

## Research Article

# A Novel Approach of Label Construction for Predicting Remaining Useful Life of Machinery

Hailong Lin <sup>1</sup>, Zihao Lei <sup>2</sup>, Guangrui Wen <sup>2</sup>, Xiaojun Tian <sup>1</sup>, Xin Huang <sup>2</sup>,  
Jinsong Liu <sup>1</sup>, Haoxuan Zhou <sup>2</sup>, and Xuefeng Chen <sup>2</sup>

<sup>1</sup>SDIC Biotechnology Investment Co., Ltd., Beijing 100034, China

<sup>2</sup>School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Zihao Lei; [zihao\\_lei@163.com](mailto:zihao_lei@163.com)

Received 14 September 2021; Accepted 6 December 2021; Published 18 December 2021

Academic Editor: Changqing Shen

Copyright © 2021 Hailong Lin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearings are key components of rotating machinery, and predicting the remaining useful life (RUL) is of great significance in practical industrial scenarios and is being increasingly studied. A precise and reliable remaining useful life prediction result provides valuable information for decision-makers, which is essential to ensure the safety and reliability of mechanical systems. Generally, the RUL label is considered to be an ideal life curve, which is the benchmark for RUL prediction. However, the existing label construction methods make more use of expert experience and seldom mine knowledge from data and combine experience to assist in constructing a health index (HI). In this paper, a novel and simple approach of label construction is proposed for predicting the RUL accurately. More specifically, the degradation index of the multiscale frequency domain is first extracted. Furthermore, the fuzzy C-means (FCM) algorithm is innovatively used to divide the degradation data into several stages to obtain the turning point of degradation. Then, a nonlinear degradation index, the RUL label with the turning point, was constructed based on principal component analysis (PCA). Finally, the recurrent neural network (RNN) is used for prediction and verification. In order to verify the effectiveness of the proposed approach, two different bearing lifecycle datasets are gathered and analyzed. The analysis result confirms that the proposed method is able to achieve a better performance, which outperforms some existing methods.

## 1. Introduction

Rolling bearings are widely used in various mechanical systems as one of the most critical components. Among them, the failure of rolling bearings is one of the most important causes of mechanical system failure [1]. Therefore, the diagnosis and prognosis of bearings play an important role in the performance of mechanical equipment [2–4]. Predicting the RUL of bearings is of great importance to prevent sudden failures in mechanical systems and has also received much attention as a key issue in prognostics and health management (PHM) [5–8].

In general, RUL prediction methods could be mainly classified into model-based, data-driven, and hybrid

methods [9]. In recent years, more and more data-driven methods have been proposed for RUL prediction. Lei et al. divided the data-driven RUL prediction into four main steps, including data acquisition, HI construction, health stage division, and RUL prediction [10].

The RUL label is considered to be the ideal life curve of the equipment, which means the remaining useful life and corresponds to each operating cycle [11]. As for RUL prediction of rolling bearings, the RUL label is commonly regarded as the benchmark of accuracy evaluation for prediction results [12]. As research on data-driven prediction approaches continues to advance, the RUL label leaves more significant impacts on the model training process. Since the prediction model is also a kind of neural network,

its parameters are also obtained by satisfying the following conditions: First, the parameters are initialized, and then the predicted label is obtained by forward propagation. Next, the loss between the predicted label and the RUL label is calculated, and finally, the parameters are updated by back-propagation gradient descent. Therefore, reasonable label plays a critical role in RUL prediction of rolling bearings and greatly affects the accuracy and generalization ability of prediction models [13, 14].

According to previous studies, for the construction of the RUL label, the following three methods are mainly classified: (1) Failure determination based on the fault threshold. As the most common method in early RUL prediction studies, the maximum allowable vibration value or experienced value is usually considered the failure threshold for real situations [15–19], and it is clear to determine the failure time with fixed thresholds. However, it lacks tolerance to sudden noise in operating conditions and is difficult to obtain an ideal fault threshold in advance for a new component. (2) The RUL label based on the ideal degradation curve. In [11, 12, 20, 21], linear function and improved piecewise function were used to fit the degradation curve as much as possible. Among them, when a linear function is used, the RUL value decreases linearly with the time period. When the piecewise function is used, the RUL value remains unchanged during the previous period and then linearly decreases. In different degradation stages, the degradation rate will be different, which is then reflected in the slope of the degradation curve. The above two labels are the most widely used in the methods of RUL prediction. The key point is to find the transition time of different failure degrees, that is, the turning point of degradation, and set a more ideal RUL value. (3) The RUL label based on the degradation index. Some recent research studies [22–25] use one specific index to construct the life curve and evaluate the bearing degradation state, such as kurtosis, skewness, and other common statistical features. However, the single indicator is difficult to contain enough information and can easily be affected by fluctuations, and this is apparently contrary to the stable and monotonous performance of RUL label.

There is no doubt that the construction of RUL labels is crucial for RUL prediction because model training and result evaluation of RUL prediction require ideal life curves as a benchmark, yet most of the actual industrial field rolling bearing-monitoring data are unlabeled data. As an important part of RUL prediction, it relies more on expert experience, which represents the current linear function or piecewise linear function, rather than constructing the RUL label based on the knowledge mined by the data itself [26]. And this is obviously insufficient for PHM in the era of big data for the mining of knowledge in the data, especially when predictive maintenance is driven by the rapid development of big data [27, 28]. In addition, it can impose limitations on the prediction accuracy in practical applications. Despite these strong motivations for research, there are few studies and certain gaps on how to construct a more reasonable RUL label, which is completely incompatible with its important status of predictive maintenance.

Considering the problems mentioned above, a novel and simple RUL label construction approach is proposed in this paper and the accurate RUL prediction of rolling bearings is achieved based on the RNN. First, ensemble empirical mode decomposition (EEMD) is used to decompose signals into several intrinsic mode functions (IMFs). The frequency-domain features and energy of the IMFs are extracted and fused to obtain the degradation index using PCA. Then, the FCM algorithm is applied to detect the turning points between normal condition, slight fault, and heavy fault. It is rather remarkable that the new RUL label is constructed based on the degradation index and the known turning points. To guarantee the robustness of the RUL label, the anomalous jump points of the degradation index are further eliminated using the linear regression method. Finally, to verify the effectiveness and superiority of the proposed method, the simplest recurrent neural network is constructed for RUL prediction. And the major contributions of this work can be summarized as follows:

- (1) An improved life label is proposed based on the health index and turning points of failure stages, which provides a new idea to improve the existing methods of constructing the RUL label based on human experience or single health index.
- (2) The proposed approach is capable of adaptively constructing the novel label that reflects changes in the rate of degradation at different stages while being more suitable for practical applications. Moreover, the construction of new label relies on the knowledge mined from the data rather than just the experience of experts.
- (3) It provides the solution for researchers to construct the RUL label for new equipment in time once the fault occurs. Therefore, it is also promising to be applied in online RUL prediction. Some experimental results also confirm the effectiveness and superiority of the proposed method.

The rest of this paper is organized as follows: Basic methods including the RNN model and FCM algorithm are introduced in Section 2. In Section 3, the principle and schematic diagram of the proposed method are presented. Section 4 shows the results of experiments on two datasets, XJTU-SY bearing data and IMS bearing data. Finally, the summary and conclusions are given in Section 5.

## 2. Theoretical Background

*2.1. RNN.* Because of the outstanding ability to handle time-series data, the RNN model is suitable to be used in RUL prediction. There are many developed versions of RNN, but the basic principles of these networks remain unchanged. Thus, the classical structure of the RNN model is applied for the prediction of RUL in this work.

The classical RNN model is shown in Figure 1. It is assumed that there is a certain part of the equipment, which

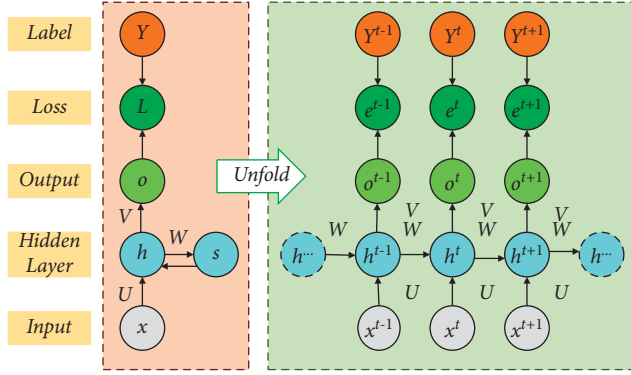


FIGURE 1: Standard and unfolded RNN.

provides a set of time-series data of sensors like vibration. The data collected from the equipment can be presented as  $X = [X^1, X^2, \dots, X^t, \dots, X^T]$ ,  $X \in R^{1 \times T}$ , where  $X^t$  is the sampling signal at time  $t$  ( $t = 1, 2, \dots, T$ ). And  $X$  is often replaced by features extracted from raw data in application. The RNN takes time-series data  $X$  as model input, and then the RUL is obtained as model output. Parameters of the RNN model are optimized via implicit function mapping and error backpropagation through time (BPTT) algorithm. Taking time order into consideration in error propagation, the RNN shows great superiority in time-series data processing.

The RNN model is described in Figure 1 at time  $t$  with its fold form on the left and unfold form on the right, where  $X^t$  denotes the input at time  $t$ ,  $h^t$  denotes the hidden state at time  $t$ ,  $O^t$  denotes the output at time  $t$ ,  $L$  denotes the loss function,  $e^t$  denotes the error, and  $Y^t$  denotes the real RUL at time  $t$ , which is often replaced by the provided RUL label. Besides, matrixes  $U, V, W$  are passing parameters of the RNN model from input nodes to hidden nodes, from hidden nodes to output nodes, and from hidden nodes at time  $t-1$  to hidden nodes at time  $t$ , respectively.

In Step 1, signals propagate forward along the arrows.  $h^t$  and  $O^t$  are given as follows:

$$\begin{aligned} h^t &= \psi(UX^t + Wh^t + b), \\ O^t &= \sigma(Vh^t + c), \end{aligned} \quad (1)$$

where  $\psi$  and  $\sigma$  are activation functions and  $b$  and  $c$  are the deviation of the input layer and output layer, respectively.

In Step 2, calculation errors propagate back forward through time at every iteration. While tanh is used in the hidden layer and softmax is used in the output layer, the error  $e^t$ , total loss  $L$ , and partial derivatives of  $U, V, W, b$ , and  $c$  are calculated as follows:

$$e^t = Y^t - O^t,$$

$$L = \sum_{t=1}^T e^t,$$

$$\left\{ \begin{aligned} \nabla U &= \frac{\partial L}{\partial U} = \sum_{t=1}^T \frac{\partial e^t}{\partial h^t} \frac{\partial h^t}{\partial U} \\ &= \sum_{t=1}^T \text{diag}\left(1 - (h^t)^2\right) \frac{\partial L}{\partial h^t} (X^t)^T, \\ \nabla V &= \frac{\partial L}{\partial V} = \sum_{t=1}^T \frac{\partial e^t}{\partial O^t} \frac{\partial O^t}{\partial V} = \sum_{t=1}^T (\hat{Y}^t - Y^t) (h^t)^T, \\ \nabla W &= \frac{\partial L}{\partial W} = \sum_{t=1}^T \frac{\partial e^t}{\partial h^t} \frac{\partial h^t}{\partial W} \\ &= \sum_{t=1}^T \text{diag}\left(1 - (h^t)^2\right) \frac{\partial L}{\partial h^t} (h^t - 1)^T, \\ \nabla b &= \frac{\partial L}{\partial b} = \sum_{t=1}^T \frac{\partial e^t}{\partial h^t} \frac{\partial h^t}{\partial b} = \sum_{t=1}^T \text{diag}\left(1 - (h^t)^2\right) \frac{\partial L}{\partial h^t}, \\ \nabla c &= \frac{\partial L}{\partial c} = \sum_{t=1}^T \frac{\partial e^t}{\partial O^t} \frac{\partial O^t}{\partial c} = \sum_{t=1}^T (\hat{Y}^t - Y^t). \end{aligned} \right. \quad (2)$$

Through the circles of Step 1 and Step 2, parameters of the RNN model are optimized and the objective function of total loss is minimized; thus, the predicted RUL is as close as possible to its real value. The training process will be stopped when there is no significant improvement on prediction accuracy or when iteration times reach the default number.

**2.2. FCM.** Clustering is a suitable technique for handling the prognosis tasks without labels, since it aims to organize a set of samples into the corresponding groups based on similarity. Therefore, clustering is actually described as the method for grouping unlabeled data. As the extended version of K-means algorithm, the FCM algorithm was proposed by Bezdek and possesses the ability of assigning each sample to each cluster in a certain degree.

Due to its superiority in dealing with the uncertainty and independence of labels, the FCM algorithm has been widely applied to fault diagnosis of rotating machinery. The FCM algorithm defines and assembles samples into certain classes via minimizing the objective function and calculating the membership degree to clustering centers. The set

of samples can be presented as  $X = [x_1, x_1, \dots, x_1]$ , including  $c (c > 1)$  hidden subsets totally. The clustering center  $V_i$ , membership degree  $V_{ij}$ , and objective function  $J$  are calculated as follows:

$$V_i = \frac{\sum_{j=1}^n V_{ij}^m x_j}{\sum_{j=1}^n V_{ij}^m}, \quad (3)$$

$$V_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij}/d_{ik})^{2/(m-1)}}, \quad (4)$$

$$J = \sum_{i=1}^c \sum_{j=1}^n V_{ij}^m d_{ij}^2, \quad (5)$$

where  $V_{ij}$ ,  $m$ , and  $d_{ij}$ , respectively, represent the membership degree, the weighted index, and the Euclidean distance from the  $j$ -th sample to the  $i$ -th clustering center. The FCM algorithm can be summarized into three steps as follows.

*Step 1.* Initialize parameters including the number of subsets  $c (c > 1)$ , fuzzy degree  $q_{ij}$ , membership degree matrix  $V$ , number of iterations, and threshold.

*Step 2.* Update clustering centers according to the membership degree matrix  $V$  and then recalculate the membership degree  $V_{ij}$ .

*Step 3.* Estimate the deviation of function  $J$ .

All these steps will be stopped if the deviation or the absolute value of function  $J$  is less than the threshold; otherwise, Step 2 and Step 3 will be repeated.

### 3. Proposed Method

In this section, an RNN-based approach is proposed for RUL prediction. Frequency-domain and time-frequency domain features are used as the input of PCA to obtain the degradation index, and the turning points obtained from the FCM algorithm are used to reconstruct the degradation index into normal condition, slight fault, and heavy fault. The ideal RUL label is then generated to train the RNN model accurately and synchronously. The proposed method mainly consists of three steps, as shown in Figure 2.

*3.1. Data Preprocessing.* Vibration energy of frequency domain contains information of spectral distribution and position change for the main frequency band. And related frequency-domain features are sensitive to bearing degradation since an imperceptible change produces a spectrum line in the corresponding frequency spectrum. Therefore, it is vital for fault prognosis to extract some indicators in the frequency domain. Besides, as a self-adaptive signal processing technique, EEMD decomposes a signal into several IMFs and one residual. The number and selection of these

IMFs depend on the signal itself. Thus, the internal oscillatory modes imbedded in the signal are denoted by IMFs.

The first twelve-dimensional frequency-domain features and the energy of the first  $N$  IMFs are extracted for each sample. The default energy of single IMF is zero when the decomposition result is less than default  $N$ . And  $N$  is set as ten in this study as the number of obtained components from EEMD is about ten for all bearing vibration signals. In order to ensure the stable performance of the training model, the min-max normalization is applied for data preprocessing as in

$$x'_i = \frac{x_i - \min x_i}{\max x_i - \min x_i}. \quad (6)$$

*3.2. RUL Label Construction.* The FCM algorithm provides the membership degree value of each sample for each clustering center. As the total membership degree value of one sample for all groups is 1, the maximum corresponds to the most likely group that the sample belongs to. Besides, the FCM algorithm has no requirements on the amount of data. Once the obtained data are provided, the algorithm clusters samples into default groups based on similarity. Considering that the bearing's life circle is often divided into three stages of normal condition, slight fault, and heavy fault, the default value of hidden clusters is set as three. With the objective function being optimized in the FCM algorithm, the membership degree values of each sample for each cluster are shown in Figure 3.

As seen in Figure 3, the membership degree from sample 1 to sample  $n1$  is 0.85 and that of the other samples is below 0.5 in the first cluster. Thus, the samples from the first to  $n1$  belong to the first group. Similarly, samples from  $n1 + 1$  to  $n2$  belong to the second group, and samples from  $n2 + 1$  to the last belong to the third group. So,  $n1 + 1$  is the turning point between normal condition and slight fault, and  $n2 + 1$  is the turning point between slight fault and heavy fault.

One degradation index is generated by dimensionality reduction of PCA, which is recently regarded as the RUL label in other research studies. Since the linear decreasing relationship is the fatal trend of real life, the new RUL label of three degradation phases is constructed by linear regression based on the degradation index, as shown in Figure 4.

As shown in Figure 4, there are several abnormal points such as the drastic rise point B and straight decrease point A. Obvious deteriorations of bearing condition appear at these moments since the vibration or noise is rapidly increasing. Thus, the value of the degradation index is usually near the extremum, which causes conflicts with the reality of rapid reduction in remaining useful life. Therefore, these abnormal points are eliminated here. One point will be regarded as abnormal once its deviation is three times larger than the average deviation of several adjacent points as in (7) and (8).  $2a$  adjacent points are used for calculation, and linear regression is applied to fit the degradation index when all abnormal points are eliminated. And  $a = 5$  in this work:

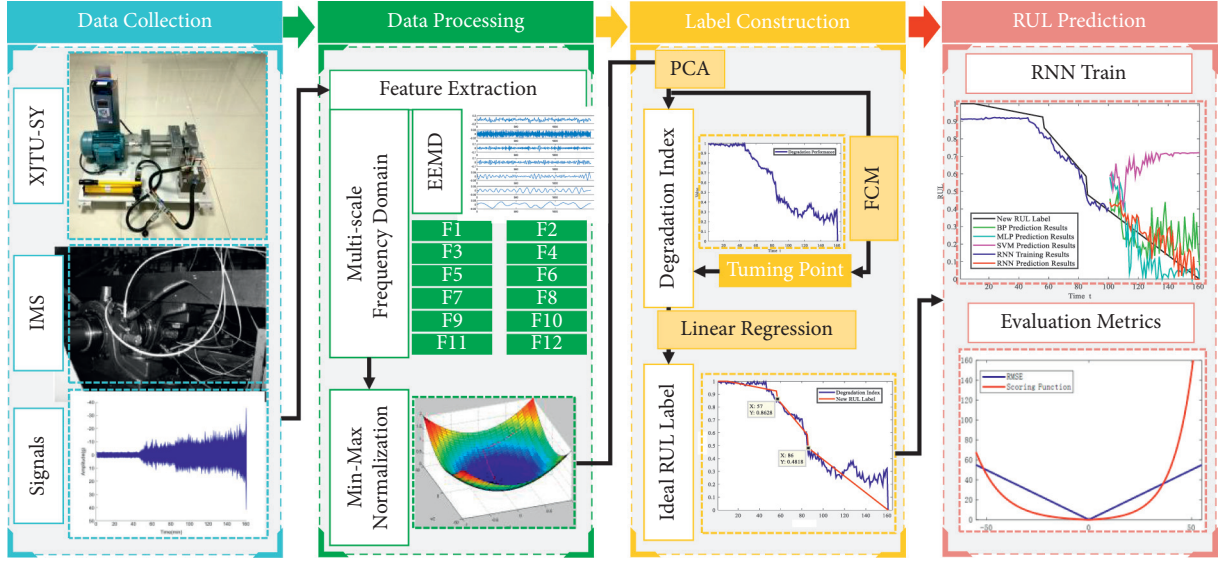


FIGURE 2: Schematic diagram of the proposed method.

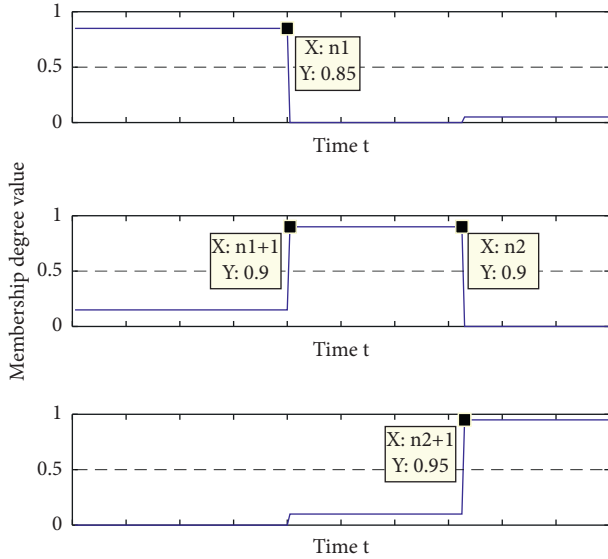


FIGURE 3: Clustering results of the FCM algorithm.

$$f(t) = f_T(t) + f_R(t), \quad (7)$$

$$M(t) = N \times \left[ \frac{(\sum_{t-a}^{t+a} f_R(t) - f_R(t))}{2a} \right], \quad (8)$$

where  $f(t)$  is the degradation index at time  $t$ ,  $f_T(t)$  is the mean trend of the degradation index at time  $t$ ,  $f_R(t)$  is the random part of the degradation index at time  $t$ , and  $M(t)$  is  $N$  times the average deviation of  $2a$  adjacent points at time  $t$ . One point named " $f(t)$ " will be replaced by  $f_T(t)$  when  $f_R(t)$  is larger than  $M(t)$ .

**3.3. RUL Prediction.** In order to simulate the online RUL prediction procedure for rolling bearings, the batch size of model training is set as one. Thus, the RNN model will be

updated for each new sample. Once the first turning point is detected by the FCM algorithm, the model is going to learn degradation modes and obtain the ability of prediction. And the operation data of previous moments are used to predict the RUL in the future. The demarcation point of training and prediction is determined by an experienced ratio of two to one, when the first turning point between normal condition and slight fault appeared in the first two-thirds of samples. Otherwise, the training set should be expanded until a set of failure samples are also learned by the RNN model.

As shown in Figure 5, the RNN model is constructed based on the frequency-domain features and energy features of the first ten IMFs. With loop iteration for parameter optimization, two hidden layers are established in the RNN model. The key work in the model is the update of weight matrixes for all nodes, which are trained by the BPTT algorithm. The square root error is used as the loss function for partial derivative calculation in the RNN model. Besides,  $\tanh$  is used as the activation function in the input layer, and  $\text{relu}$  is applied in the output layer.

Suppose  $X^t$  is a feature vector extracted from a single sample at time  $t$  in the input layer. Outputs of the input layer, the two hidden layers, and the output layer at time  $t$  are, respectively, shown as follows:

$$I^t = U * X^t, \quad (9)$$

$$H_1^t = \tanh(I^t + W * H_1^{t-1} + b_1), \quad (10)$$

$$H_2^t = \tanh(H_1^t + W * H_2^{t-1} + b_2), \quad (11)$$

$$O^t = \text{relu}(V * H_2^t + c), \quad (12)$$

where  $U, V, W, b$ , and  $c$  are as the same as mentioned above, and original values of these parameters are initialized randomly from  $-1$  to  $1$ .

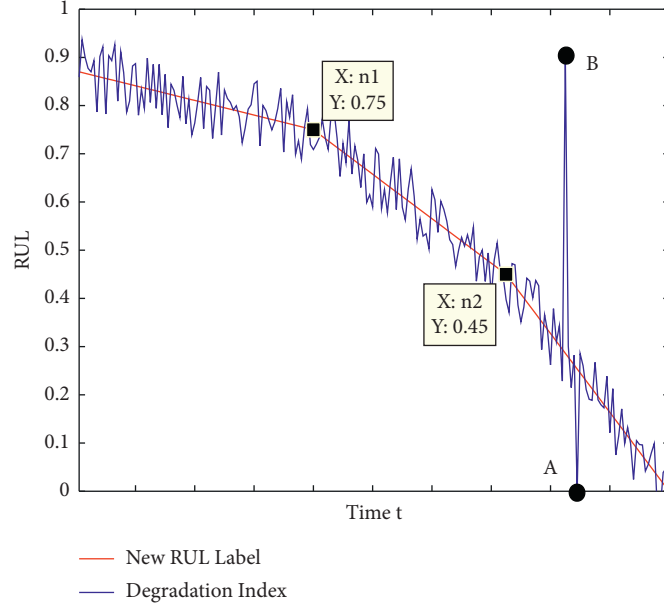


FIGURE 4: Reconstructed RUL label.

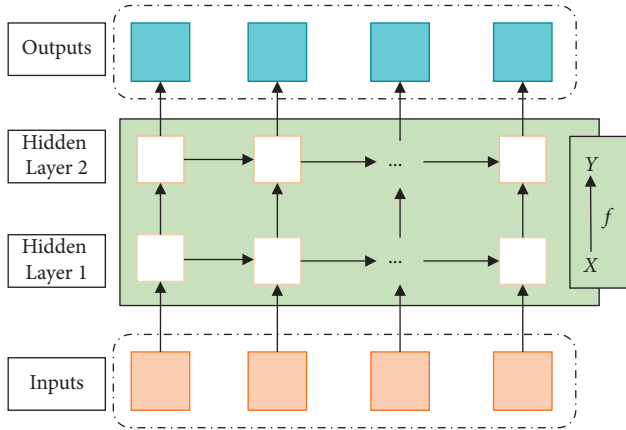


FIGURE 5: Structure of the RNN model.

The ideal RUL at time  $t$  is obtained as  $Y^t$  according to the new RUL label. The loss of output is calculated with the ideal RUL, and the partial derivative of model parameters is then updated. Through the repeated forward calculation and BPTT algorithm, the training process of the RNN model is completed. Compared with traditional training algorithms based on gradient descent, the decoupled extended Kalman filter algorithm shows significant advantages on computation and performance. And it is applied in prediction model training. Evaluation is finally conducted based on the prediction results and the ideal RUL of test data.

#### 4. Experimental Study

In experiments, XJTU-SY bearing data and IMS bearing data are used. Compared with other approaches including the backpropagation (BP) network, support vector machine (SVM), and multilayer perceptron (MLP) and traditional functions including linear RUL function and piecewise RUL

function, the prediction results of the proposed method show a significant superiority.

##### 4.1. Case Study 1: XJTU-SY Datasets

**4.1.1. Data Description.** XJTU-SY Bearing Datasets are provided by the Institute of Design Science and Basic Component, Xi'an Jiaotong University (XJTU), and the Changxing Sumyoung Technology (SY). They contain complete run-to-failure data of fifteen rolling bearings by conducting several accelerated degradation experiments [12]. Bearings of type LDK UER204 were operated in totally three conditions, and five bearings were tested under each operating condition. The sampling rate was kept at 25.6 kHz, and the sampling interval was equal to one minute. The bearing test bed is shown in Figure 6.

**4.1.2. Evaluation Metrics.** In this section, the prediction performance of the proposed method is evaluated quantitatively by employing scoring function and root mean square error (RMSE), which are described, respectively, as follows.

(1) *Scoring Function.* This is different when the measurement runs ahead of the real value or when the predicted RUL value lags behind the real value. The definition of scoring function is shown as follows:

$$\text{score} = \begin{cases} \sum_{i=1}^n (e^{h_i/13} - 1), & h_i < 0, \\ \sum_{i=1}^n (e^{h_i/10} - 1), & h_i \gg > 0, \end{cases} \quad (13)$$

where  $h_i$  is the difference between  $x_i$  and  $h_i$  as mentioned above.

Furthermore, a smaller score means a better prediction result. Besides, scoring function gives a different penalty

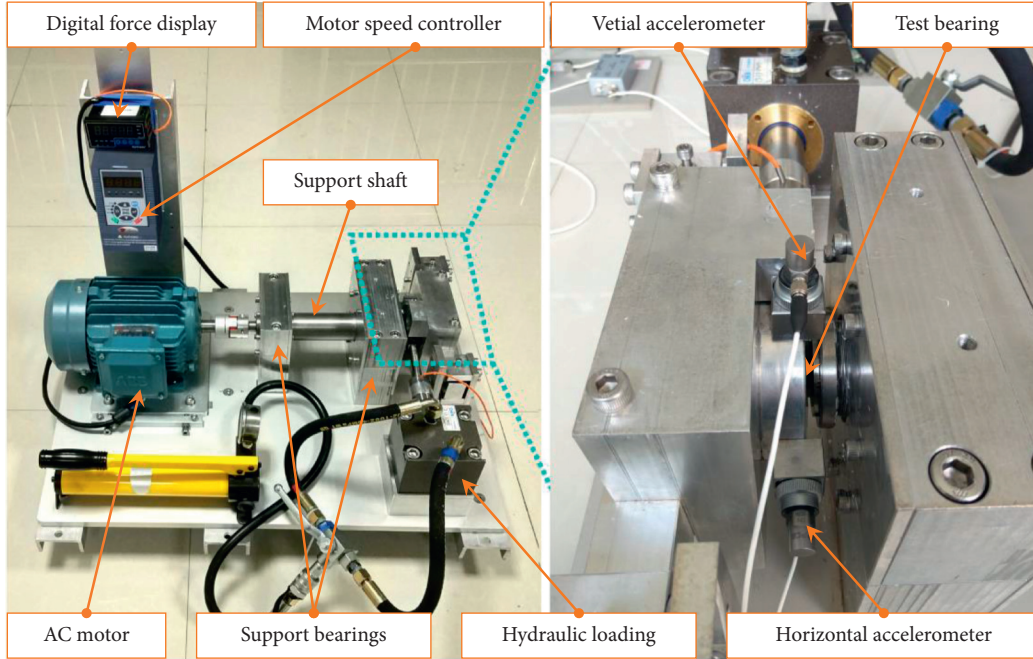


FIGURE 6: Test bed of XJTU-SY bearing datasets.

when the model underestimates the RUL and when the model overestimates the RUL because there remains little time for maintenance if the predicted RUL is larger than the real one. Faults or disasters are going to happen soon, and prediction gets no significance at all.

(2) *RMSE*. This is widely used for RUL prediction, which is the root of error square divided by the number of samples. The definition of RMSE is shown as follows:

$$\text{RMSE} = \sqrt{\frac{\sum (x_i - d_i)^2}{n}}, \quad (14)$$

where  $x_i$  represents the measurement,  $d_i$  is the real value, and  $n$  refers to the number of samples.

This statistical index reflects the extent from measurements to real values and gives the same value no matter the error is positive or negative. Thus, a smaller RMSE value means a better prediction result.

**4.1.3. Experimental Results and Analysis.** Operation data in the vertical axis of bearing 2-2 are extracted for this experiment. The operating conditions include a rotation speed of 2250 RPM and hydraulic loading of 11 KN. After a constant operation of two hours and forty-one minutes, fault occurred in the outer race, and 161 sampling files are obtained. The former 100 sampling files are used for model training to predict the RUL for the next 61 test samples. The degradation index obtained by PCA is shown in Figure 7. In previous studies, based on expert experience, the degradation stages of bearings are generally divided into three stages: normal condition, slight fault, and severe fault [4]. Therefore, the degradation process can be divided into stages according to the FCM algorithm mentioned in Section 2.2,

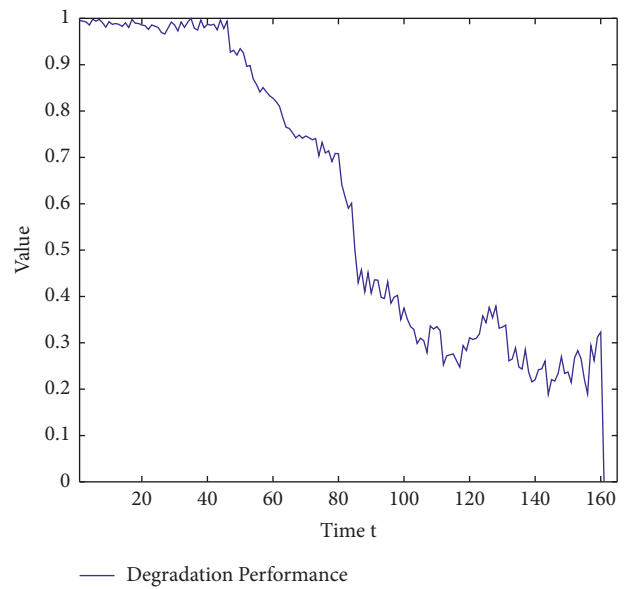


FIGURE 7: Degradation index of bearing 2-2.

and then the membership degree of each sample to each cluster center can be obtained.

As mentioned above, three clustering centers are set in the FCM algorithm with the maximum iteration as 100 and the error as  $1e-6$ . The result of membership degree function is shown in Figure 8. The membership degree from sample 1 to sample  $n_1$  is 0.85, and that of the other samples is below 0.5 in the first cluster. Thus, the samples from the first to  $n_1$  belong to the first group. Similarly, samples from  $n_1 + 1$  to  $n_2$  belong to the second group, and samples from  $n_2 + 1$  to the last belong to the third group. So,  $n_1 + 1$  is the turning point between normal condition

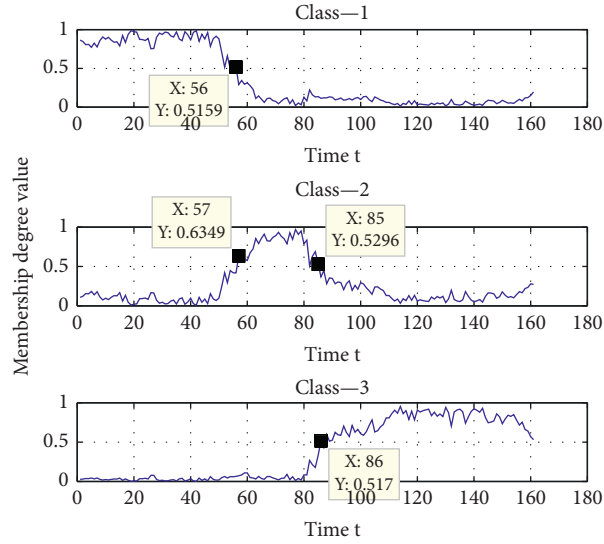


FIGURE 8: Clustering results of bearing 2-2.

and slight fault, and  $n_2 + 1$  is the turning point between slight fault and heavy fault. It can be found that bearing 2-2 is operating normally from the beginning to the 56th sample. The slight fault occurred at the 57th sample, and the heavy fault started at the 86th sample. Based on the turning points, linear regression analysis is applied to construct the new RUL label, as shown in Figure 9.

With the novel RUL label constructed, the RNN model is applied for RUL prediction. Through 10 test runs, the model structure and parameters corresponding to the best prediction result are obtained. The RNN model with two hidden layers is able to discover the hidden patterns from the inputs, and the suitable maximum iteration number is 2000. In Figure 10, the results of RUL prediction at any time step of the test set are given. The prediction results of the test set are given by the red line, while the results of model training are given by the blue line. The results of RUL prediction are fluctuating around the ideal RUL, which proves the prediction ability of the RNN model.

Different models including the classic RNN, back-propagation (BP) network, multilayer perceptron (MLP), and support vector regression (SVR) are compared in this paper in Figure 11. Considering the unpredictable performance of BP networks and MLP, the best network structure is obtained by 10 loop iterations. Finally, the three-layered node 22-27-1 is used in the BP network, which means 22 nodes in the input layer, 27 nodes in the hidden layer, and 1 output node. Two hidden layers with the structure of 22-25-7-1 are used in the MLP. In the SVR model, RBF kernel function is used with gamma value 4, cost value 1.5, and  $p$  value  $1e-5$  in the loss function.

Linear and piecewise RUL functions are both discussed. Bearing life decreases linearly from one to zero between the first and the last sample in linear function. As shown in Figure 12, the turning point of piecewise function is set at the 70th sample and the original RUL is determined at 130 finally after repeated experiments in the

experienced range. The structures of models remain the same as before.

The RMSE and scoring function are adopted to evaluate the performances of models. The comparison of prediction results with different labels and models is shown in Table 1.

As shown in Table 1, the novel RUL label shows a better performance than linear and piecewise RUL functions among all four models, which strongly proves the superiority of the proposed method for RUL label construction. Besides, the RNN obtained the best result of RUL prediction with both the novel RUL label and the piecewise RUL label, followed by the MLP and BP network. And SVR shows little potential for prediction in this experiment.

Turning points are the key information of the proposed method. As the degradation of rolling bearings is usually continuous, the adjacent samples in the time series are usually clustered into the same group until the turning point appears. The influence on clustering results of fluctuations in input data is discussed in this part. Since the first turning point has been known, 10, 20, and 30 samples are added, respectively. Clustering results with three training sequence lengths are shown in Figure 13.

As seen in Figure 13, misjudgment takes place when there exist only 67 samples. And the first turning point comes earlier than the 57th sample. It is because that the clustering centers are farther away from the correct positions when there are fewer samples for training. And the small fluctuations in clustering centers will cause large changes in calculated membership degree values according to the principles of the FCM algorithm. Samples that actually belong to the first group but are far away from the calculated center are misjudged. Thus, the situation will be improved with more samples provided. Through experiments, it is noted that about twenty samples are required for generating accurate turning points, which is not a serious problem with the reality of high-frequency sampling rate and relatively slow degradation rate at the two turning points.



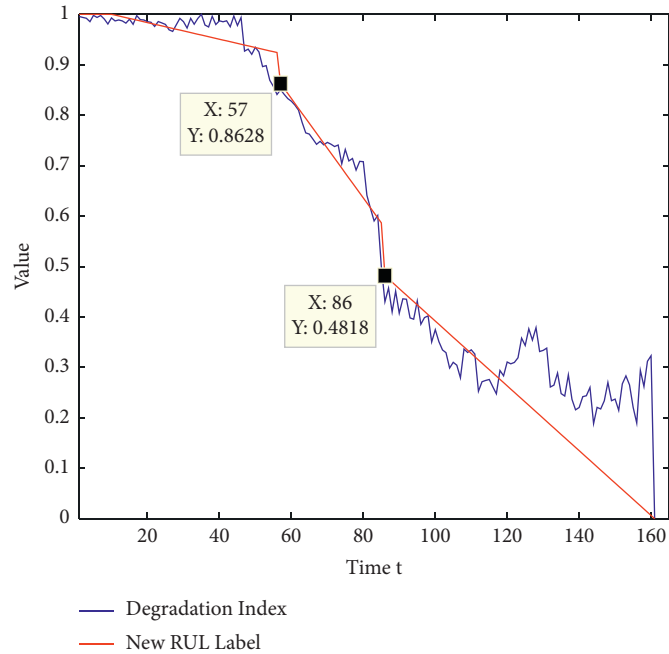


FIGURE 9: Novel RUL label of bearing 2-2.

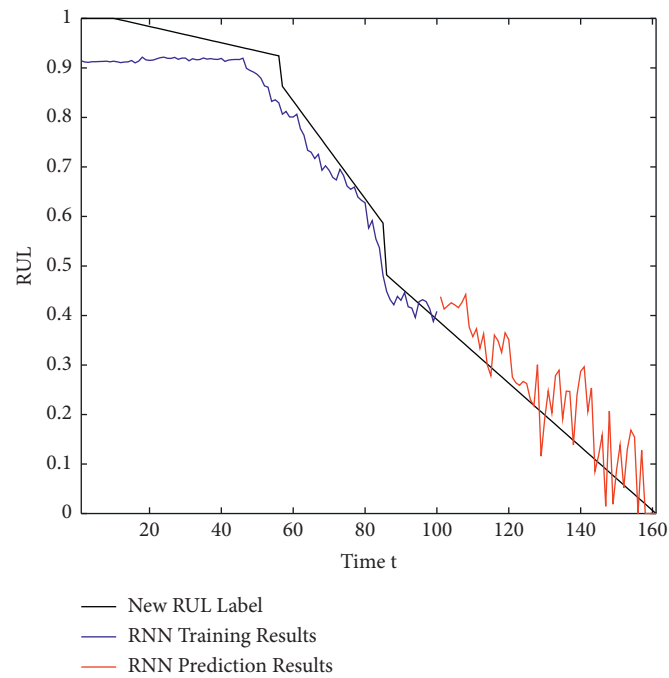


FIGURE 10: Prediction results of the RNN model.

4.2. Case Study 2: IMS Bearing Data

4.2.1. Data Description. IMS (Intelligent Maintenance Systems of NSF I/UCR Center) bearing data include three sets of degradation experiments among four rolling bearings. These experiments were carried out using bearings of type Rexnord ZA-2115. The test bench structure is shown in Figure 14. As the most commonly used dataset in bearing life prediction research studies, Set 2 is applied for verification of the proposed method.

Set 2 describes a situation that the outer race failure occurred in bearing 1 at the end of experiment, containing 984 sampling files in total. The rotation speed was kept constantly at 2000 RPM, and a radial load of 6000 lbs was applied onto the shaft and bearing by a spring mechanism. Besides, eight accelerometers were mounted, respectively, in vertical and horizontal directions of four bearings with a sampling rate of 20 kHz and a sampling interval of 10 minutes.

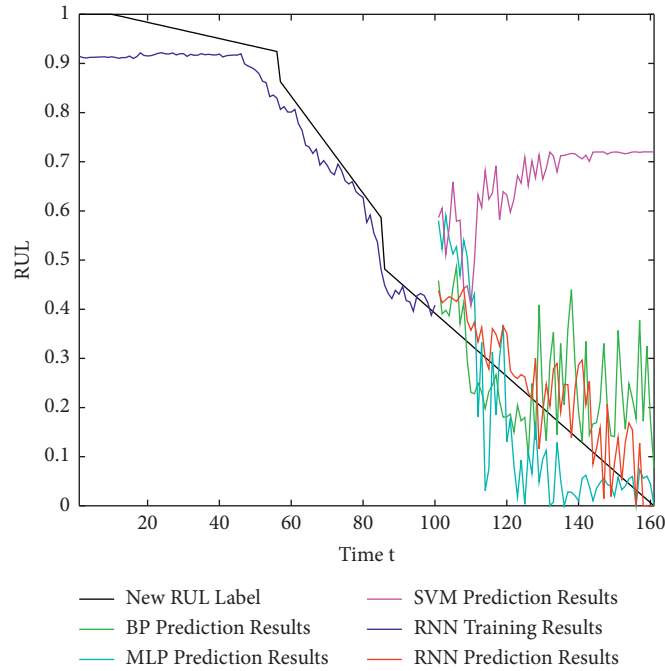


FIGURE 11: Prediction results of four models.

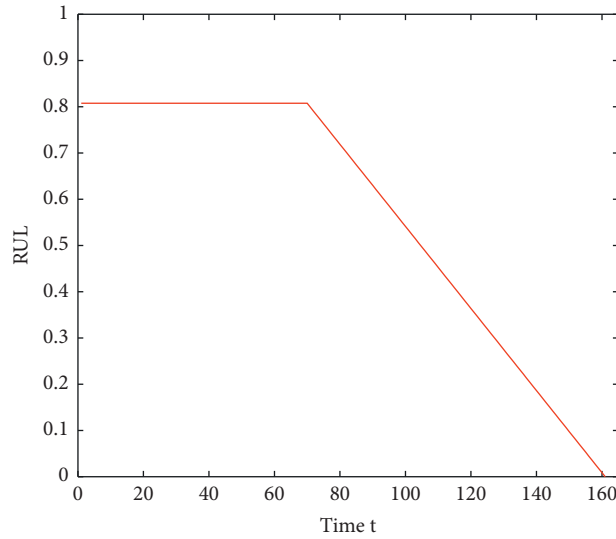


FIGURE 12: Piecewise RUL function.

TABLE 1: Evaluation results of three labels and four models (XJTU-SY).

Models	New RUL label		Linear RUL label		Piecewise RUL label	
	Score	RMSE	Score	RMSE	Score	RMSE
RNN	0.4173	0.2250	0.4492	0.5724	1.2704	1.6062
BP network	0.6192	0.4582	1.6552	1.9789	1.4326	1.7585
MLP	0.5369	0.2647	1.5889	1.9935	1.4882	1.8286
SVR	2.9445	3.6695	3.1515	3.9295	3.3848	4.2084

4.2.2. *Experimental Results.* 700 samples are used for model training to predict the next 284 outputs. PCA is applied for the feature’s dimensionality reduction. The FCM algorithm

is then applied for clustering, as shown in Figure 15. As seen in Figure 15, the bearing operates normally from the beginning to the 533rd sample and gets into slow

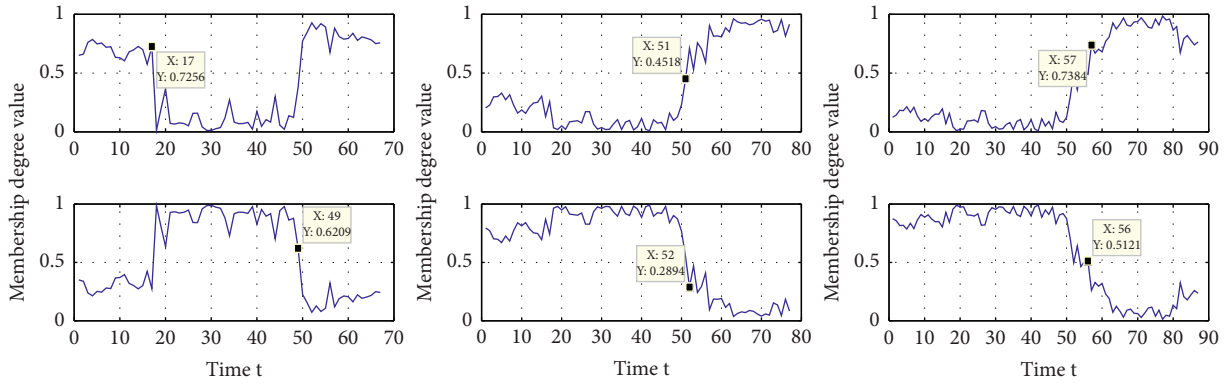


FIGURE 13: Clustering results with different training lengths.

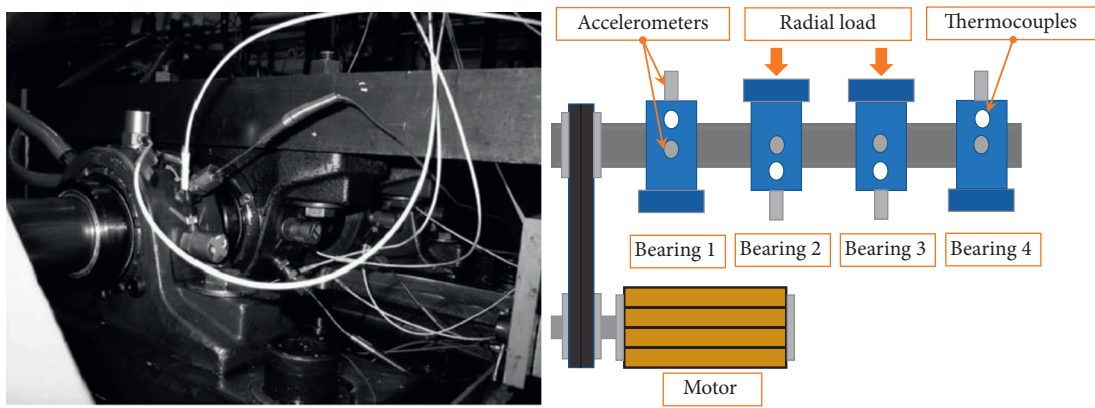


FIGURE 14: IMS test rig.

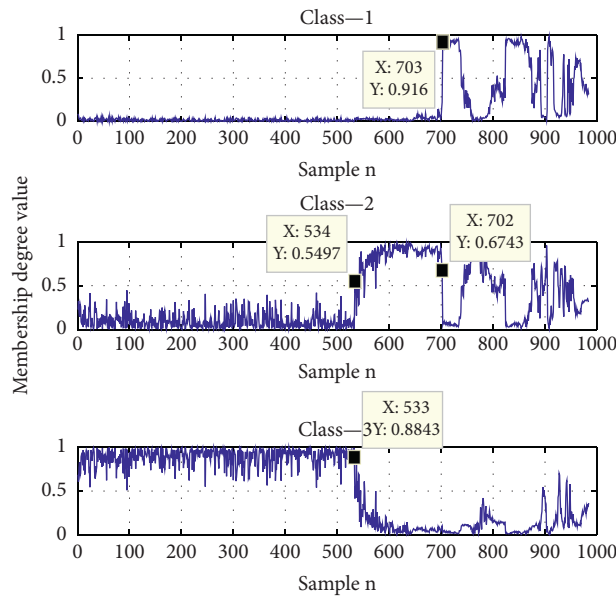


FIGURE 15: Clustering results of bearing 1.

degradation at the 534th sample and rapid degradation at the 703rd sample. With the turning points known, the linear regression algorithm is applied, and a novel RUL label appears in Figure 16.

It is noted that the RUL label at a training sequence length shorter than 703 is different from that in Figure 16, since the FCM algorithm has merely detected the first turning point. And then the RUL label is constructed

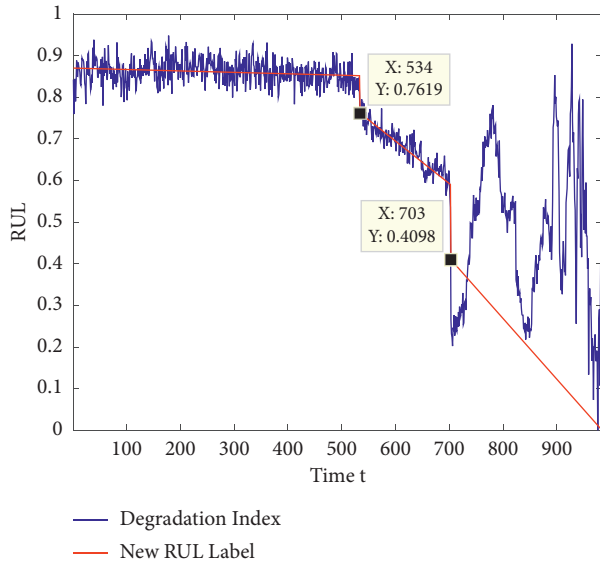


FIGURE 16: Novel RUL label of bearing 1.

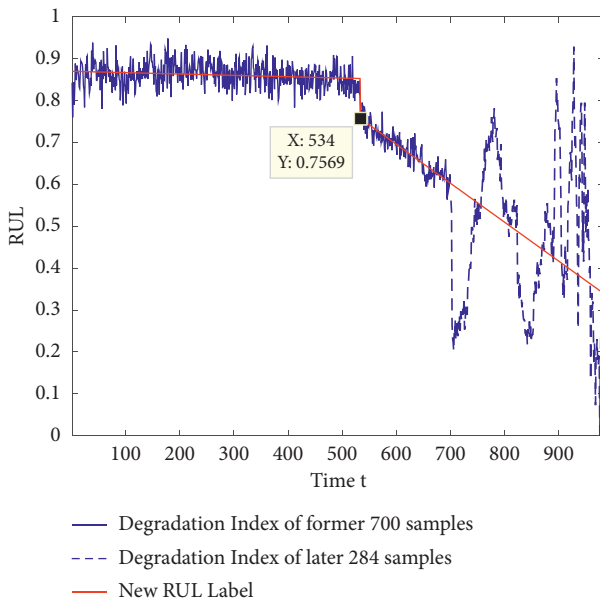


FIGURE 17: RUL label at a training length of 700.

through operation data before the next turning point. When there are 700 samples for model training, the RUL label is shown in Figure 17.

Four models of RNN, BP network, MLP, and SVR are compared. Here, the BP network with nodes 22-27-1 is constructed. The MLP with two hidden layers of 22-25-7-1 is obtained. A standard RNN structure is used with two hidden layers and the maximum iterations of 2000. In the SVR model, a polynomial is used as the kernel function with a cost value 1.5 and  $p$  value  $1e-4$  in loss function. These models run for 10 times, respectively, and compared in the best results.

Besides, linear and piecewise RUL functions are used for prediction, where the turning point of piecewise function is

TABLE 2: Evaluation results of three labels and four models (IMS).

Models	New RUL label		Linear RUL label		Piecewise RUL label	
	Score	RMSE	Score	RMSE	Score	RMSE
RNN	3.2199	0.6838	5.7834	2.9054	4.1641	1.8924
BP network	3.5863	3.0051	8.3087	5.1212	4.9563	3.2139
MLP	3.9781	3.5016	6.1778	5.3195	5.1935	4.7358
SVR	4.6046	1.2288	5.8675	5.9114	6.1047	3.1722

set at the 450th sample and the original RUL is 750. For evaluation, the RMSE and scoring function of different models and labels are calculated and shown in Table 2. Similar conclusions are easily achieved to the first experiment in XJTU-SY bearing data. The proposed label better agrees with the real RUL, and the RNN performs the best among four models.

## 5. Conclusions

An adaptive method of RUL training label construction based on the FCM algorithm is proposed in this paper, and the whole proposition is demonstrated by comparing the new constructed RUL label and the current linear RUL function and piecewise RUL function. Experiments carried out on one recent bearing degradation dataset, XJTU-SY bearing data, and a widely used dataset, IMS bearing data, strongly proved the superiority of the proposed method. The conclusions of this work are as follows:

- (1) The RUL label constructed by the proposed method better fits the hidden degradation mode than linear and piecewise RUL functions through a comparison of four common models for remaining useful life estimation. And the recurrent neural network performs the best in these models with its excellent capability of time-series data processing.
- (2) With the FCM algorithm for turning point detection, the fault of bearings is found synchronously, which provides a solution for online RUL prediction with no experience of failure threshold and ending time of life to use. Besides, it takes several samples for the algorithm to identify the turning points accurately.
- (3) The influence of the training sequence length is discussed in both experiments. As there exists more degradation information in the larger training set, the FCM algorithm shows higher accuracy on cluster and the RNN provides better performance on prediction.

The proposed method based on the FCM algorithm is proved to be powerful and accurate for RUL label construction, but the computing performance of the RNN model still needs to be improved. For example, the RUL can be predicted by the long short-term memory (LSTM) network, gated recurrent unit (GRU), bidirectional LSTM network, bidirectional GRU, transformer, and so on. And more improvements on algorithms should be developed. Besides, the linear degradation mode based on current

research studies is applied in this work. And nonlinear degradation curves could be used in the future research studies, as many degradation progresses are nonlinear in application.

As for further work, it is not idealistic to divide the bearing degradation process into several stages since it is commonly considered continuous and gradual. Another thing is that imbalance of degradation process division and rare failure data may lead to the poor performance of the assessment model. Thus, further work will pay attention to the continuous label of degradation process, and the imbalance of faulty data will also be considered.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Hailong Lin and Zihao Lei contributed equally to this work.

## Acknowledgments

This work was supported in part by the National Key R&D Program of China (no. 2020YFB1710002), in part by the National Natural Science Foundation of China (no. 51775409), and in part by the Equipment Pre-Research Fund of China (no. 61420030301).

## References

- [1] H. Zhou, J. Chen, G. Dong, and R. Wang, "Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 65-79, 2016.
- [2] Z. Huang, Z. Lei, G. Wen et al., "A multi-source dense adaptation adversarial network for fault diagnosis of machinery," *IEEE Transactions on Industrial Electronics*, 2021.
- [3] Z. Lei, G. Wen, S. Dong et al., "An intelligent fault diagnosis method based on domain adaptation and its application for bearings under polytropic working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-14, 2021.
- [4] S. Dong, G. Wen, Z. Lei, and Z. Zhang, "Transfer learning for bearing performance degradation assessment based on deep hierarchical features," *ISA Transactions*, vol. 108, pp. 343-355, 2021.
- [5] W. T. Sui, D. Zhang, X. M. Qiu, W. Zhang, and L. Yuan, "Prediction of bearing remaining useful life based on mutual information and support vector regression model," in *Proceedings of the 2019 the 5th International Conference on Electrical Engineering, Control And Robotics*, IOP Conference Series-Materials Science and Engineering, Guangzhou, China, 2019.
- [6] Y. Hu, H. Li, P. Shi et al., "A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process," *Renewable Energy*, vol. 127, pp. 452-460, 2018.
- [7] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7762-7773, 2015.
- [8] D. Wang and K.-L. Tsui, "Two novel mixed effects models for prognostics of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 1-13, 2018.
- [9] K. Javed, R. Gouriveau, and N. Zerhouni, "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mechanical Systems and Signal Processing*, vol. 94, pp. 214-236, 2017.
- [10] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799-834, 2018.
- [11] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 88-95, IEEE, Dallas, TX, USA, 19 June 2017.
- [12] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401-412, 2020.
- [13] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proceedings of the 2008 International Conference on Prognostics and Health Management*, pp. 59-64, IEEE, Denver, CO, USA, 6 October 2008.
- [14] G. S. Babu, P. Zhao, and X.-L. Li, Eds., in *Proceedings of the International conference on database systems for advanced applications*, Springer, Cham, 25 March 2016.
- [15] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694-700, 2004.
- [16] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98-109, 2017.
- [17] W. Mao, J. He, J. Tang, and Y. Li, "Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network," *Advances in Mechanical Engineering*, vol. 10, no. 12, 2018.
- [18] Y. Zhang, R. Xiong, H. He, and Z. Liu, Eds., in *Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin)*, IEEE, Harbin, China, 9 July 2017.
- [19] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng, and X. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 529-539, 2018.
- [20] M. Yuan, Y. Wu, and L. Lin, Eds., in *Proceedings of the 2016 IEEE international conference on aircraft utility systems (AUS)*, IEEE, Beijing, China, 10 October 2016.
- [21] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Transactions on industrial informatics*, vol. 12, no. 3, pp. 924-932, 2016.
- [22] J. Wu, C. Wu, Y. Lv, C. Deng, and X. Shao, "Design a degradation condition monitoring system scheme for rolling

- bearing using EMD and PCA,” *Industrial Management & Data Systems*, vol. 117, 2017.
- [23] J. Zhang, P. Wang, R. Yan, and R. X. Gao, “Long short-term memory for machine remaining life prediction,” *Journal of Manufacturing Systems*, vol. 48, pp. 78–86, 2018.
- [24] L. Cao, Z. Qian, and Y. Pei, Eds., in *Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, IEEE, Chongqing, China, 26 October 2018.
- [25] Y. Zhang, B. Tang, Y. Han, and L. Deng, “Bearing performance degradation assessment based on time-frequency code features and SOM network,” *Measurement Science and Technology*, vol. 28, no. 4, Article ID 045601, 2017.
- [26] T. Gangavarapu, A. Jayasimha, G. S. Krishnan, and S. Kamath, “Predicting ICD-9 code groups with fuzzy similarity based supervised multi-label classification of unstructured clinical nursing notes,” *Knowledge-Based Systems*, vol. 190, Article ID 105321, 2020.
- [27] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, Butterworth-Heinemann, Oxford, UK, 2016.
- [28] X. Zhao and M. Jia, “A novel deep fuzzy clustering neural network model and its application in rolling bearing fault recognition,” *Measurement Science and Technology*, vol. 29, no. 12, Article ID 125005, 2018.