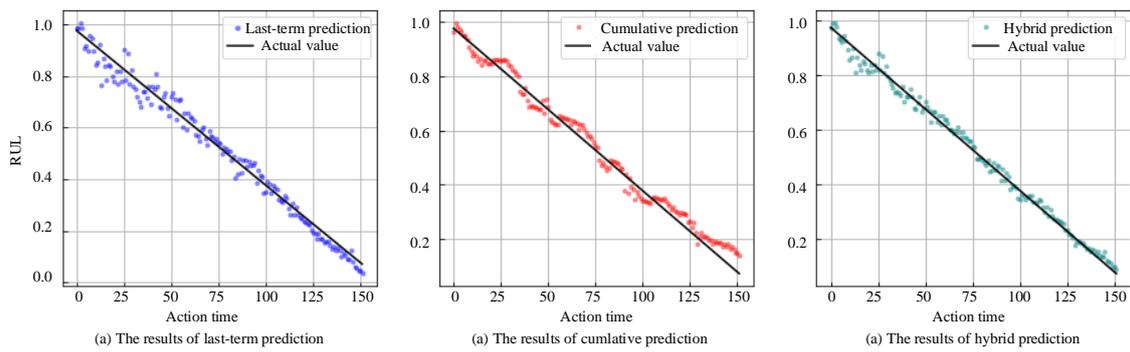
On the other hand, the time window is used to extract energy and statistical features. It is worth noting that during decomposing the data of the time window based on ensemble empirical mode decomposition, the number of IMF components  $N_{imf}$  is different for different local data. The reason is that the local data of the sample have various fluctuations, and there is noise in the data, so  $N_{imf}$  is added as an energy feature.

The size of latest stage sequence  $L$  is a key parameter affecting the performance of model. For the sample with periodic mode switching, a multiple of the period is selected as the time window length. For samples with irregular mode switching, the time window length is selected by traversing. In order to avoid late prediction as much as possible, we use the scores to choose the length of time window. Figure 9 shows the effects of  $L$  on prediction accuracy, and the sizes of the time windows for the sub-datasets are 30, 55, 40 and 45 respectively. The reason is that a larger time window can cover more information to get better results, but extending the window length too much could adversely affect near-term forecasts.



**Figure 9.** The scores of different time window sizes for four sub-datasets.

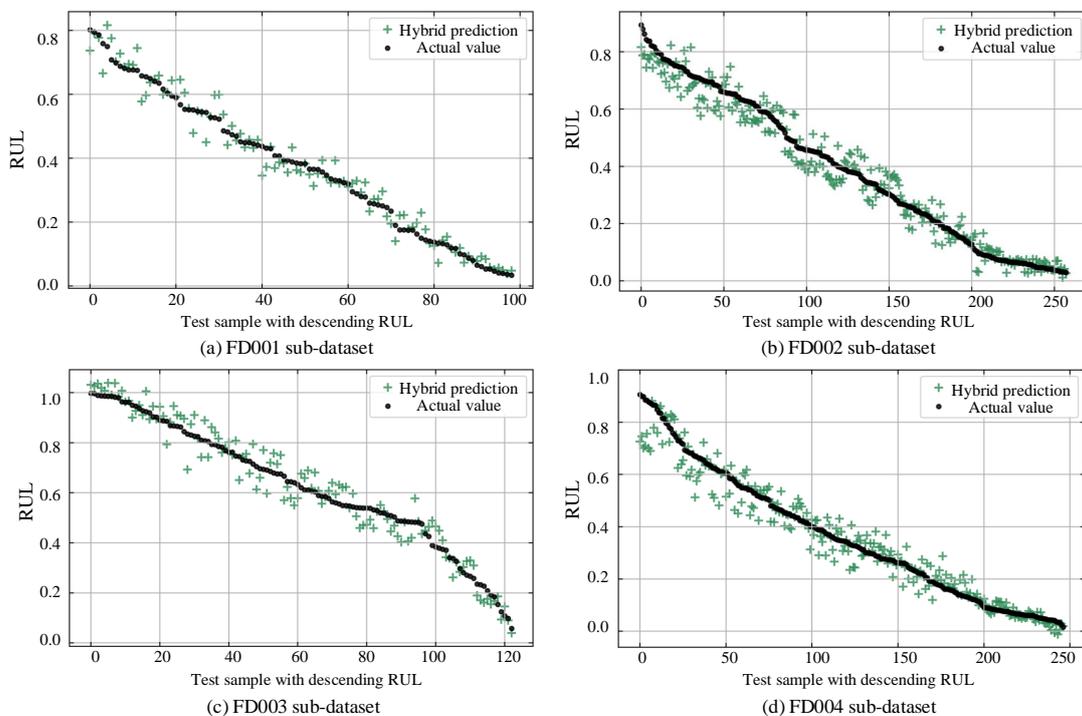
Finally, the model averaging is used to fuse the prediction results. After several experiments, the parameters of each sub-dataset are finally determined. Figure 10 illustrates the comparison of results for long-term cumulative degradation prediction, latest stage degradation prediction and fusion prediction for the same sample.



**Figure 10.** The results of latest stage degradation prediction, long-term cumulative degradation prediction and hybrid prediction for number 148 in FD002. (a) The comparison between latest stage cumulative degradation prediction and actual values. (b) The comparison between long-term degradation prediction and actual value. (c) The comparison between hybrid prediction and actual values.

The experimental results show that the long-term cumulative prediction model is more accurate in predicting the remaining useful life of the system at the early stage of aging, and the latest stage degradation prediction model is more accurate at predicting the system in the late stage of aging. The hybrid prediction model combines the advantages of the two and has better overall prediction performance. No matter whether the system is in an early or late stage of the fault, it has a stabler prediction effect on the life cycle of the system.

The estimated results based on the proposed model and the actual values of remaining useful life for the four sub-datasets’ test samples are compared in Figure 11. In order to observe the performance of the prediction model intuitively, the test samples are arranged in descending order of the actual values.



**Figure 11.** The prediction results of the hybrid model for four sorted sub-datasets (FD001, FD002, FD003, FD004) are shown in (a–d) respectively.

Figure 11 shows that the prediction results of the proposed model are close to the actual value of remaining useful life, and it is worth noting that the prediction ability of the model in the early stage of aging is weaker than that in the late stage of aging. The result is that the fault characteristics in the early stage of aging are not obvious, and with the aging of the system, the prediction accuracy is gradually enhanced. In addition, with the increase in the number of faults and the number of operational modes, it is more difficult to obtain an accurate prediction. However, the overall prediction performance of the model remains stable. When it is difficult to predict accurately, the prediction results tend to an early prediction as opposed to a late prediction.

The proposed method is compared with other popular methods in this dataset to verify the superiority of the proposed method. The recurrent neural network (RNN), the long-short term memory network (LSTM), stacked autoencoders (SAEs), convolutional neural networks (CNN), the gradient boosting decision tree (GBDT) and support vector regression (SVR) as representative methods are evaluated in Table 3.

**Table 3.** The evaluation metrics values of each dataset based on different models.

Methods	Metric	FD001	FD002	FD003	FD004
SVR	RMSE	0.147	0.241	0.205	0.272
	Score	1.139	3.874	1.757	4.461
CNN	RMSE	0.132	0.234	0.193	0.243
	Score	1.304	2.893	1.418	3.36
LSTM	RMSE	0.149	0.223	0.215	0.231
	Score	1.256	3.312	1.862	3.975
SAEs	RMSE	0.145	0.259	0.233	0.294
	Score	1.672	4.023	2.047	4.351
GBDT	RMSE	0.124	0.213	0.218	0.302
	Score	1.562	2.421	1.592	3.461
Proposed model	RMSE	<b>0.119</b>	<b>0.172</b>	<b>0.184</b>	<b>0.217</b>
	Score	<b>1.042</b>	<b>2.213</b>	<b>1.387</b>	<b>2.84</b>

The tabular data shows that the model proposed in this paper has better evaluation metric values and has been proven to perform well in both overall prediction and late prediction. Comparing the predictions between four sub-datasets, it was found that to accurately predict the remaining useful life of the system working in multiple modes is more challenging than that of the system working in single mode.

#### 4. Conclusions

In this paper, a hybrid degradation modeling and prognostic method is proposed for a multi-modal system. Firstly, the cumulative dynamic differential feature is constructed to describe the physical characteristics with the existence of mode switching. The constructed health indicator is based on the three properties, including monotonicity, threshold similarity, and trend similarity. Based on the cumulative dynamic differential health indicator, the long-term cumulative degradation assessment model is constructed by gated recurrent unit. Then, the time window is used to extract energy and statistical features to obtain local features. The latest degradation effect is predicted based on light gradient boosting machine. Finally, the model is averaged to construct a decision-level fusion model.

In the experimental analysis section, the theoretical experiment uses synthetic data to study the sensitivity of the proposed method to different types of input data. Most of systems show a slow aging trend in the early stage and a fast aging trend in the later stage. The synthetic data with exponential functional form are constructed according to based on exponential functional and Gaussian random variable. The experimental result shows the proposed method has good performances on samples with different trends. In addition, experimental analysis on the C-MAPSS turbofan aero-engine dataset,

which includes four different working environment sub-datasets, demonstrates the effectiveness and generality of the proposed method. Compared with other popular methods, the proposed model has shown very promising prognostic performance.

In the future, the proposed method will be extended in the following ways: 1) How can one reconstruct data samples while working with different mode switchings from historic data samples? The accuracy of prediction depends on the completeness of the dataset. An effective reconstruction sample model could supplement the dataset to improve the accuracy of prediction. 2) How can one construct the health indicators which have better prediction performance? This paper constructed a health indicator based on three properties, including monotonicity, threshold similarity and trend similarity. Exploring more properties of the health indicators and studying optimization solution of complex multi-objective programming are the challenges. 3) A complex system may have different kinds of failures during useful life. The type and number of failures have an impact on the aging process of the system. Thus, how to construct a prediction model considering the impact of both the type and number of failures on remaining useful life is a significant issue.

**Author Contributions:** conceptualization, J.P. and S.W.; methodology, S.W. and Y.C.; software, S.W.; validation, S.W.; formal analysis, W.Y.; investigation, X.Z.; resources, D.G.; data curation, B.C.; writing—original draft preparation, S.W.; writing—review and editing, Y.C.; visualization, Y.Y.; supervision, Z.H.; project administration, J.P.; funding acquisition, J.P. and S.W.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant numbers 61873353, 61672537, 61672539 and 61803394) and the Fundamental Research Funds for the Central Universities of Central South University (2018zzts540).

**Acknowledgments:** This research was funded by the National Natural Science Foundation of China (grant numbers 61873353, 61672537, 61672539 and 61803394) and the Fundamental Research Funds for the Central Universities of Central South University (2018zzts540).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

RUL	Remaining useful life
HI	Health indicator
CDD-HI	Cumulative dynamic differential health indicator
GBDT	Gradient boosting decision tree
LightGBM	Light gradient boosting machine
RNN	Recurrent neural network
GRU	Gated recurrent unit
DTW	Dynamic time warping
LSTM	Long-short term memory network
EEMD	Ensemble empirical mode decomposition
IMF	Intrinsic mode function
CNN	Convolutional neural networks
SVR	Support vector regression
SAEs	Stacked autoencoders

## References

1. Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Networks Learn. Syst.* **2016**, *28*, 2306–2318, doi:10.1109/TNNLS.2016.2582798.
2. Yang, D. Physics-of-failure-based prognostics and health management for electronic products. In Proceedings of the 2014 15th International Conference on Electronic Packaging Technology (ICEPT2014), Chengdu, China, 12–15 August 2014; pp. 1215–1218.
3. Huang, B.; Cohen, K.; Zhao, Q. Active anomaly detection in heterogeneous processes. In Proceedings of the 2018 International Conference on Acoustics, Speech, and Signal Processing (ICASSP2018), Calgary, AB, Canada, 15–20 April; pp. 3924–3928.

4. Li, N.; Lei, Y.; Guo, L.; Yan, T.; Lin, J. Remaining useful life prediction based on a general expression of stochastic process models. *IEEE Trans. Ind. Electron.* **2017**, *64*, 5709–5718, doi:10.1109/TIE.2017.2677334.
5. Kang, M.; Kim, J.; Kim, J.M. Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis. *IEEE Trans. Power Electron.* **2014**, *30*, 2786–2797, doi:10.1109/tpel.2014.2358494.
6. Xiao, L.; Chen, X.; Zhang, X.; Liu, M. A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition. *J. Intell. Manuf.* **2017**, *28*, 1893–1914, doi:10.1007/s10845-015-1077-x.
7. Glowacz, A.; Glowacz, Z. Diagnosis of stator faults of the single-phase induction motor using acoustic signals. *Appl. Acoust.* **2017**, *117*, 20–27, doi:10.1016/j.apacoust.2016.10.012.
8. Huang, B.; Cohen, K.; Zhao, Q. Active anomaly detection in heterogeneous processes. *IEEE Trans. Inf. Theory* **2018**, *65*, 2284–2301, doi:10.1109/TIT.2018.2866257.
9. Zio, E.; Di Maio, F. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliab. Eng. Syst. Saf.* **2010**, *95*, 49–57, doi:10.1016/j.res.2009.08.001.
10. Cheng, Y.; Peng, J.; Gu, X.; Zhang, X.; Liu, W.; Yang, Y.; Huang, Z. RLCP: A Reinforcement Learning Method for Health Stage Division Using Change Points. In Proceedings of the 2018 IEEE International Conference on Prognostics and Health Management (ICPHM2018), Seattle, WA, USA, 11–13 June 2018; pp. 1–6.
11. Yang, Y.; Xiong, L.; Liu, W.; Gao, K.; Huang, Z. An Energy-Based Nonlinear Pressure Observer for Fast and Precise Braking Force Control of the ECP Brake. *Int. J. Precis. Eng. Manuf.* **2018**, *19*, 1437–1445, doi:10.1007/s12541-018-0170-4.
12. Li H.; Zhang X.; Peng J.; He J.; Huang Z.; Wang J. Cooperative CC-CV charging of supercapacitors using multi-charger systems. *IEEE Trans. Ind. Electron.* **2020**, doi: 10.1109/TIE.2019.2962485
13. Wu, Y.; Yuan, M.; Dong, S.; Lin, L.; Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, *275*, 167–179, doi:10.1016/j.neucom.2017.05.063.
14. Khelif, R.; Chebel-Morello, B.; Malinowski, S.; Laajili, E.; Fnaiech, F.; Zerhouni, N. Direct remaining useful life estimation based on support vector regression. *IEEE Trans. Ind. Electron.* **2016**, *64*, 2276–2285, doi:10.1109/TIE.2016.2623260.
15. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long short-term memory network for remaining useful life estimation. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM 2017), Chiba, Japan, 13–17 April 2018; pp. 88–95.
16. Hu, C.; Youn, B. D.; Wang, P.; Yoon, J. T. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliab. Eng. Syst. Saf.* **2012**, *103*, 120–135, doi:10.1016/j.res.2012.03.008.
17. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834, doi:10.1016/j.ymsp.2017.11.016.
18. Liao, H.; Tian, Z. A framework for predicting the remaining useful life of a single unit under time-varying operating conditions. *Iie Trans.* **2013**, *45*, 964–980, doi:10.1080/0740817X.2012.705451.
19. Lei, Y.; Li, N.; Lin, J. A new method based on stochastic process models for machine remaining useful life prediction. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2671–2684, doi:10.1109/TIM.2016.2601004.
20. Li, H.; Wang, Y.; Wang, B.; Sun, J.; Li, Y. The application of a general mathematical morphological particle as a novel indicator for the performance degradation assessment of a bearing. *Mech. Syst. Signal Process.* **2017**, *82*, 490–502, doi:10.1016/j.ymsp.2016.05.038 .
21. Hu, J.; Tse, P. A relevance vector machine-based approach with application to oil sand pump prognostics. *Sensors* **2013**, *13*, 12633–12686, doi:10.3390/s130912663.
22. Benkedjouh, T.; Medjaher, K.; Zerhouni, N.; Rechak, S. Health assessment and life prediction of cutting tools based on support vector regression. *J. Intell. Manuf.* **2015**, *26*, 213–223, doi:10.1007/s10845-013-0774-6.
23. Lei, Y.; Li, N.; Gontarz, S.; Lin, J.; Radkowski, S.; Dybala, J. A model-based method for remaining useful life prediction of machinery. *IEEE Trans. Reliab.* **2016**, *65*, 1314–1326, doi:10.1109/TR.2016.2570568.
24. Guo, L.; Li, N.; Jia, F.; Lei, Y.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109, doi:10.1016/j.neucom.2017.02.045.
25. Zhang, Y.; Xiong, R.; He, H.; Liu, Z. A lstm-rnn method for the lithium-ion battery remaining useful life prediction. In Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin 2017), Harbin, China, 9–12 July 2017; pp. 1–4.

26. Song, Y.; Li, L.; Peng, Y.; Liu, D. Lithium-Ion Battery Remaining Useful Life Prediction Based on GRU-RNN. In Proceedings of the 2018 12th International Conference on Reliability, Maintainability, and Safety (ICRMS 2018), Shanghai, China, China, 17–19 October 2018; pp. 317–322.
27. Zheng, C.; Liu, W.; Chen, B.; Gao, D.; Cheng, Y. A data-driven approach for remaining useful life prediction of aircraft engines. In Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC 2018), Maui, HI, USA, 4–7 November 2018; pp. 184–189.
28. Wang, S.; Zhang, X.; Gao, D.; Chen, B.; Cheng, Y.; Yang, Y.; Peng, J. A Remaining Useful Life Prediction Model Based on Hybrid Long-Short Sequences for Engines. In Proceedings of the In 2018 21st International Conference on Intelligent Transportation Systems (ITSC 2018), Maui, HI, USA, 4–7 November 2018; pp. 1757–1762.
29. Singh, S. K.; Kumar, S.; Dwivedi, J.P. A novel soft computing method for engine RUL prediction. *Multimed. Tools Appl.* **2019**, *78*, 4065–4087, doi:10.1007/s11042-017-5204-x.
30. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 2016 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining (ACM 2016), Amsterdam, Holland, 15–19 October 2016; pp. 785–794.
31. Ke, G.; Meng, Q.; Finley, T. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 2017), Long Beach, USA, 3–9 December 2017; pp. 3146–3154.
32. Mosallam, A.; Medjaher, K.; Zerhouni, N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *J. Intell. Manuf.* **2016**, *27*(5), 1037–1048, doi:10.1007/s10845-014-0933-4.
33. Pan, Y.; Er, M.J.; Li, X.; Machine health condition prediction via online dynamic fuzzy neural networks. *Eng. Appl. Artif. Intell.* **2014**, *35*, 105–113, doi:10.1016/j.engappai.2014.05.015.
34. Liao, L.; Köttig, F.; Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans. Reliab.* **2014**, *63*, 191–207, doi:10.1109/TR.2014.2299152.
35. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks* **1994**, *5*, 157–166, doi:10.1109/72.279181.
36. Zhang, X.; Xiao, P.; Yang, Y. Remaining Useful Life Estimation Using CNN-XGB with Extended Time Window. *IEEE Access* **2019**, *7*, 15438–154397, doi:10.1109/ACCESS.2019.2942991.
37. Coble, J.B.; Hines, J.W. Prognostic algorithm categorization with PHM challenge application. In Proceedings of the 2008 International Conference on Prognostics and Health Management (ICPHM2008), Denver, CO, USA, 6–9 October 2008; pp. 1–11.
38. Saxena, A.; Goebel, K.; Simon, D. Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the 2008 International Conference on Prognostics and Health Management (ICPHM2008), Denver, CO, USA, 6–9 October 2008; pp. 1–9.
39. Frederick, D.K.; DeCastro, J.A.; Litt, J.S. *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)*; NASA Glenn Research Center: Cleveland, OH, USA, 2007.
40. Saxena, A.; Goebel, K. Turbofan Engine Degradation Simulation Data Set. NASA Ames Prognostics Data Repository. 2008. Available online: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-datarepository/> (accessed on 20 November 2018).
41. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52, doi:10.1016/0169-7439(87)80084-9.

