Research Article

Remaining Useful Life Prediction of Rolling Bearings Based on Multiscale Convolutional Neural Network with Integrated Dilated Convolution Blocks

Ran Wang [iD],[1] **Ruyu Shi,**[1] **Xiong Hu,**[1] **and Changqing Shen** [iD][2]

[1]*College of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China*
[2]*School of Rail Transportation, Soochow University, Suzhou, Jiangsu 215131, China*

Correspondence should be addressed to Changqing Shen; cqshen@suda.edu.cn

Remaining useful life (RUL) prediction is necessary for guaranteeing machinery's safe operation. Among deep learning architectures, convolutional neural network (CNN) has shown achievements in RUL prediction because of its strong ability in representation learning. Features from different receptive fields extracted by different sizes of convolution kernels can provide complete information for prognosis. The single size convolution kernel in traditional CNN is difficult to learn comprehensive information from complex signals. Besides, the ability to learn local and global features synchronously is limited to conventional CNN. Thus, a multiscale convolutional neural network (MS-CNN) is introduced to overcome these aforementioned problems. Convolution filters with different dilation rates are integrated to form a dilated convolution block, which can learn features in different receptive fields. Then, several stacked integrated dilated convolution blocks in different depths are concatenated to extract local and global features. The effectiveness of the proposed method is verified by a bearing dataset prepared from the PRONOSTIA platform. The results turn out that the proposed MS-CNN has higher prediction accuracy than many other deep learning-based RUL methods.

1. Introduction

Prognostics and health management (PHM) are crucial for the mechanical system. RUL prediction is one of the important tasks in modern industry PHM. Maintenance costs can be reduced if the remaining useful life of the machinery can be known in advance. Bearings are the critical parts of the mechanical system [1, 2]. The failure of bearings may lead to a severe accident. Thus, the bearings RUL prediction has drawn more and more attention in the study of PHM.

The bearing RUL prediction methods can be roughly divided into two types: model-based approaches and data-driven approaches [3]. With the development of modern industrial technology, an enormous amount of condition monitoring data are recorded, data-driven methods such as deep learning have the powerful data-processing ability when faced with massive data [4]. Since DL-based

approaches can extract features from the input data without much prior knowledge, they have become more and more popular in RUL prediction and fault diagnosis [5]. A prediction framework constituted by deep autoencoders (AE) is proposed in Reference [6]. AE is used to retain sufficient information when compressing features. Shen et al. [7] proposed a contractive autoencoder-based rotating machinery fault diagnosis method. Robust features can be learned by the contractive autoencoder automatically. A deep long short-term memory (DLSTM) network was proposed in Ref. [8]. Multisensor condition monitoring data are fused to get more useful information for accurate RUL prediction. Long short-term memory (LSTM) is also applied in discovering the potential patterns in Ref. [9]. Xiang et al. [10] proposed a novel LSTM framework. Attention-guided ordered neurons are applied in this framework to achieve the accurate gear remaining useful life prediction. A double

CNN model architecture was implemented in Ref. [11]. The fault occurrence time (FOT) is determined by the first CNN; RUL prediction is accomplished on the second CNN. Guo et al. [12] proposed a health indicator construction method to monitor the health state of machinery. Convolutional neural network is designed to learn features and construct a mapping from HI between features. A recurrent convolutional neural network was designed in Ref. [13]. The temporal dependencies of different degradation states can be captured by the recurrent convolutional layers.

Among the variety of DL techniques, CNN has gained more attention because of two outstanding characteristics, i.e., spatially shared weights and local perception. At first, CNN was widely used in image recognition and achieved tremendous success. Nowadays, it is also popular in fault diagnosis [14] because it can accomplish feature extraction and fault classification automatically. However, there are still two shortcomings in traditional CNN. (1) The scales of the convolution kernel are very important for the performance of CNN. Kernels with a bigger size can extract features in a bigger receptive field, whereas kernels with a smaller size extract feature in a smaller receptive field. The performance of the network is directly affected by the scales of the convolution kernel. Single size of the convolution kernels can lead to the extracted information incompletely. (2) The ability of traditional CNN is not strong enough to extract local and global features simultaneously. In the conventional CNN, just the feature maps of the final layer before flatten are treated as final features while features extracted in previous low-level layers are omitted. Although the global features extracted by the high-level layers are more invariant than features extracted by the low-level layers, detailed local features extracted by the low-levels are also contribute to prognosis and classification [15, 16].

In order to overcome the aforementioned problems and learn more representative features, various multiscale CNN models have been proposed and applied in machinery fault diagnosis and RUL prediction [17]. The final convolutional layer and the final pooling layer are merged to construct a multiscale layer in Ref. [18]. Global features from the final pooling layer and local features from the final convolutional layer are utilized for classification. The results show that the network with the proposed multiscale layer can improve the recognition accuracy. Chang et al. [19] were inspired by the inception model and proposed a concurrent convolution neural network method to enhance wind turbine fault diagnosis accuracy. Also, some multiscale CNN methods have also been applied in RUL prediction. A multiscale deep convolutional neural network (MS-DCNN) with three MS-blocks was applied in Ref. [20]. In order to extract features in different receptive fields, three distinct sizes of convolution kernels are implemented in each layer parallelly. Kernels with a bigger size can extract features in a bigger receptive field, but it also leads to more weights of convolution kernels to train. Too many weights are generally hard to train for the network. Different moving steps in convolution operations are used to obtain features in different scales [21]. A multiscale convolutional neural network that merges the final convolutional layers and the final pooling layer was designed

to extract the local and global features for bearing RUL prediction [15]. But the detailed information learned by the low layers is still lost in this architecture.

In this paper, a novel MS-CNN method for bearing RUL prediction is introduced. An integrated dilated convolution block is constructed to extract features in different receptive fields from the complex signal. Then, several stacked integrated dilated convolution blocks are concatenated to construct a multiscale feature extractor. The two advantages of the proposed MS-CNN are summarized as follows:

(1) An integrated dilated convolution block is constructed to extract features in different receptive fields from complex signals without increasing the weight of convolution kernels.

(2) A multiscale feature extractor is constructed to avoid the loss of information at different levels. The multiscale feature extractor can make full use of the global features obtained from the higher layers and the local features obtained from the lower layers.

The rest of this paper is introduced as follows: in Section 2, the relevant theoretical backgrounds are introduced, including convolutional neural networks and dilated convolution. The proposed MS-CNN bearing remaining useful life prediction method is introduced in Section 3. In Section 4, a public dataset about rolling bearings is used to verify the superiority of the proposed method. The dataset is collected from the PRONOSTIA platform. Finally, the conclusions are summarized in Section 5.

2. Theoretical Background

2.1. Convolutional Neural Network. CNN is a feed-forward neural network. The traditional CNN is mainly composed of three kinds of layers: the convolutional layers, the pooling layers, and the fully connected layers [22].

In the convolutional layers, features can be learned by several convolution kernels. The convolutional operation is a linear operation. In order to increase the nonlinear of two adjacent layers, the nonlinear activation function is carried out to solve the problem. The output of the convolutional layer can be written as

$$Z_j^r = \varphi \left(\sum_i Z_i^{r-1} * w_j^r + b_j^r \right), \tag{1}$$

where $*$ denotes the convolution operation, Z_j^r is the jth output of layer r, Z_i^{r-1} is the ith input from the previous $r-1$ layer, w_j^r is the weight vector of convolution kernel in rth layer, and b_j^r is the bias of the jth output. $\varphi(\cdot)$ is the nonlinear activation function.

In the pooling layer, the output of the convolutional layers can be compressed to improve computational efficiency [14], which is a form of down-sampling. The pooling operation includes many pooling functions, such as L^2 norm pooling, max pooling, and average pooling. Pooling operation is adopted as

$$Z_m^n(i) = \text{pool}\left\{Z_m^{n-1}(igp + s)\right\}, \qquad (2)$$

where Z_m^{n-1} is the mth input feature map in $n-1$ layer, Z_m^n is the mth output in the layer n, p is the pooling size and s denotes the stride, and pool ( ) denotes pooling function.

After several convolutional and pooling operations, the feature maps of the previous layer are flattened and then sent to the fully connected layer. RUL prediction is a regression task. Thus, the fully connected layers can be regarded as the regression layer, which builds a connection from the feature maps learned by previous layers to the final result.

2.2. Dilated Convolution.

Dilated convolution (atrous convolution) can enlarge the receptive field without increasing network parameters [23]. Recently, dilated convolution has been implemented in many areas such as bearing fault diagnosis [24], semantic image segmentation [25], and sound classification [26].

Compared with the ordinary convolution operation, a hyperparameter called dilation rate is added to the dilated convolution. As shown in Figure 1, different dilation rates can be seen as inserting different sizes of holes between each convolution kernel parameter. When applied in one-dimension CNN, it can be calculated as

$$y_i = \varphi\left(\sum_{k=1}^{K} x_{i+r\cdot k}\omega_k + b\right), \qquad (3)$$

where y_i represents the output of the ith element in the convolution, x_i is the ith element of input, ω are the weights of the filters, and the length of the filter is K. r is the dilation rate, $r = 1$ in dilated convolution is equal to the ordinary convolution. One zero is inserting in the adjacent convolution weight when the dilation rate $= 2$.

3. Proposed Bearing RUL Method

In this section, the framework of the proposed multiscale convolutional neural network is introduced in detail. Machinery vibration signals are sent to the network directly as the input data. The procedure of the proposed MS-CNN-based RUL prediction method is introduced lately. Four steps are implemented to get the bearing remaining useful life prediction: FOT determination, data preprocessing, RUL prediction, and smoothing operation. The proposed MS-CNN could establish a relationship between monitoring signal and remaining useful life without much prior knowledge. It is easy to be generalized in industrial applications.

3.1. Proposed MS-CNN Architecture.

The framework of the multiscale convolutional neural network is shown in Figure 2. The MS-CNN consists of two modules, a multiscale feature extractor and the regression layer. The multiscale feature extractor is constructed to improve the network's learning ability. Features in different receptive fields are taken into consideration by an integrated dilated convolution block. Then, the integrated dilated convolution blocks in different levels are all concatenated to establish the proposed multiscale feature extractor. The regression layer is designed to construct the mapping relationship between features and corresponding real lifetime.

3.1.1. The Integrated Dilated Convolution Block.

From the structure of the traditional CNN, features are extracted by the convolutional operation and pooling operation. The scale of the convolution kernel can affect the learning ability of the network. A single-sized convolution kernel in each convolutional layer may lead to the information learned by this layer incompletely. Inspired by the inception network, an integrated dilated convolution block is introduced, which is shown in Figure 3.

Different convolution kernels with different dilation rates are concatenated to learning multiscale features in different receptive fields in this paper. Compared with the convolution with different sizes of convolution kernels. The integrated dilated convolution block can learn different scales of information without increasing the parameter of the network. Too large dilated rate may cause the loss of detailed information. Too many dilated rates integrated into one block may lead to redundant information. Taking consideration of the characteristic of the vibration signal and inspired by inception networks, the structure of the integrated dilated convolution block has three kinds of dilated rates.

When the input data are sent to the integrated dilated convolution block, three kinds of convolution operations are performed on the input data synchronously. Features extracted from n filters are then concatenated into a features vector. The features vector can be recorded as follows:

$$g^l(x) = \left[g_1^l(x), g_2^l(x), \ldots, g_n^l(x)\right], \qquad (4)$$

$g_i^l(x)$ denotes the feature map learned by different convolution kernels in the lth layers.

3.1.2. The Proposed MS-CNN.

In the traditional CNN structure, the feature maps of the last pooling layer are treated as final features for classification or regression. Local features extracted by previous layers are usually discarded. Although the global features extracted by high-level layers are much more representative and robust than those extracted by low-level layers, local features contain some detailed information and are useful for prognosis.

In this paper, the outputs of different integrated dilated convolution blocks are concatenated to extract local and global features synchronously. The proposed MS-CNN is shown in Figure 2. The concatenated feature maps can contain not only the invariant and stable global feature but also the detailed information. The concatenated feature maps can be expressed as

$$G = \left\{g^1 \oplus g^2 \oplus \cdots \oplus g^l\right\}, \qquad (5)$$

where the $\oplus$ denotes the concatenation operation and g^l indicates the feature maps of the lth integrated dilated convolution block.
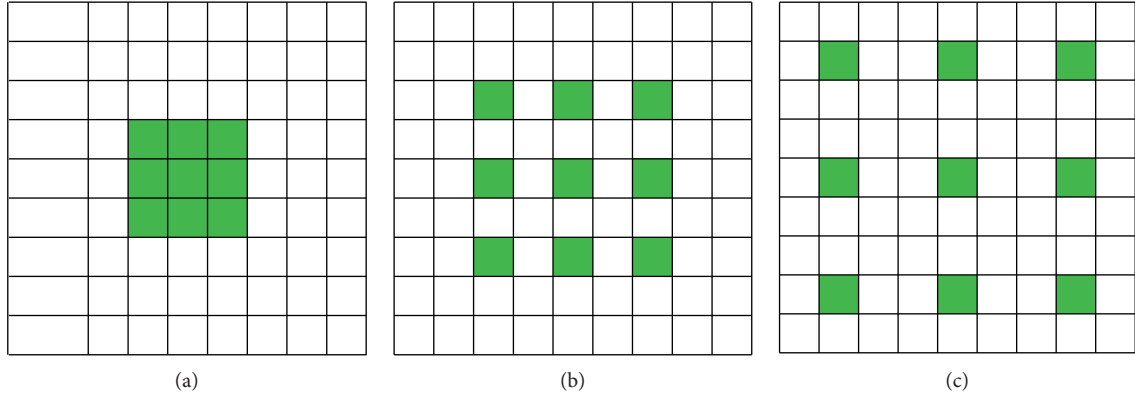
FIGURE 1: Different dilation rates in dilated convolution: (a) dilation rate = 1, (b) dilation rate = 2, and (c) dilation rate = 3.
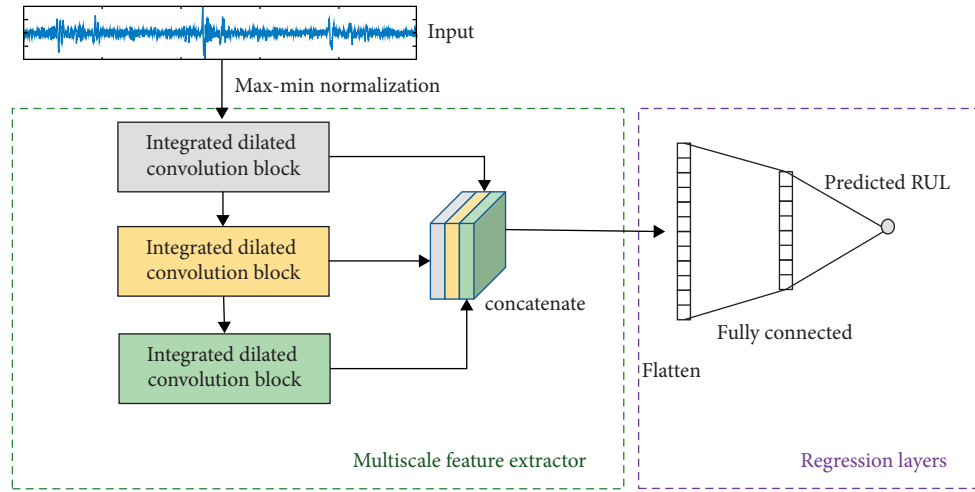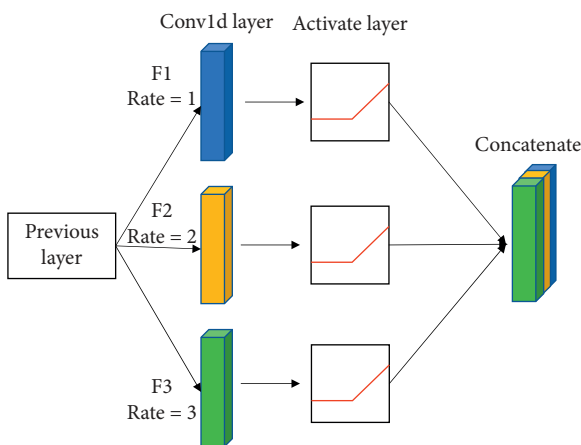

FIGURE 2: The proposed MS-CNN.

FIGURE 3: The integrated dilated convolution block.

To keep the RUL prediction ranging from 0 to 1, the sigmoid activation function is applied in the last fully connected layer. The sigmoid activation function can be expressed as

$$S(x) = \frac{1}{1 + e^{-x}}, \tag{6}$$

where the X represents the input feature map of the final fully connected layer.

MSE loss function is utilized to update the parameters of the whole network, which is expressed as

$$\text{loss}_{\text{MSE}} = \frac{1}{N} \sum_{i=0}^{N-1} (y_i - f(x_i))^2, \tag{7}$$

where N denotes the number of the training samples and $f(x_i)$ denotes the predicted RUL value of the input x_i, y_i is the true label.

3.2. Framework of the Proposed Bearing RUL Method. The flow chart of the proposed method is displayed in Figure 4. Firstly, signals of machinery are collected by sensors, then the fault occurrence time (FOT) is determined according to the collected signals. Then, the data from FOT to failure are divided into two categories: training samples and testing samples. The training samples are applied to train the
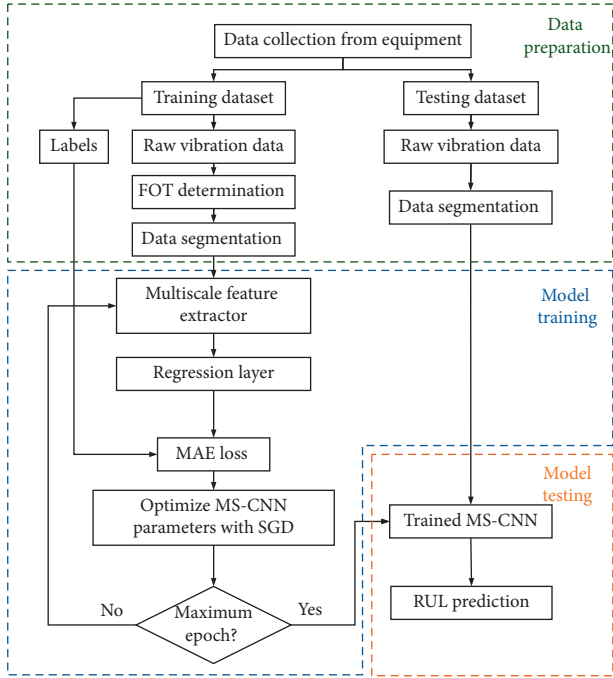
FIGURE 4: The flow chart of the proposed method.

FIGURE 5: The FOT determined by the applied method.

network. The superiority of the MS-CNN is confirmed by testing samples. At last, a smoothing operation is implemented to get a continuous RUL result.

3.2.1. Step 1: FOT Determination. As shown in Figure 5, in the early stage, vibration signals recorded by the sensors usually undergo a stable period, which means the monitored component is in a healthy stage. Data in the healthy stage include unrelated information to RUL prediction. Thus, the determination of FOT before RUL prediction is necessary. In this paper, kurtosis is applied to find the FOT. The formula of kurtosis is as the following equation:

$$k_u = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \overline{x})^4}{\sigma^4}, \tag{8}$$

where $\overline{x}$ is the mean and standard deviation of the signal, σ is the standard deviation of the signal, and N is the number of signal data.

Kurtosis is very sensitive to the amplitude value. The degradation can be well reflected by kurtosis. Thus, it is usually very helpful for detecting incipient faults [27].

Laida criterion (also known as 3σ rules) is applied to detect the FOT. We assumed that the signal in the early period is in the health stage. The mean μ and the standard deviation σ of kurtosis in the health stage are calculated, and then, 3σ is used as the FOT indicator.

Although it is very effective to regard kurtosis as health index to detect early fault points, this health index is not stable enough because it is affected by noise and outliers. Thus, local linear regression is applied after calculating kurtosis from the vibration signal, which can remove the FOT misjudgment caused by outliers. After local linear
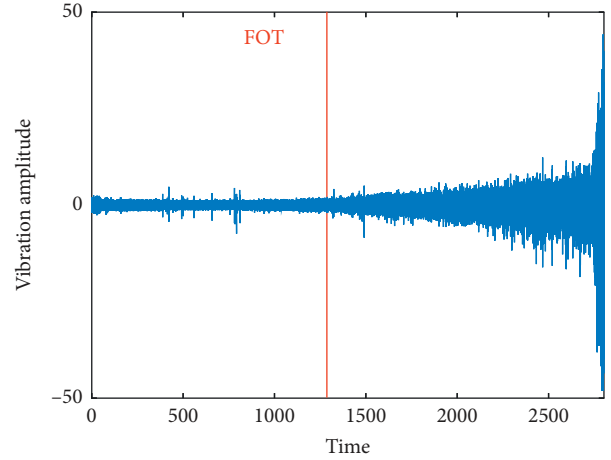
regression smoothing operation, when the smoothed kurtosis of the time t falls out of the 3σ interval, it can be regarded as the FOT as follows:

$$\left| k_u^i - \mu \right| > 3\sigma, \tag{9}$$

where k_u^i denotes the kurtosis of the ith data sequence. The FOT of bearing1_1 determined by the applied method is shown in Figure 5. The amplitude of the vibration signal keeps in a stable stage before FOT and degradation starts after FOT. The method of FOT determination is proved to be effective and accurate. FOT determination is only determined in the training processing. In the online testing processing, lacking of the whole lifetime vibration signal leads to the FOT determination impossible.

3.2.2. Step 2: Data Preprocessing. To speed up the training convergence, normalizing the raw data is a common and effective operation. Firstly, the selected sensor data is divided into time segmentation, and the length of the segmentation is L. Each time segmentation is a sample that can be represented as $X = [x_1, x_2, x_3, \ldots, x_L]$. Max-min normalization is implemented to ensure the data within [0, 1] as the following equation:

$$\widehat{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \tag{10}$$

where the x_{max} is the max value of the sample X. x_{min} is min value of the sample X. $\widehat{x}_i$ represents the value of x_i after max-min normalization.

The label of the training data is constructed as the reliability in the range of [0, 1]. FOT is regarded as the start of degradation. The label can be described as the linearly degrading process from FOT to complete failure, as shown in Figure 6. Then, the samples and the corresponding labels are treated as the input and the output of the proposed MS-CNN.

3.2.3. Step 3: RUL Prediction Based on the Proposed MS-CNN. The proposed method based on MS-CNN has two processes: offline training and online testing. In the training
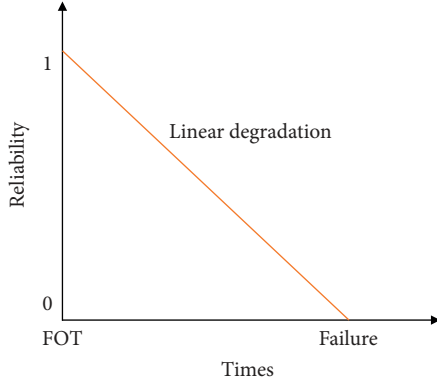
FIGURE 6: The label of the training sample.

period, the training dataset and corresponding label are used to train the proposed MS-CNN. After data preparation, training data segmentation undergoes a multiscale feature extractor. More representative and comprehensive features can be learned by the multiscale feature extractor. The output of the multiscale feature extractor is sent to the regression layer to construct the relationship between features and RUL. MSE loss function is applied in the network. Small training samples can lead to overfitting. The network has bad performance on the testing dataset. Dropout is adopted in the fully connected layers to avoid overfitting by setting some hidden neurons to zero and turned off in the testing process. Adam algorithm is applied to update the parameters of the network. Different from the stochastic gradient descent algorithm, the Adam algorithm can adjust the learning rate adaptively without setting the learning rate in advance. When the online testing data at a moment are sent to the network, the prognostic RUL of that testing data can be predicted by the trained network.

In order to assess the predicted result quantitatively, two error indicators are applied in this paper, i.e., the mean absolute error (MAE) and root mean squared error (RMSE) [28].

$$
\begin{aligned}
\text{MAE} &= \frac{\sum_{i=1}^{N} \left| L_i - E_i \right|}{N}, \\
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{N} \left(L_i - E_i \right)^2}{N}},
\end{aligned}
\tag{11}
$$

where N denotes the number of the samples, E_i is the corresponding RUL of the ith sample predicted by the proposed MS-CNN, L_i is the corresponding actual RUL.

3.2.4. Step 4: Smoothing. The RUL results predicted by the network are discrete and fluctuant. However, in the real industrial applications, the actual RUL of bearing is always continuous. The remaining useful life of the bearing decreases as time goes by. Thus, the smoothing operation is applied to smooth the predicted RUL.

At the time n, the prediction RUL of this moment is recorded as RUL(n), the RUL at five time-point moments before n such as $n-5, n-4, \ldots, n-1$, and n undergo a

moving average filter. The smoothing operation can be implemented as equation (12). The regression RUL result at the time n is regarded as the final predicted result. The smoothing operation can make the predicted result according to the actual condition.

$$
R(n) = \begin{cases} \dfrac{1}{n} \left(R(n) + \cdots + R(2) + R(1) \right), & \text{if } n < 5, \\[3mm] \dfrac{1}{5} \left(R(n-4) + R(n-3) + \cdots + R(n) \right), & \text{otherwise.} \end{cases}
\tag{12}
$$

4. Experiment

In this section, experiments are carried out to verify the superiority of the proposed MS-CNN method. The experimental dataset was acquired from the PRONOSTIA platform. The dataset is introduced in Section 4.1. In Section 4.2, the parameters of the proposed MS-CNN are displayed. What is more, the results of the proposed method on experimental data are displayed. Different DL-based bearing remaining useful life prediction methods are compared in Section 4.3. A PC with Intel Core i7-5557U CPU, 4-GB RAM is used in all the experiments' implementation. All the results in this study are tested ten times to reduce random errors.

4.1. Data Description. The PRONOSTIA platform (as shown in Figure 7) performed the accelerated degradation test of the rolling bearing. The platform is mainly divided into three parts: the rotating part, the degradation part, and the measurement part. Degradation of the bearings can be accelerated by the degradation part. In order to measure the vibration signal of bearings, two acceleration sensors were installed on the vertical and horizontal axis. However, the amplitude of the vertical vibration signals is lower than the horizontal ones. The degradation trend was better captured by the sensors placed on the horizontal axis. Therefore, only the horizontal vibration signals are used in the paper. The frequency of sampling in the experiment is 25.6 kHz [29]. A sample contains the data collected in 0.1 s every 10 s. 17 run-to-failure bearings were acquired. The first two bearings of each working condition were applied for the training process, and the rest of the bearings were used as the testing data. All the datasets are shown in Table 1.

4.2. Proposed Method on Experimental Data. Row vibration signal of each run-to-failure bearing is divided into time-series segmentation. Each segmentation is a sample, and it contains 2560 data points. After the determination of the FOT, each sample is normalized by the max-min normalization method as described in Section 3.2.2. Training samples with corresponding labels were utilized for training the MS-CNN.

The parameters of the MS-CNN are displayed in Table 2. KS represents the size of the convolution kernel, r represents the dilation rate, n is the number of filters, and s represents the stride. N represents the number of neurons in the fully
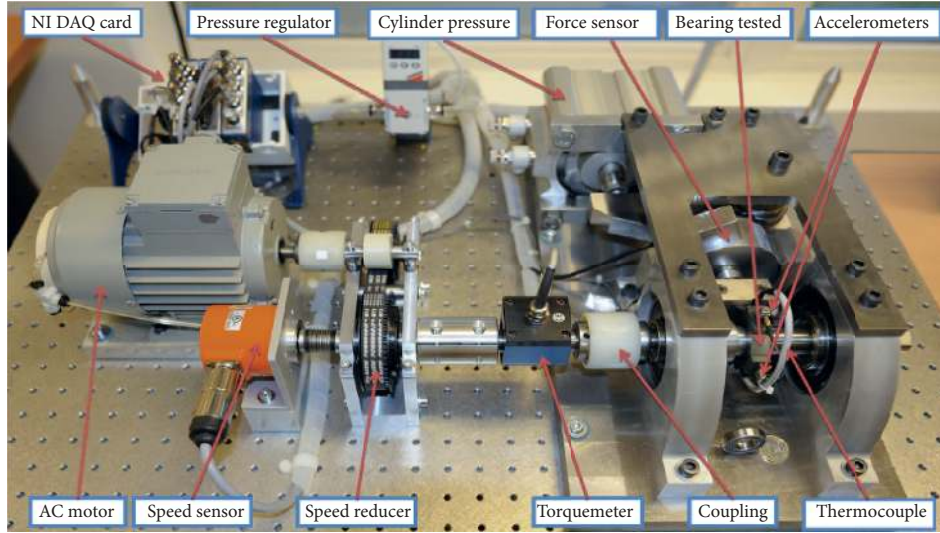
FIGURE 7: The PRONOSTIA platform for bearing accelerated degradation tests.

TABLE 1: Datasets of IEEE PHM 2012 prognostic challenge.

Data set	Operation conditions		
	Condition 1	Condition 2	Condition 3
Load (N)	4000	4200	5000
Speed (rpm)	1800	1650	1500
Training set	Bearing 1_1	Bearing 2_1	Bearing 3_1
	Bearing 1_2	Bearing 2_2	Bearing 3_2
Test set	Bearing 1_3	Bearing 2_3	Bearing 3_3
	Bearing 1_4	Bearing 2_4	
	Bearing 1_5	Bearing 2_5	
	Bearing 1_6	Bearing 2_6	
	Bearing 1_7	Bearing 2_7	

TABLE 2: Parameters of MS-CNN.

Layer	Parameter
1st fused layer	$KS = 16 \times 1; \; r = 1; \; n = 4; \; s = 1$
	$KS = 16 \times 1; \; r = 2; \; n = 4; \; s = 1$
	$KS = 16 \times 1; \; r = 3; \; n = 4; \; s = 1$
2nd fused layer	$KS = 8 \times 1; \; r = 1; \; n = 8; \; s = 1$
	$KS = 8 \times 1; \; r = 2; \; n = 8; \; s = 1$
	$KS = 8 \times 1; \; r = 3; \; n = 8; \; s = 1$
3rd fused layer	$KS = 4 \times 1; \; r = 1; \; n = 16; \; s = 1$
	$KS = 4 \times 1; \; r = 2; \; n = 16; \; s = 1$
	$KS = 4 \times 1; \; r = 3; \; n = 16; \; s = 1$
1st FC layer	$N = 256$; action function: ReLU
2nd FC layer	$N = 64$; action function: ReLU
3rd FC layer	$N = 1$; action function: sigmoid

connected layer. The dropout strategy is adopted to avoid overfitting with a coefficient of 0.2. MSE is employed as the loss function. Adam optimization algorithm is applied to update the parameters of the model in the back-propagation process.

The loss function curve for the training is shown in Figure 8. The training loss curve declined rapidly from the beginning to the 20th epoch. Then, keeping a slow decline
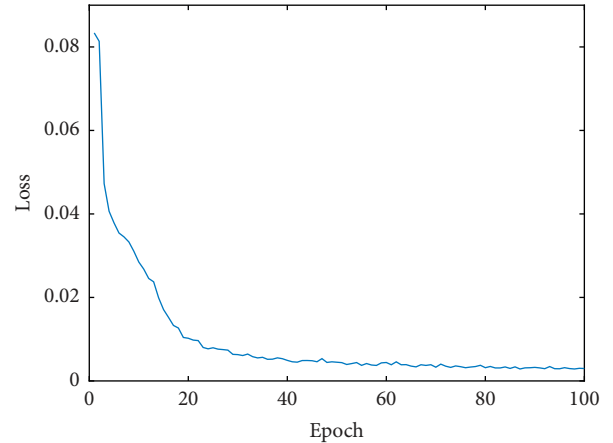

FIGURE 8: Loss function curve for the training set.

trend among the 20th epoch and 80th epoch, the loss curve is stable from the 80th epoch to the 100th epoch. Thus, the epoch of the network is determined as 80.

The performance of the network is related to depth. Effects on the number of integrated dilated convolution blocks are discussed in this section. As shown in Figure 9, the performance of the network is improved when the number of fused dilation convolution layer increases at first. When the number of the integrated dilated convolution blocks exceeds three, the MAE of the result increased. Too many integrated dilated convolution blocks may lead to an overfitting problem, the training data will have high performance, but the testing data will have a bad performance. On the other hand, as the depth of the network increases, the training time can be longer. Thus, in this study, the number of fused layers is determined to 3.

The row vibration signal of bearing1_3 is shown in Figure 10. Bearing1_3 shows a failure behavior in a gradual degradation trend. The remaining useful life prediction result of bearing 1_3 is shown in Figure 11. The yellow line in
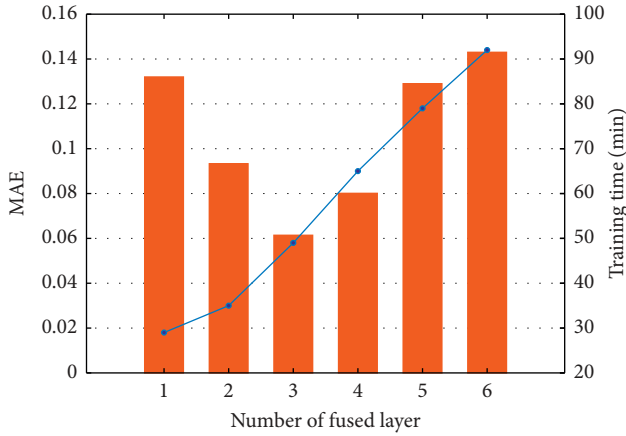
FIGURE 9: Effect of the number of fused dilated convolution layer on the bearing1_3.
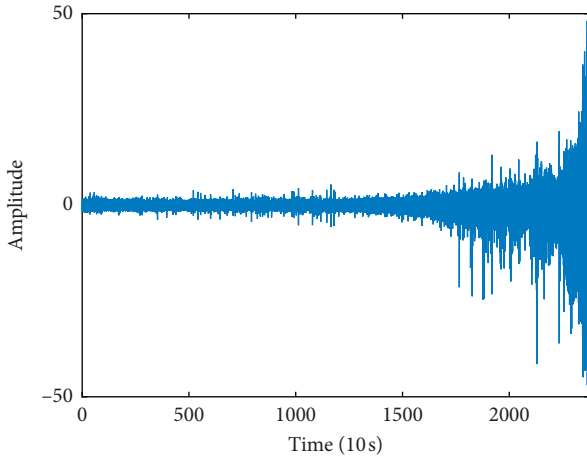

FIGURE 10: Vibration signal of bearing 1_3.

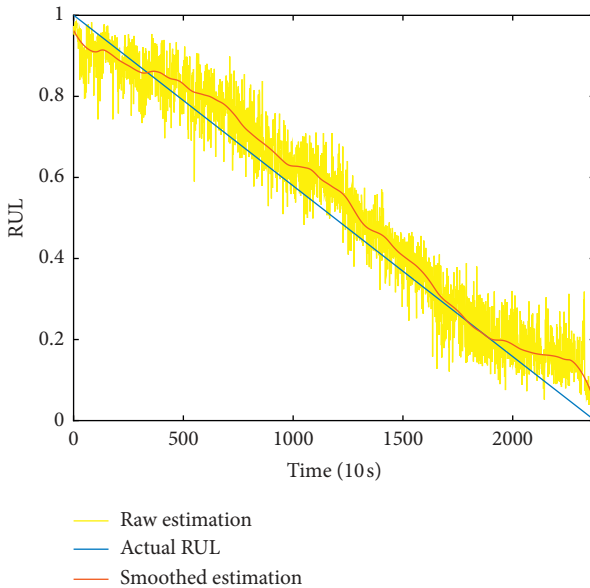Raw estimation
Actual RUL
Smoothed estimation

FIGURE 11: RUL estimation result of bearing 1_3.

the figure is the row estimation result without smoothing operation. The row estimation result is discontinuous and fluctuates in a larger range. The red line is the estimation result with smoothing operation. Smoothed estimation shows a steady and continuous RUL result of the testing data, which is consistent with the actual RUL. The row vibration signal of bearing1_7 is shown in Figure 12. Bearing1_7 shows a sudden failure behavior. RUL prediction result of bearing 1_7 is shown in Figure 13. Although the estimation result is not completely in line with the actual RUL in the early stage, degradation can be effectively reflected in the near-failure stage. Bearings in different failure behaviors were used to turn out the superiority of the proposed MS-CNN method.

4.3. Comparison Results. Several commonly used deep learning models are used for comparison, including deep neural network (DNN), convolutional neural network (CNN), long short-term memory (LSTM), and multiscale convolutional neural network which merge the final convolutional layers with the final pooling layer [15]. All the comparative methods have tuned the parameters to optimal values relatively.

4.3.1. DNN. Six hidden layers were utilized in the deep neural network to share the same depth as the proposed MS-CNN. And the number of neurons in each layer was 300, 200, 200, 100, 100, and 1. Dropout was applied in each layer and the dropout rate of 0.2 is used.

4.3.2. CNN. A traditional CNN was applied to compare with the proposed MS-CNN. CNN has three convolutional layers and three pooling layers. The kernel size of the three convolutional layers were 16×1, 8×1, and 4×1. The number of the kernels was 4, 8, and 16. Three fully connected layers were implemented to stay the same as the proposed MS-CNN.

4.3.3. LSTM. LSTM is one of the variant algorithms of recurrent neural networks. LSTM framework was designed to contain three LSTM layers and three fully connected layers. The number of neurons in LSTM layers was 128, 64, and 64. The three fully connected layers were the same as the proposed method.

4.3.4. MS-CNN. The third convolution layer and the third pooling layer were concatenated to form a fused layer as proposed in Ref. [15]. The kernel size and the number of the kernel are the same as the CNN structure. After the fused layer, three fully connected layers were constructed the same as the proposed MS-CNN.

The prognostic result of different methods for testing data bearing1_3 is shown in Figure 14. The degradation can be reflected by the proposed method. The predicted RUL of bearing1_3 by the proposed method is the closest to real-life than those other approaches. The DNN method shows the
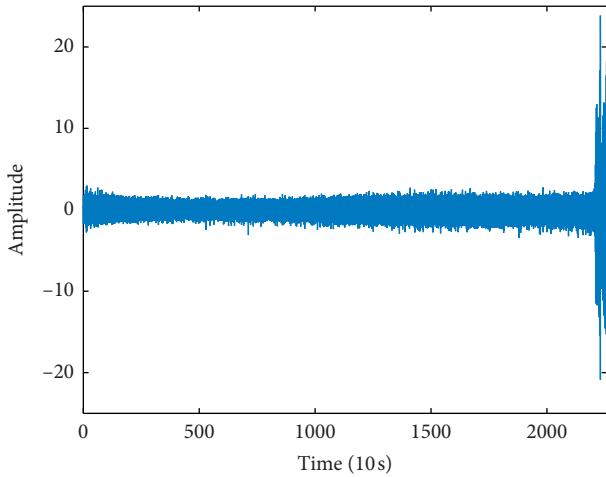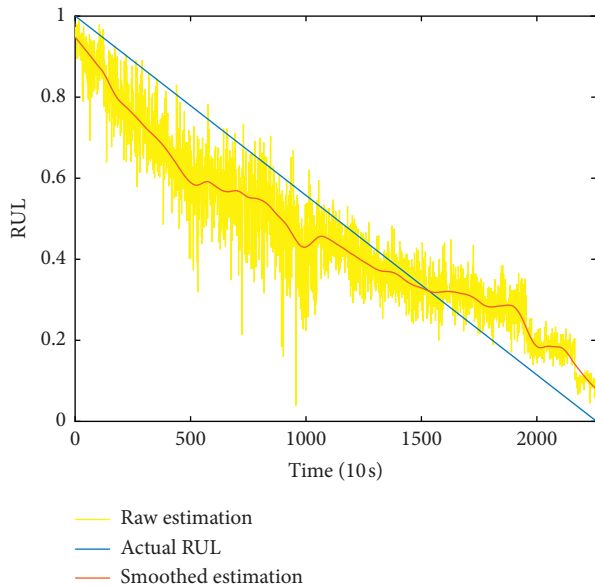
FIGURE 12: Vibration signal of bearing 1_7.

FIGURE 14: RUL estimation result of bearing 1_3 with different methods.

FIGURE 13: RUL estimation result of bearing 1_7.
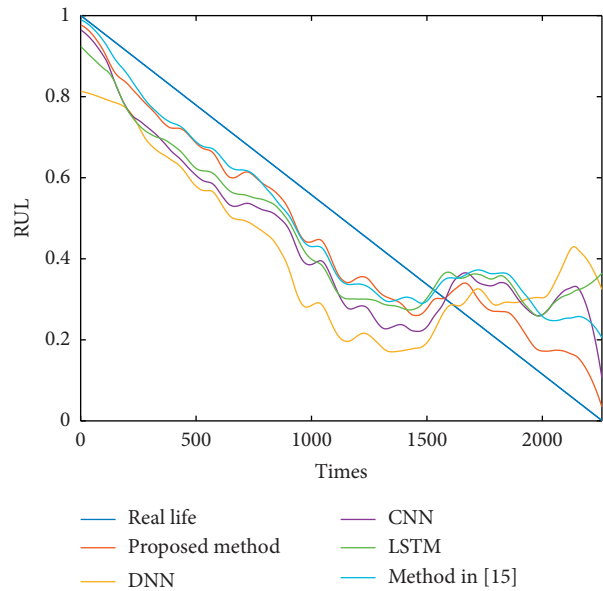
FIGURE 15: RUL estimation result of bearing 1_7 with different methods.

worst result among the comparing methods. The prognostic result of different methods for testing data bearing1_7 is shown in Figure 15. The RUL of bearing1-7 predicted by the proposed MS-CNN is not consistent with the actual RUL in the early stage. That is because the signal in the early stage shows a stable state. In the late stage, degradation can be reflected by the proposed MS-CNN. Since the accurate RUL prediction in the near-failure stage is more important in real industries, the proposed MS-CNN is promising in real industrial implementation.

The numerical comparison result of all the testing data is shown in Table 3. It can be seen that the MAE and RMSE of the proposed MS-CNN are almost the lowest among the comparison methods. Although the proposed MS-CNN gets bad performance in bearing 2_4, the method is still robust in many different tasks. DNN method gets the biggest errors than other methods. MS-CNN in Ref. [15] has smaller errors
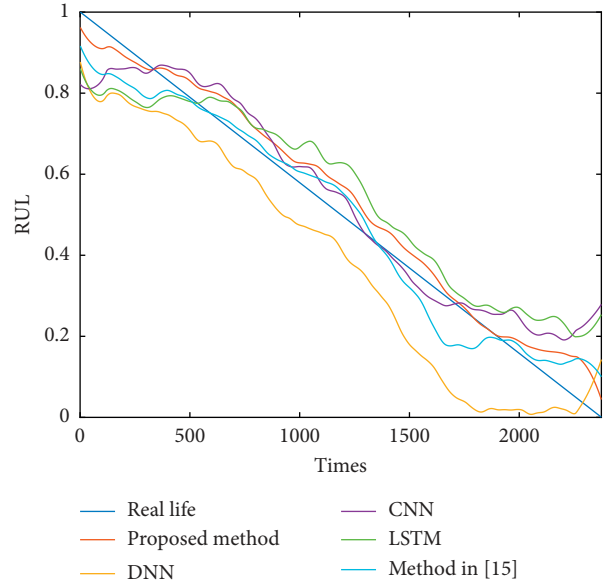
than CNN. That is because the combination of the final convolutional layer and the final pooling layer makes use of the local and global features learned by the high-level layer. But MS-CNN in Ref. [15] has bigger errors than the proposed method. The result shows which can be suggested that the detailed information extracted by low-level layers is useful for Prognostic. LSTM method shows worse performance than the CNN method. It is not suitable for extracting features from plenty of original data. What is more, the LSTM method consumes much more time to train than other methods, and it is not suitable for industrial

TABLE 3: Performance comparisons with different methods.

Bearing	Proposed method		DNN		CNN		LSTM		MS-CNN in [15]	
	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)	MAE (%)	RMSE (%)
1-3	6.14	7.83	17.31	22.54	8.50	10.52	11.21	14.68	6.75	8.98
1-4	19.29	24.62	34.14	36.27	26.31	33.28	26.71	32.37	24.38	27.47
1-5	17.46	23.98	28.74	32.15	24.84	32.22	25.15	32.05	18.15	22.25
1-6	18.19	21.96	32.46	38.43	22.86	26.81	24.31	29.08	20.33	22.95
1-7	10.87	17.96	19.68	32.6	12.09	17.96	16.85	23.19	10.31	16.36
2-3	17.54	23.23	40.51	46.19	25.35	31.36	28.18	34.58	25.18	30.25
2-4	29.81	34.98	42.73	55.72	32.78	39.04	32.69	40.70	26.14	28.69
2-5	21.71	28.23	37.46	41.03	22.56	29.41	26.32	32.75	23.62	30.16
2-6	19.20	24.35	29.45	39.85	22.20	27.77	29.89	36.22	20.70	25.71
2-7	27.20	32.64	41.20	44.21	38.29	43.02	28.37	33.68	22.14	30.26
3-3	13.78	17.97	28.05	23.75	18.01	21.83	16.86	20.37	15.46	20.07

applications. The results proved that the proposed method could provide reliable remaining useful life estimations in different failure behaviors.

5. Conclusions

In this paper, an MS-CNN-based method for bearing prognostic is proposed to overcome the shortcomings of traditional CNN. The effectiveness of the proposed method was verified on a public dataset. Some contributions are summarized as follows: (1) the integrated dilation convolution block can extract features in different receptive fields from the raw signal without increasing the parameters of the network; (2) the integrated dilation convolution block in different depths are concatenated, avoiding the loss of detailed information learned by the lower layer. The proposed architecture can show a high accuracy than other deep learning methods mentioned in this paper. However, the structure of the network is designed subjectively. Our future study is supposed to concentrate on optimizing the structure of the network automatically.

Data Availability

The data used in this paper are available, which can be downloaded from GitHub - wkzs111/phm-ieee-2012-data-challenge-dataset: dataset that was used during the PHM IEEE 2012 Data Challenge, built by the FEMTO-ST Institute.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] H. T. Shi and X. T. Bai, "Model-based uneven loading condition monitoring of full ceramic ball bearings in starved lubrication," *Mechanical Systems and Signal Processing*, vol. 139, Article ID 106583, 2020.

[2] S. Zhou, M. Xiao, P. Bartos et al., "Remaining useful life prediction and fault diagnosis of rolling bearings based on short-time Fourier transform and convolutional neural network," *Shock and Vibration*, vol. 2020, Article ID 8857307, 14 pages, 2020.

[3] M. Grasso, S. Chatterton, P. Pennacchi, and B. M. Colosimo, "A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis," *Mechanical Systems and Signal Processing*, vol. 81, pp. 126–147, 2016.

[4] Y. Chen, G. Peng, Z. Zhu et al., "A novel deep learning method based on attention mechanism for bearing remaining useful life prediction," *Applied Soft Computing*, vol. 86, Article ID 105919, 2020.

[5] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.

[6] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *Journal of Manufacturing Systems*, vol. 48, pp. 71–77, 2018.

[7] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.

[8] J. Wu, K. Hu, Y. Cheng et al., "Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network," *ISA Transactions*, vol. 97, pp. 241–250, 2019.

[9] 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.

[10] S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, "LSTM networks based on attention ordered neurons for gear remaining life prediction," *ISA Transactions*, vol. 106, pp. 343–354, 2020.

[11] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Transaction on Industrial Electronics*, vol. 66, pp. 3208–3216, 2019.

[12] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, pp. 142–150, 2018.

[13] B. Wang, Y. Lei, T. Yan et al., "Recurrent convolutional neural network: a new framework for remaining useful life prediction of machinery," *Neurocomputing*, vol. 379, pp. 117–129, 2019.

[14] R. Wang, F. Liu, F. Hou, W. Jiang, Q. Hou, and L. Yu, "A non-contact fault diagnosis method for rolling bearings based on acoustic imaging and convolutional neural networks," *IEEE Access*, vol. 8, pp. 132761–132774, 2020.

[15] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3208–3216, 2019.

[16] P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition," *Neurocomputing*, vol. 225, pp. 188–197, 2017.

[17] X. Wang, C. Q. Shen, M. Xia et al., "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Reliability Engineering & System Safety*, vol. 202, Article ID 107050, 2020.

[18] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, Columbus, OH, USA, June 2014.

[19] Y. Chang, J. Chen, C. Qu, and T. Pan, "Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels," *Renewable Energy*, vol. 153, pp. 205–213, 2020.

[20] H. Li, W. Zhao, Y. Zhang et al., "Remaining useful life prediction using multi-scale deep convolutional neural network," *Applied Soft Computing*, vol. 89, p. 13, Article ID 106113, 2020.

[21] C. Z. Wu, F. Z. Feng, S. J. Wu et al., "A method for constructing rolling bearing lifetime health indicator based on multi-scale convolutional neural networks," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 41, no. 11, p. 11, Article ID 526, 2019.

[22] B. Wang, Y. G. Lei, N. P. Li et al., "Deep separable convolutional network for remaining useful life prediction of machinery," *Mechanical Systems and Signal Processing*, vol. 134, p. 18, Article ID 106330, 2019.

[23] Y. Han, B. Tang, and L. Deng, "An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes," *Computers in Industry*, vol. 107, pp. 50–58, 2019.

[24] W. Sun, R. Zhao, R. Yan et al., "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1350–1359, 2017.

[25] L.-C. Chen, G. Papandreou, I. Kokkinos et al., "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2016.

[26] Y. Chen, Q. Guo, X. Liang, J. Wang, and Y. Qian, "Environmental sound classification with dilated convolutions," *Applied Acoustics*, vol. 148, pp. 123–132, 2019.

[27] N. Li, Y. Lei, J. Lin et al., "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.

[28] A. Saxena, J. Celaya, B. Saha et al., "Metrics for offline evaluation of prognostic performance," *International Journal of Prognostics and Health Management*, vol. 1, pp. 2153–2648, 2010.

[29] P. Nectoux, R. Gouriveau, K. Medjaher et al., "PRONOSTIA: an experimental platform for bearings accelerated degradation tests," in *Proceedings of the Conference on Prognostics and Health Management*, pp. 1–8, Denver, CO, USA, June 2012.