Di Wang, *Member, IEEE*, Xiaochen Xian, *Member, IEEE*, and Changyue Song, *Member, IEEE*

[[1]](#footnote-1)

Joint Learning of Prognostics and Failure Mode Recognition for Degradation Processes

*Abstract*—To avoid unexpected failures of units in manufacturing systems, failure mode recognition and prognostics are critically important in prognostics health management (PHM). Most existing methods either ignored the effects of various failure modes on RUL prediction or implemented failure mode recognition and RUL prediction in two separate parts, which obtain inaccurate RUL prediction results and lack information about failure modes of units as a result. To address the issue, this paper proposes a joint learning model of failure mode recognition and RUL prediction for degradation processes based on multiple sensor signals. As a deep neural network, the proposed joint learning model first extracts degradation features by considering engineering knowledge to ensure good interpretability, and then takes the extracted features as model inputs. By conducting failure mode recognition and RUL prediction as a collaborative way that can fully characterize complex relationship between RUL and failure modes, The proposed model outputs recognized failure modes and predicted RUL of units simultaneously. A case study on the degradation of aircraft gas turbine engines is presented to evaluate the proposed model performance.

*Note to Practitioners*—The paper aims to develop a joint learning method for failure mode recognition and RUL prediction of operating units based on multiple sensor signals. Specifically, the developed method addresses a challenging question in practice, i.e., how to effectively conduct failure mode recognition and RUL prediction as a joint task based on interpretable extracted degradation features that are related to engineering knowledge. To implement this method in practice, four steps are included as follows: *First*, collecting multiple sensor signals, failure time, and failure modes of historical units. *Second*, construct the joint learning model based on features extracted from sensor signals by considering engineering knowledge of degradation. *Third*, estimating model parameters using data of historical unit. *Fourth*, recognize the failure mode and predict the RUL of an in-service unit. The proposed method is expected to be widely applied to practical situations, especially for manufacturing systems with complex structures and unknown failure thresholds.

*Index Terms*— Joint learning, failure mode recognition, remaining useful lifetime prediction, feature extraction

# INTRODUCTION

U

NEXPECTED failures of manufacturing systems often lead to severe consequences including disruption of logistics, production downtime, threatening equipment safety, and customer dissatisfaction. As a result, prognostics health management (PHM) of manufacturing systems is critically important to avoid unexpected failures [1], which has played significant roles in various industrial use cases including aircraft engines [2], batteries [3], and semiconductors [4].

With the development of sensing technology, multiple types of sensor signals are collected for PHM of manufacturing systems, which provide abundant information about degradation characteristics of operating units in manufacturing systems, including correlations among different types of sensors, multiple potential failure modes, possible degradation mechanisms, and more. By fully considering the degradation characteristics of operating units, failure mode recognition and remaining useful lifetime (RUL) prediction of the operating units based on collected multiple sensor signals has become a significant issue in PHM [5]. Specifically, failure mode recognition and RUL prediction is to recognize the specific failure mode that causes the operating unit to degrade among multiple failure modes and predict the RUL of the operating unit under multiple failure modes as well.

Various methods for RUL prediction, failure mode recognition, as well as RUL prediction- and failure mode recognition-joint task have been proposed in the literature. For RUL prediction, data-driven methods are widely applied, which can be broadly divided into two categories, namely, statistical and machine learning methods. Statistical fusion methods for prognostics are commonly used including decision-level fusion, feature-level fusion and data-level fusion based on different fusion levels [6]. In particular, health index (HI)-based data-level fusion methods are highly desired in practice by constructing a composite HI to characterize the degradation status of a unit from multiple types of sensor data for better prognostics [7, 8]. However, most of existing statistical fusion methods make prognostics by assuming only one potential failure mode exists and ignoring the differences between multiple failure modes. In many real-world applications, multiple failure modes may have distinct influences on an operating unit, leading to different degradation patterns of the unit. Therefore, accurate RUL prediction under multiple failure modes has been a challenge for existing statistical fusion methods. Machine learning methods for RUL prediction have attracted great attention especially for complex systems in recent years due to its simplicity, flexibility, and general applicability [9]. Tian et al. [10] applied a feed-forward neural networks (FFNNs) by taking the cycle time and available degradation signals as the inputs and the normalized life percentage as the output. However, FFNNs cannot fully characterize time evolutive features of degradation signals. To address this issue, Zhu et al. [11] adopted a convolutional neural network (CNN) that extracted 2D features from vibration signals to predict RULs of bearings. Recurrent neural network (RNN) in [12] and its variants including long short-term memory (LSTM) in [13] extracted features from time sequences of sensor signals and environmental conditions to predict RUL. Other models like restricted Boltzmann machine (RBM) [14] and deep belief networks (DBN) [15] have also been employed to predict RUL. Existing machine learning-based RUL prediction methods behave like black boxes to extract features from sensor signals directly without consideration of engineering knowledge, leading to poor interpretability of the extracted features and huge computational time for feature extraction from original sensor signals.

For failure mode recognition, some studies have been proposed including physical-based models [16], hidden markov models [17] and Gaussian process models [18], but these studies can only analyze a single sensor signal. For studies based on multiple sensor signals, deterministic methods such as proportional hazard model [19] and minimum cut set model [20], as well as probabilistic methods such as response surface model [21] explored multi-failure response traits, but correlations among different types of sensors are ignored. Various machine learning models such as K-nearest neighbor (KNN) [23], decision tree (DT) [24], random forest (RF) [25], support vector machine (SVM) [22], logistic regression (LR), Bayes classification (BC), and artificial neural network (ANN) [26], took multiple failure mode recognition as a multiclassification problem, but these methods cannot make prognostics simultaneously.

For RUL prediction- and failure mode recognition-joint task, statistical methods have been applied. For example, Bichon et al. proposed Gaussian process surrogate models for the efficient estimation of the reliability of a system with multiple failure modes [27]. This method is based on a single type of sensor signal, but cannot fuse multiple sensor signals for a more accurate prognostic result under multiple failure modes. The same problem also exists in [28]. Chehade et al. [5] proposed a data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes. The method recognizes failure modes and predict the RUL in two separate parts, reducing the efficiency and accuracy of the method. Liu et al. developed an optimum condition-based maintenance policy for continuously monitored degrading systems with multiple failure modes by assuming a pre-specified failure threshold [29]. However, the failure threshold may not be straightforward to determine in practice, because units of different failure modes may require different failure thresholds. Deep learning methods for solving prognostics problems with multiple failure modes can be broadly categorized into two types: One type of methods takes multiple sensor signals and the failure mode of a unit together as the input and outputs the RUL as the prognostic result, such as Bayesian deep learning framework in [30]. It is difficult to understand how each failure mode affects the degradation of units using these methods. The other type of methods separate failure mode recognition and RUL prediction into two parts, in which the failure mode recognition is a supervised classification problem and the RUL prediction is another supervised prediction problem [31]. To the best of our knowledge, few studies have made collaborative strategies to recognize multiple failure modes and predict the RUL simultaneously.

In conclusion, the challenges and research gaps of failure mode recognition and RUL prediction are as follows: *First*, degradation patterns of units under multiple failure modes are complex, and may require different failure thresholds with different failure modes, which is difficult to determine in practice. Some of related existing methods make prognostics by assuming there is only one potential failure mode with one prespecified failure mode, failing to determining multiple failure thresholds. *Second*, sensor signals contain abundant interpretable degradation features that are related to engineering knowledge, and different types of sensors have correlations among each other. Existing deep learning methods extract features in “black boxes” by training massive raw sensor signals, which results in poorly interpretable extracted features and huge model computational time. *Third*, failure mode recognition and RUL prediction are integrated and collaborative, and thus should regarded as a joint task when conducting PHM. Unfortunately, most of existing methods handle these two issues separately and hierarchically.

In this paper, we develop a joint learning model of failure mode recognition and RUL prediction for degradation processes based on multiple sensor signals. Our proposed method focuses on operating units with multiple failure modes and aims to recognize the specific failure mode and predict the RUL of an in-service unit. The proposed joint learning model is a deep neural network, in which degradation features of an in-service unit extracted from multiple sensor signals by considering engineering knowledge are taken as the model inputs and the model outputs are the probabilities of all failure modes and the RUL distribution based on multiple failure modes of the in-service unit simultaneously. In particular, due to the limited available sensor signals of in-service units, we enhance the efficiency of their degradation features by learning common information from historical units and incorporate the information into the degradation features of the in-service units. The proposed method is expected to be suitable to degradation processes under different failure modes and thus leads to superior failure mode recognition and RUL prediction results.

The contributions of the proposed method are as follows: *First*, we capture degradation patterns under multiple failure modes in the deep neural network and directly predict the distribution of RUL, which does not need to determine different failure thresholds and widens the application of the proposed method to more practical situations where failure thresholds are unknown. *Second*, we extract degradation features engineering knowledge to ensure good interpretability of the extracted features and characterize correlations among multiple sensors by the deep neural network, which ensures model accuracy and largely reduce computational time as well. *Third*, the proposed deep neural network is a joint learning model that conducts failure mode recognition and RUL prediction as a collaborative way, which can capture complex relationship between RUL and failure modes ingeniously.

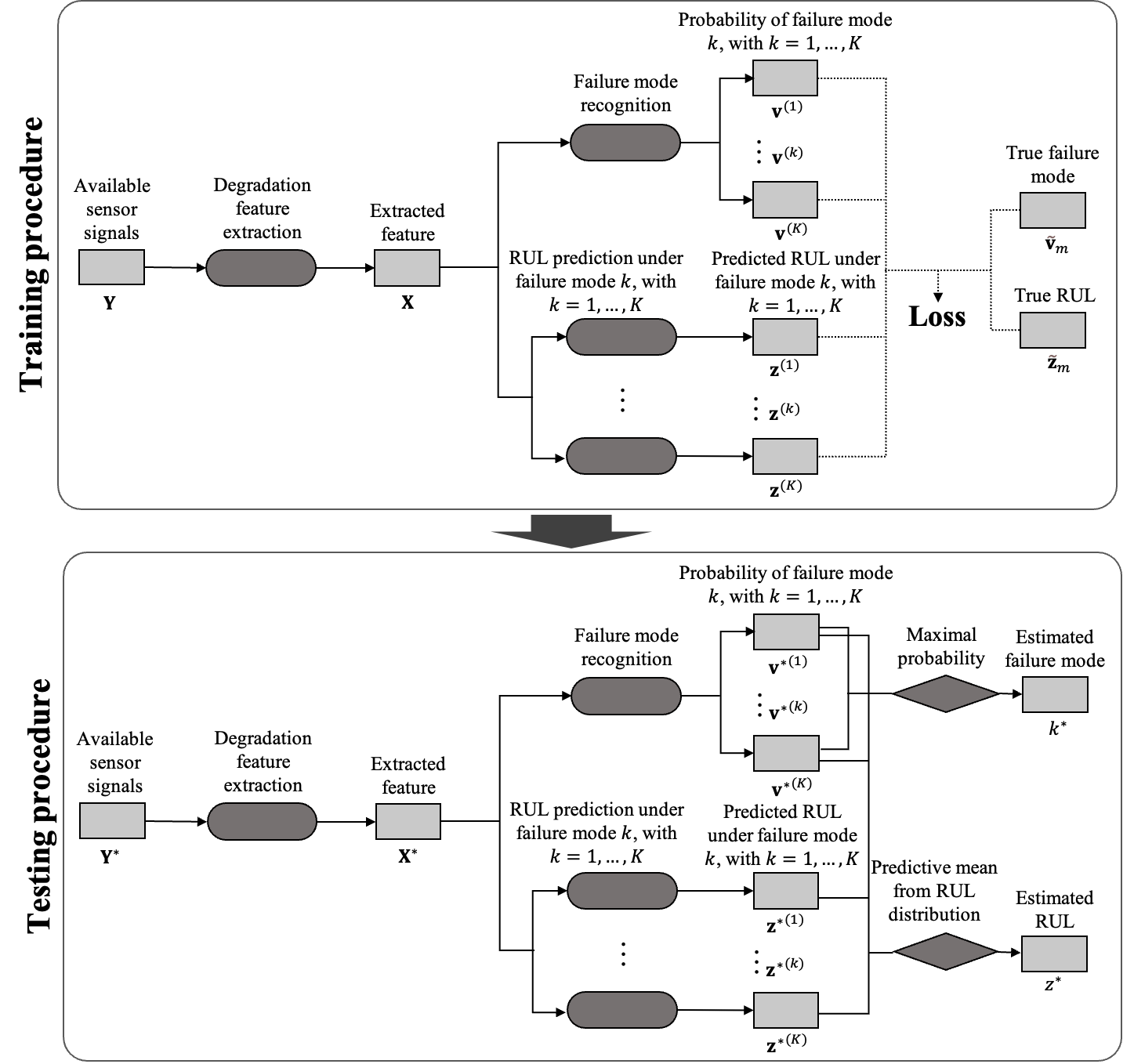


Fig. 1. The framework of the proposed method

The remainder of this paper is organized as follows. Section II introduces the proposed joint learning method that can recognize the failure mode and predict the RUL simultaneously in a complete deep learning framework. Sections III presents a case study of aircraft gas turbine engines to evaluate the proposed method. Section IV concludes this paper and outlines future directions.

# Methodology

There are two types of units, i.e., historical units, in which sensor signals are collected until failure, and in-service units, in which sensor signals are available up to some time points before failure. The historical units have a large number of available sensor signals, and thus show clear degradation trends and small uncertainties. Whereas for the in-service units, the number of available sensor signals rely on their degradation stages. For example, if an in-service unit is in the early degradation stage, the collected available sensor signals will be limited, leading to unclear degradation trend and large uncertainty. In this paper, we aim to recognize the failure mode and predict the RUL of in-service units under multiple failure modes given corresponding available sensor signals up to some time points before failure.

Fig. 1 presents the framework of the proposed method. We illustrate the proposed framework from two aspects, i.e., the offline model training procedure and the online model testing procedure, respectively. In the model training procedure, we first extract features of in-service units for each type of available sensor signals by considering engineering knowledge of degradation. Then, the extracted features of in-service units are taken as the model inputs of the proposed deep neural network, and the deep neural network jointly learns the probabilities of multiple failure modes and the estimated RUL under these failure modes as model outputs. The true RUL value and true failure mode of in-service units are provided as the labels of model inputs to construct the loss function for model training. In the model testing procedure, the purpose is to predict the RUL and failure mode of a new in-service unit. We first obtain the extracted features of the in-service unit and take them as the inputs of the proposed deep neural network to predict the probabilities of multiple failure modes and RUL distribution under these failure modes. Finally, we achieve the failure mode with the highest probability as the point estimator of the failure mode and the mean value of RUL distribution as the point estimator of RUL.

In Section II-A, we provide the details of feature extraction procedure for each type of sensor signals of in-service units. Section II-B introduces the offline training procedure for the proposed deep neural network. Section II-C provides failure mode recognition and RUL prediction for a new in-service unit by implementing the online model testing procedure.

## Feature Extraction by Considering Engineering Knowledge of Degradation

The first step is to extract features of in-service units for each type of available sensor signals to capture the progression of sensor signals by considering engineering knowledge of degradation. Since the available sensor signals of in-service units can be limited, their extracted features may be unreliable. We enhance the extracted features of in-service units by learning population-level information from the extracted features of historical units as well.

The sensor signals of a unit are available, which we denote as . Specifically, for unit , we decompose each type sensor signal , with , as a general mixed-effect degradation model given by Lu and Meeker [32] that is commonly applied in engineering practice, i.e.,

, (1)

where is the sensor signal time , and is the corresponding noise term that is assumed to follow the Gaussian distribution with mean zero and variance . The degradation model is a polynomial form with the basis function at time and the vector of the random effect parameters for the -th sensor signal of unit . is estimated by least squares method, i.e., , when we write (1) as a matrix form in which , , and with the last observed time . We consider the zero-order, first-order, and second-order derivatives of the degradation status at the last observed time , i.e., to characterize the progression of each sensor signal. Then, the original extracted features for the -th sensor signal are designed as

(2)

where , with , is standardized value of by decimal scaling.

For an in-service unit , its original extracted features for the -th sensor signal can be calculated by (1) and (2), with with . Due to the limited available sensor signals of in-service units, their original extracted features may be unreliable. To address this issue, we enhance the extracted features of in-service units by learning population-level information from historical units. We denote as the set of known historical units, in which sensor signals are available until failure. The original extracted features for the -th sensor signal of unit , are obtained as by (1) and (2), with and . Recall that the features of units are extracted from the mixed-effect degradation model. Mixed-effect models can incorporate population-level information by assuming a prior distribution of the random-effect parameter, i.e., , which can be estimated on the basis of historical units. If available sensor signals of in-service unit are provided, the posterior distribution of the random-effect parameter can be updated as , where the prior probability distribution is given as . We denote as the random variable of the feature for the -th sensor signal of in-service unit , where can be regarded as a deterministic mapping of , i.e., . The enhanced feature in-service unit considering population-level information from historical units is calculated by the expectation of , i.e.,

(3)

In (3), does not have a well-defined distribution because of multiple failure modes, and thus may not have an analytical expression. Instead, we can approximate by the corresponding sample average using a numerical way. Particularly, the random-effect parameters of historical units can be regarded as random samples from , and thus we can approximate based on historical units as follows

(4)

Therefore, enhanced feature is approximated by

(5)

where

(6)

In (6), is a constant that ensures , and is calculated by considering from the noise term , where and the noise variance is estimated by

(7)

If the historical unit has a similar degradation trend as the in-service unit , the likelihood ) will be large, leading to higher weight of unit .

Finally, the extracted feature of in-service unit is denoted as .

## Offline Training: Joint Learning by the Deep Neural Network

The proposed deep neural network aims to recognize the failure mode and predict the RUL of an in-service unit with available sensor signals up to some time point, which is a regression- and classification-integrated problem. We assume there are totally failure modes, and denote as the index of the -th failure mode. In the deep neural network, the extracted feature of an in-service unit is taken as the model input variable. Specifically, for in-service unit , the model input is . The formulation of the deep neural network is as follows:

(9)

(10)

. (11)

The deep neural network first learns common information for both failure mode recognition and RUL prediction by , where denote the parameter set for common information learning. Then, failure mode recognition module and RUL prediction module are constructed simultaneously. The failure mode recognition module outputs the vector of the probabilities of each failure mode by , where denotes the parameter set with respect to failure mode recognition. Here, , where is the probability of failure mode and we write to facilitate understanding. The RUL prediction module include submodules that predict RUL for each failure mode respectively. i.e., , where is the parameter set for RUL prediction under failure mode .

We introduce the equations of and , in detail. The formulation of the is

(12)

where is the value of hidden layer , with , , and ; and are the weight matrix and bias matrix at hidden layer , respectively; is the relu activation function that is commonly used in practice. We denote the parameter set as to simplify the notation. Based on , the formulation of is

(13)

In (13), is the value of hidden layer with , , and and are the weight matrix and bias matrix at hidden layer , respectively. at the output layer, and the probabilities of failure modes are obtained by a softmax function, where and are the weight matrix and bias of the output layer respectively, and represents the summation of all the elements of . We denote the parameter set as to simplify the notation. In addition, based on , the formulation of , with , is

(14)

where is the value of hidden layer , with , , and are the weight matrix and bias matrix at hidden layer , and and are the weight matrix and bias of the output layer. We denote the parameter set as to simplify the notation.

We denote as the set of in-service units for model training. For the unit , is the extracted feature obtained from its sensor signals , and is taken as the model input. We have the true failure mode value and true RUL value as the labels of model inputs. We denote the training data set as and the model parameter set as. The likelihood function for all training data is as follows

(15)

In (15), without loss of generality, we assume , with the noise term for . Then, the negative log-likelihood function is obtained as

(16)

where the constant is omitted.

We model the distribution of RUL from weight uncertainties by looking for the posterior distribution over using Bayesian theorem . In deep neural networks, the posterior distribution is generally not tractable. To address this issue, variational inference is adopted by defining an approximate variational distribution that is computationally tractable. The model training procedure is to minimize the Kullback–Leibler (KL) divergence between and i.e.,

(17)

where we only consider uncertainties of weights in , with . Specifically, at the output layer of , we assume the approximate variational distributions and as Gaussian distributions, i.e., and . We take (17) as the loss function, and adopt the backpropagation algorithm [33] to estimate model parameters.

## Online Testing: Failure Mode Recognition and RUL Prediction

For an in-service unit with available sensor signals , our goal is to recognize its failure mode and predict its RUL . Its feature are first extracted by the procedure in Section II-A to obtain the model input of the proposed deep neural network. Then, the proposed deep neural network jointly learns the failure mode and RUL. For failure mode recognition, the proposed deep neural network outputs the probabilities of multiple failure modes as , with . We achieve the failure mode with the highest probability as the point estimator of the failure mode as follows:

(18)

For RUL prediction, the proposed deep neural network outputs the RUL distributions under multiple failure modes. We denote the approximate predictive distribution of RUL as . The predictive mean value of RUL can be approximated by the sample mean as

(19)

with random samples through stochastic forward passes , where is sampled from , for . We achieve the point estimator of RUL as .

# Case Study

We conduct a case study on the degradation dataset of aircraft gas turbine engines to evaluate model performance.

## Overview of Dataset

The dataset is generated on Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) developed by NASA, which simulates the degradation processes of turbofan engines as shown in Fig. 3 [34]. 21 sensor signals of aircraft engines in their service life cycles are available to monitor the degradation status of the aircraft engines. The dataset contains four sub-datasets, and we use the sub-dataset named FD003 for case study. In FD003, each operating unit is subject to one of two potential failure modes: failure due to the high-pressure compressor (HPC) or failure due to the engine’s fan. FD003 have one training set and one testing set with 100 units respectively. In the training set, the sensor signals of each unit are collected until failure and the true failure time data are recorded. Failure modes of the units are provided by [5]. In the testing set, the sensor signals of each unit are available up to some time points before failure and the true RUL data are recorded. Failure modes of the units are unknown.

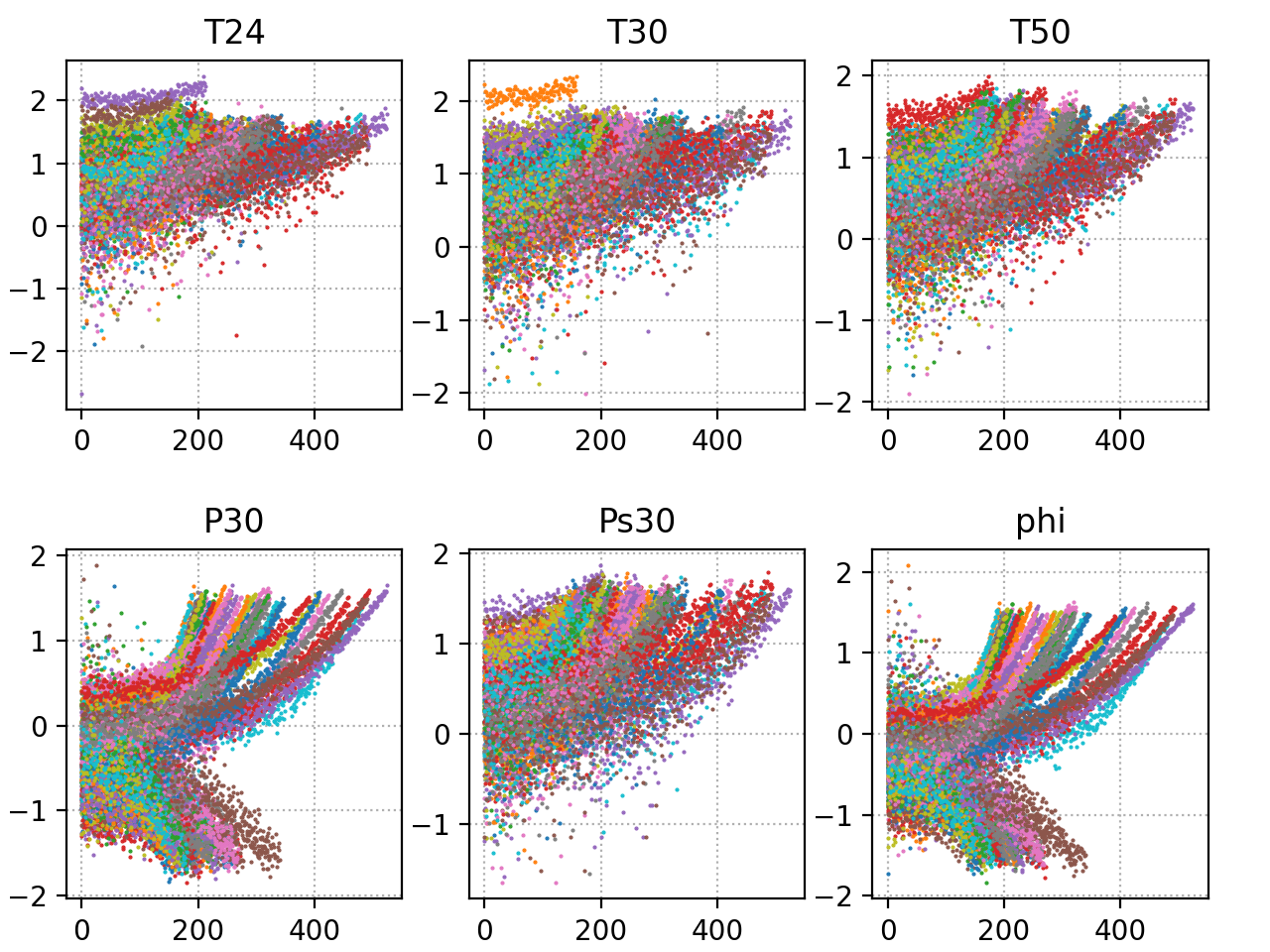


Fig. 3. An illustration of the selected sensor signals of the units in the training set.

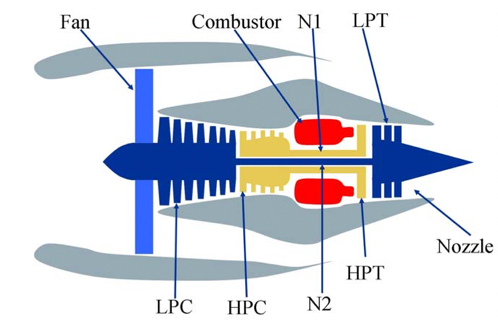


Fig. 2. Simplified diagram of engine simulated by C-MAPSS.

TABLE I

Detailed information of the selected sensors in the case study.

|  |  |  |
| --- | --- | --- |
| Symbol | Description | Units |
| T24 | Total temperature at LPC outlet | °R |
| T30 | Total temperature at HPC outlet | °R |
| T50 | Total temperature at LPT outlet | °R |
| P30 | Total pressure at HPC outlet | psia |
| Ps30 | Static pressure at HPC outlet | psia |
| phi | Ratio of fuel flow to Ps30 | pps/psi |

We consider the following preprocessing procedures for the collected sensor signals. Nine candidate sensors out of 21 sensors, namely, T24, T30, T50, P30, Ps30 and Phi, are selected on the basis of the criteria that the sensor signals should show a consistent increasing (or decreasing) trend under the corresponding failure modes in all units of the training set that have failed. The description of the selected sensors is shown in Table I. Without loss of generality, all collected sensor signals of units in both the training set and the testing set are preprocessed through z-score normalization and logarithm transformation [1]. Fig. 3 presents each selected sensor signal of the units in the training set. This figure displays sensor signals of units that fail due to two different failure modes. It clearly shows that the sensor signals tend to be highly correlated and that some sensors are sensitive to certain types of failure modes (i.e., Ps30, Phi, BPR, W31 and W32) while others (i.e., T24, T30, T50) are not. This motivates the use of deep neural network that fuse multiple and dependent sensor information to better infer the failure mode of a unit.

## Evaluation Metrics

We set three metrics to evaluate model performance.

* For failure mode recognition of in-service units, we use *accuracy* as the metric to measure the efficiency of failure mode recognition. *Accuracy* is defined as the ratio of units whose failure mode is correctly recognized.
* For RUL prediction of in-service units, we consider *mean absolute error* (MAE) and the *RUL prediction error* as the metric to evaluate prognostic performance. Specifically, the *RUL prediction error* of a new in-service unit is defined as the absolute difference between the predicted RUL and the true RUL divided by the true failure time as follows

. (20)

## Degradation Feature Extraction

To illustrate the effectiveness of the feature extraction model, we present extracted features of two randomly selected units in Fig. 3. For each unit, extracted features of all sensor signals are shown. X-axes represent the three extracted features respectively, and y-axes represent RUL. In these sub-figures, for an in-service unit with the last observed time , the blue points are original features of historical units at . The grain triangle points are original features of the in-service unit, and the red star points are enhanced features of the in-service unit. Fig. 3(a) shows an in-service unit at late degradation stage with available time 133 and RUL 50. Since the in-service unit at late degradation stage has a clear degradation trend, the original features directly extracted from sensor signals of the in-service unit are reliable, and close to the enhanced features learnt from both the in-service unit and the historical units. The extracted features of all sensor signals are related to RUL, and sensor signals are correlated with each other, which indicates that establishing a deep neural network is necessary to characterize the correlations among sensor signals and the relationship between the extracted features and RUL. Fig. 3(b) shows an in-service unit at early degradation stage with available time 56 and RUL 135. We can see that the original features directly extracted from sensor signals of the in-service unit are far from those of the historical units. The original features are unreliable given limited number of available sensor signals. Instead, the enhanced features learnt from both the in-service unit and the historical units are within the same region of the historical units and thus are reliable. Therefore, we consider enhanced features as the extracted features in the proposed method.

## Hyperparameter Selection

|  |
| --- |
|  |
| (a) |
|  |
| (b) |

Fig. 3. Extracted features of two randomly selected in-service units. (a) An in-service unit at late degradation stage with available time 133 and RUL 22, and (b) an in-service unit at early degradation stage with available time 56 and RUL 137. (Note: the blue points are historical units. The grain triangle points are original features of in-service units, and the red star points are enhanced features of in-service units.)

The structure of the proposed neural network mainly depends on two hyperparameters, i.e., the number of hidden layers and the number of hidden neurons at each hidden layer. We investigate the effects of these two hyperparameters by five-fold cross validation based on data under a randomly selected failure mode, where the model training process is reduced to learn a conventional neural network that takes the extracted feature of an in-service unit as model input and the corresponding RUL as model output. Figure 4 presents the model performance for RUL prediction by setting various hidden layers and hidden neurons based on data under failure mode ‘HPC’. As the number of hidden neurons increases, the MAEs for model training and validation both become smaller. Especially, when the number of hidden neurons is large, MAEs for model training are smaller than the ones for model validation, indicating the overfitting of the proposed model. In addition, MAEs become smaller as the number of hidden layers increases. We find that the model with the structure of four hidden layers and 15 hidden neurons have the comparable MAEs for model training and validation, and thus the number of hidden layers is set as four and the number of hidden neurons at each hidden layer is set as 15 for RUL prediction in our proposed model.

After hidden layers and hidden neurons are determined, we further investigate the effect of the number of common layers that learns common information for both failure mode recognition and RUL prediction by . We implement five-fold cross validation based on data under the other failure mode except the selected mode, where the model training process is also reduced to learn the conventional neural network. Figure 5 presents the model performance for RUL prediction by setting various numbers of common layers based on data under failure mode ‘fan’. We find that the model with one common layer has the smallest MAE for RUL prediction, and thus the number of common layers is set as one. Similarly, the number of hidden layers is set as two and the number of hidden neurons at each hidden layer is set as 5 for failure mode recognition in our proposed model. For other hyperprior parameters, it is worth noting that different values of the learning rate and mini-batch size are used and we find small changes in these parameters do not have a significant effect on RUL prediction performance. We set the learning rate by default in Keras and the mini-batch size as 50.

TABLE II

Results of proposed method and benchmark methods of type (1).

|  |  |  |
| --- | --- | --- |
| Method | Accuracy of failure mode recognition | RUL prediction error |
| KNN+ANN | 0.85 | 0.1521 |
| DT+ANN | 0.93 | 0.1521 |
| SVM+ANN | 0.92 | 0.1521 |
| LR+ANN | 0.88 | 0.1521 |
| BC+ANN | 0.65 | 0.1521 |
| ANN+ANN | 0.94 | 0.1521 |
| Proposed | 0.97 | 0.0667 |

Note: the method for failure mode recognition is on the left of “+”, and the method for RUL prediction is on the right.

TABLE III

Training time of the proposed method and the benchmark method of type (2).

|  |  |  |
| --- | --- | --- |
| Method | Without feature extraction | Proposed |
| Training time /s | 720 | 38 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| (a) | (b) | (c) | (d) |

Fig. 4. Hyperparameter selection for failure mode 1. (Note: the number of hidden layers are (a) 1, (b) 2, (c) 3, and (d) 4.)



Fig. 5. Hyperparameter selection for failure mode 2.

## Results and Comparison

To thoroughly evaluate the proposed method, we compare the proposed method with three types of benchmark methods: (1) conventional methods that recognize failure modes and predict RUL separately. The methods for failure mode recognition include KNN [22], DT [23], SVM [24], LR, BC and ANN [25], which take failure mode recognition as a multi-classification problem. Due to the nonlinear relationship between sensor signals and RUL of units, we take ANN as a benchmark for RUL prediction. (2) To verify the effectiveness of feature extraction by considering engineering knowledge, we set a benchmark method that has similar structure to the proposed method but disregards the feature extraction procedure. (3) To show the superiority of the proposed method that predicts RUL by fully considering the effects of various failure modes, we set another benchmark method that implements feature extraction procedure and then predicts RUL directly without considering the failure modes of units.

Table II lists the results of the proposed method and benchmark methods of type (1), including mean value of the accuracies of in-service units to recognize failure modes and mean value of RUL prediction errors. For failure mode recognition, although DT, SVM and ANN achieve good performance, the proposed method performs best. The RUL of the benchmark methods are predicted by employing ANN, and the prediction errors are the same since the procedures of failure mode recognition and RUL prediction are implemented separately and have no mutual influence. Compared with ANN, the proposed method predicts RUL more accurately, which indicates the effectiveness of joint learning by conducting failure mode recognition and RUL prediction collaboratively to capture complex relationship between RUL and failure modes ingeniously.

Figure 6 presents the results of failure mode recognition and RUL prediction under different levels of actual RUL using the proposed method and the benchmark method of type (2). For the levels of actual RUL, levels “25, 50, 75, 100, 125” refer to the in-service units that have actual RUL less than or equal to 25, 50, 75, 100 and 125, and level “” refers to all in-service units. In Figure 6(a), we provide the accuracy for failure mode recognition under each level of the actual RUL. When the actual RUL is less than 100, in other words, the observed time points of in-service units are relatively sufficient, the proposed and benchmark methods can both recognize failure modes correctly. Whereas, when the actual RUL becomes large, the proposed method has more accuracy than the benchmark method, indicating that feature extraction by considering engineering knowledge of degradation in the proposed method is effective to better recognize failure modes. In Figure 6(b), the bar and error bar in each level of the actual RUL represent the mean value and standard error of the RUL prediction errors, respectively. Here, the standard error is the one standard deviation of the RUL prediction error divided by the square root of the number of units. Compared with the benchmark method, the proposed method consistently provides much lower prediction errors for RUL. It can be seen that the superiority of the proposed method is more significant in RUL prediction by implementing the creative feature extraction procedure. In addition, Table III presents the training time of the proposed and benchmark methods. It is shown that training time of the proposed method is much shorter than the benchmark method. This is because that the benchmark method implements feature extraction in an NN structure that behaves aimlessly and must rely on a large number of training data, and thus takes much computational time. Whereas, the proposed method considers engineering knowledge of degradation to extract features more accurately and largely reduce computational time for model training as well.

|  |  |
| --- | --- |
|  |  |
| (1) | (2) |

Fig. 6. Comparison of accuracies for failure mode recognition and RUL prediction errors under different levels of actual RUL using the proposed method and the benchmark method of type (2).

To validate the performance of the proposed method that predicts RUL by the estimated mean value of the joint RUL distribution under multiple failure modes, Figure 7 illustrates the estimated mean and uncertainties on the RULs of in-service units using the proposed method and the benchmark method of type (3). The benchmark method adopts a BNN to achieve RUL distribution directly without considering the effects of failure modes, and predicts RUL by the estimated mean value of the RUL distribution. In Figure 7, The “\*” markers show the mean plus/minus three standard deviation of RULs estimated by the benchmark method. The “+” markers show the mean plus/ minus three standard deviation of RULs estimated by the proposed method. We can see that the proposed method gives RUL predictions closer to the true values than the benchmark, especially when the actual RUL is small (less than 50). In addition, it is shown from standard deviation of RULs that the prediction uncertainties of the proposed method are well-controlled and overall increase as the predicted values of RULs increase. Overall, the mean values of RUL prediction errors for all in-service units using the proposed method and the benchmark of type (3) are 0.0667 and 0.1025, respectively.

# Conclusion

Accurate failure mode recognition and prognostics are critically essential to avoid unexpected failures in manufacturing systems. Existing methods ignored the effects of various failure modes on RUL prediction or implemented failure mode recognition and RUL prediction in two separate parts, leading to the lack of information about failure modes and inaccurate RUL prediction results of units. In this paper, we develop a joint learning model of failure mode recognition and RUL prediction for operating units with multiple failure modes. As a deep neural network, the proposed joint learning model first extracts degradation features by considering engineering knowledge to ensure good interpretability of the extracted features, and then takes the extracted features as model inputs. The proposed model conducts failure mode recognition and RUL prediction as a collaborative way and outputs predicted failure modes and RUL of units by fully characterizing complex relationship between RUL and failure modes. The proposed model is expected to be widely applied to practical situations with complex structures of manufacturing systems and unknown failure thresholds.



Fig. 7. RUL prediction errors of in-service units using the proposed method and the benchmark method of type (3).

In our future work, we consider two related topics. *First*, the proposed method is trained based on historical units with known failure modes. Future works will focus on failure mode recognition and RUL prediction for scenarios where failures modes of historical units and in-service units are both unknown. *Second*, as the proposed method only considers one operational condition for failures mode recognition and RUL prediction, in future work we will consider scenarios with multiple operational conditions.

References

1. K. Liu, N. Gebraeel, and J. Shi, “A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis,” IEEE Trans. Autom. Sci. Eng., vol. 10, no. 3, pp. 652–664, Apr. 2013.
2. H. Yan, K. Liu, X. Zhang, and J. Shi, “Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions,” IEEE Trans. Rel., vol 65, no. 3, pp. 1416–1426, Jul. 2016.
3. D. Liu, J. Zhou, H. Liao, Y. Peng, and X. Peng, “A health indicator extraction and optimization framework for lithiumion battery degradation modeling and prognostics,” IEEE Trans. Syst. Man Cybern. Syst., vol. 45, no. 6, pp. 915–928, Jun. 2015.
4. L. Yang and J. Lee, “Bayesian belief network-based approach for diagnostics and prognostics of semiconductor manufacturing systems,” Robot. Cim-Int. Manuf., vol. 28, no. 1, pp. 66–74, Feb. 2012.
5. A. Chehade, C. Song, K. Liu, A. Saxena, and X. Zhang, “A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes,” *J. Qual. Technol.*, vol. 50, no. 2, pp. 150–165, Apr. 2018.
6. Jardine, A., D. Lin, and D. Banjevic. 2006. A review on machinery diagnostics and prognostics implementing condition- based maintenance. Mechanical Systems and Signal Processing 20 (7):1483–1510.
7. K. Liu and S. Huang, “Integration of data fusion methodology and degradation modeling process to improve prognostics,” IEEE Trans. Autom. Sci. Eng., vol. 13, no. 1, pp. 344–354, Jan. 2016.
8. K. Liu, A. Chehade, and C. Song, “Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis,” IEEE Trans. Autom. Sci. Eng., vol. 14, no. 3, pp. 1504–1514, Jul. 2017.
9. Khan, S. and Yairi, T. (2018) A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing, 107, 241–265.
10. Tian, Z., Wong, L. and Safaei, N. (2010) A neural network approach for remaining useful life prediction utilizing both failure and sus- pension histories. Mechanical Systems and Signal Processing, 24(5), 1542–1555.
11. Zhu, J., Chen, N. and Peng, W. (2018) Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Transactions on Industrial Electronics, 66(4), 3208–3216.
12. Guo, L., Li, N., Jia, F., Lei, Y. and Lin, J. (2017) A recurrent neural net- work based health indicator for remaining useful life prediction of bearings. Neurocomputing, 240, 98–109.
13. Huang, C.G., Huang, H.Z. and Li, Y.F. (2019) A bi-directional LSTM prognostics method under multiple operational conditions. IEEE Transactions on Industrial Electronics, 66(11), 8792–8802.
14. Liao, L., Jin, W. and Pavel, R. (2016) Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment. IEEE Transactions on Industrial Electronics, 63(11), 7076–7083.
15. Zhang, C., Lim, P., Qin, A.K. and Tan, K.C. (2017) Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2306–2318.
16. Simani, S., C. Fantuzzi, and R. Patton. 2013. Model-based fault diagnosis in dynamic systems using identification techniques. London: Springer.
17. Bunks, C., D. McCarthy, and T. Al-Ani. 2000. Condition-based maintenance of machines using hidden Markov models. Mechanical Systems and Signal Processing 14 (4): 597–612. doi:10.1006/mssp.2000.1309.
18. Barron J. Bichon John M. McFarland, Sankaran Mahadevan, Efficient surrogate models for reliability analysis of systems with multiple failure modes, 96 (2011) 1386–1395.
19. R. Ahmad, S. Kamaruddin, I.A. Azid, I.P. Almanar, Failure analysis of machinery component by considering external factors and multiple failure modes – a case study in the processing industry, Eng. Fail. Anal. 25 (2012) 182–192.
20. N.C. Xiao, H.C. Huang, Y.F. Li, L.P. He, T.D. Jin, Multiple failure modes analy- sis and weighted risk priority number evaluation in FMEA, Eng. Fail. Anal. 18 (2011) 1162–1170.
21. D. Zhang, S. Tang, J. Che, Concurrent subspace design optimization and analysis of hypersonic vehicles based on response surface models, Aerosp. Sci. Technol. 42 (2015) 39–49.
22. N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *Am. Stat.*, vol. 46, no. 3, pp. 175–185, Aug. 1992.
23. T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
24. M.D. Moura, E. Zio, I.D. Lins, Failure and reliability prediction by support vector machines regression of time series data, Reliab. Eng. Syst. Saf. 96 (11) (2011) 1527–1534.
25. L.K. Song, G.C. Bai, C.W. Fei, J. Wen, Reliability-based fatigue life prediction for complex structure with time-varying surrogate modeling, Adv. Mater. Sci. Eng. (2018) 3469465.
26. Barron J. Bichon, John M. McFarland, Sankaran Mahadevan, Efficient surrogate models for reliability analysis of systems with multiple failure modes, Reliability Engineering and System Safety 96 (2011) 1386–1395.
27. Lu-Kai Song, Guang-Chen Bai, Cheng-Wei Fei. Multi-failure probabilistic design for turbine bladed disks using neural network regression with distributed collaborative strategy. Aerospace Science and Technology 92 (2019) 464–477.
28. Xiao Liu, Jingrui Li, Khalifa N. Al-Khalifa, Abdelmagid S. Hamouda , David W. Coit & Elsayed A. Elsayed (2013) Condition-based maintenance for continuously monitored degrading systems with multiple failure modes, IIE Transactions, 45:4, 422-435.
29. M. Kim, and K. Liu, “A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics,” IISE Trans., vol. 53, no. 3, pp. 326–340, Jan. 2021.
30. Lu-Kai Song, Guang-Chen Bai, Cheng-Wei Fei, Multi-failure probabilistic design for turbine bladed disks using neural network regression with distributed collaborative strategy, Aerospace Science and Technology, 92 (2019) 464–477.
31. J. C. Lu and W. Meeker, “Using degradation measures to estimate a time-to-failure distribution,” *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.
32. Y. Chauvin, D. E. Rumelhart, “Backpropagation: theory, architectures, and applications,” *J. Am. Soc. Inf. Sci.*, vol. 48, no. 1, pp. 88–89, Jan. 1997.
33. Y. Liu, D. K. Frederick, J. A. DeCastro, J. S. Litt, and W. W. Chan, “User’s guide for the commercial modular aero-propulsion system simulation (C-MAPSS)”, *NASA* *Glenn Research Center*, Mar. 2012.

1. This work was supported in part by the National Science Foundation of China under Grant 72101148; in part by the Shanghai Sailing Program under Grant 21YF1420100, and in part by the National Science Foundation of Shanghai under Grant 22ZR1433000.

   Di Wang is with the department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China (email: d.wang@sjtu.edu.cn)

   Xiaochen Xian is with the department of Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA (email: xxian@ufl.edu)

   Changyue Song is with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, USA (e-mail: csong14@stevens.edu) [↑](#footnote-ref-1)