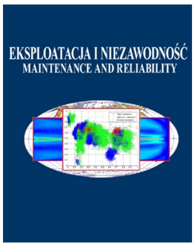

Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 23 (2021), Issue 4

journal homepage: http://www.ein.org.pl

Article citation info:
Wu C, Sun H, Lin S, Gao S. Remaining useful life prediction of bearings with different failure types based on multi-feature and deep convolution transfer learning. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2021; 23 (4): 685–694, http://doi.org/10.17531/ein.2021.4.11.

Indexed by: Web of Science Group

Remaining useful life prediction of bearings with different failure types based on multi-feature and deep convolution transfer learning

Chenchen Wu[a,b], Hongchun Sun[a,b,*], Senmiao Lin[a,b], Sheng Gao[a,b]

[a] School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[b] Key Laboratory of Vibration and Control of Aero-Propulsion Systems of Ministry of Education, Northeastern University, Shenyang 110819, China

Highlights

- Spatial pyramid pooling extracts multi-scale degradation features of bearings.
- TL solves the inconsistent distribution of degraded data for different failed bearings.
- The SPP-CNNTL model shows a better prediction effect on the RUL of the bearing.

Abstract

The accurate prediction of the remaining useful life (RUL) of rolling bearings is of immense importance in ensuring the safe and smooth operation of machinery and equipment. Although the prediction accuracy has been improved by a predictive model based on deep learning, it is still limited in engineering because lots of models use single-scale features to predict and assume that the degradation data of each bearing has a consistent distribution. In this paper, A deep convolutional migration network based on spatial pyramid pooling (SPP-CNNTL) is proposed to obtain higher prediction accuracy with self-extraction of multi-feature from the original vibrating signal. And to consider the differences of the data distribution in different failure types, transfer learning (TL) added with maximum mean difference (MMD) measurement function is used in the RUL prediction part. Finally, the data of IEEE PHM 2012 Challenge is used for verification, and the results show that the method in this paper has high prediction accuracy.

Keywords

rolling bearings, Remaining useful life (RUL), Convolutional Neural Networks (CNN), Transfer learning (TL).

1. Introduction

As one of the most important components in rotating machinery, rolling bearings play a vital role in the safe operation of mechanical equipment [5]. According to relevant statistics, 45% to 55% of the failure cases of rotating machinery are caused by the failure of rolling bearings [19]. Accurate RUL prediction technology can ensure both the safety of operator and equipment in good condition, and it is of a certain significance for the predictive maintenance.

The current methods used to predict RUL can be summarized into four categories [12]: physical model-based methods [11], statistical model-based methods [29], artificial intelligence-based methods [22], and hybrid methods [26]. The physical model-based methods describe the degradation process of machinery through the failure mechanism of mechanical equipment and mathematical model. Although this method can theoretically explain the degradation state of machinery, as the complexity of the mechanical system becomes higher and higher, it is difficult to establish an ideal degradation model. These statistical model-based methods can achieve predictions under different working conditions, but it is usually assumed that the degraded signal follows a parameterized process model, which may not be the case in reality [33]. The data-driven method gets rid of the shackles of traditional methods, and the degraded state of the bearing can be described based on the obtained bearing operating data. Therefore, the data-driven-based forecasting methods get wide attention. Recently, common models of data-driven methods gain very good effectiveness, such as Artificial Neural Network (ANN) [1], Support Vector Machine (SVM) [20, 24], Extreme Learning Machine (ELM) [28], etc. But each of these models is a shallow neural network that is of bad extraction ability and it is unable to directly mine the degraded information from the original data.

As a branch of machine learning, in recent years, deep learning emerges for its powerful feature extraction ability. Great progress has been made in image recognition, target detection, medicine, and other fields [13, 22, 23]. At present, the commonly used deep learning models in the mechanical field include Long Short-Term Network (LSTM) [30], Convolution Neural Network (CNN) [32], Stacked Denoise Autoencoding (SDA) [31], and Deep Belief Network (DBN) [21]. For instance, Wang et al [25] recurrent convolution layers were constructed to simulated the temporal correlation between different degradation states, and the variational inference was combined to measure the uncertainty of RUL prediction. It indicates that this neural network is obviously superior to other methods in terms of RUL prediction accuracy and convergence. Hinchi et al [9] use the convolutional layer to extract the features from the original data, and the

(*) Corresponding author.

E-mail addresses: C. Wu - 2079879663@qq.com, H. Sun - hchsun@mail.neu.edu.cn, S. Lin - 490230707@qq.com, S. Gao - wsgs1415926@gmail.com degradation process is captured by the LSTM layer to predict RUL. Wang et al [27] Transform original one-dimensional signal into the grey-scale image and use 2D-CNN network for feature extraction. Then the double Gaussian model is used to fit and predict the degradation curve. The results indicate that the method can predict the RUL of bearing, and this measurement has pretty good accuracy. Compared with the shallow neural network, the mentioned deep learning model has made some progress in the field of bearing RUL prediction., but two issues remain as follows:

1. Only the last layer feature is taken for the prediction of bearing RUL in most of the literature. Because the last feature is the most abstract feature, which makes the generalization ability of the network model worse, thus, the forecasting results of bearing RUL under various failure types cannot be accurate enough.
2. The impact of inconsistent bearing data distribution on the deep learning prediction model is not considered. Because the traditional deep learning model is suitable for the situation where the data distribution of the training set and the test set are consistent, however, even under the same working conditions and the same type of bearings, each bearing will show inconsistent degradation trends during the full life test of the bearing, resulting in bearing data that does not meet the assumptions of deep learning applications.

As a new learning paradigm in machine learning, transfer learning broadens the applicable conditions of deep learning. At present, it has been applied in the field of reliability. For example, Guo L et al [8] proposed a domain adaptive module to solve the difference between different bearing data distributions so as to realize bearing fault diagnosis across experimental platforms. Dong S et al [6] proposed a bearing degradation assessment model based on transfer learning and deep hierarchical feature extraction. Experiments show that the model can accurately identify the degraded stage of the bearing. Zhu J et al [33] applied the domain adaptive module proposed in Literature 23 to the field of bearing RUL prediction and successfully realized bearing RUL prediction under different working conditions. It can be seen that most applications of transfer learning in the mechanical field are dedicated to solving classification problems [6, 14], while regression problems have not been widely used [18]. However, transfer learning has great potential for simple prediction regression problems [15].

Therefore, in order to solve the above problems, a framework for RUL prediction of bearings based on SPP-CNNTL is proposed. First, the degradation stage of the bearing is divided by a binary classification network. This method avoids human error caused by manual threshold division. Then, for the data in the degradation stage of the bearing, the frequency spectrum is extracted as input, and one-dimensional CNN is used as the feature extraction network. The SPP layer is used as the last pooling layer of CNN to achieve convolutional features observed from different directions. In addition, transfer learning based on the MMD function is introduced in the CNN model to solve the problem of low prediction accuracy caused by inconsistent bearing data distribution of different fault types. Finally, the method in this paper is verified by the IEEE PHM 2012 data set, and the results show that the prediction accuracy of bearing RUL is better than other models.

The contributions of this article are summarized as follows:

1. The spatial pyramid pooling layer is used to realize multi-scale feature extraction of input data, avoiding the shortcomings of insufficient bearing degradation information extracted.
2. Transfer learning is used to solve the problem of inconsistent distribution of bearing degradation data and failure data, so as to realize the deep learning model to predict the RUL of different failed bearings.
3. Propose an end-to-end prediction framework applicable to different faulty bearings, and promote the development of predictive maintenance technology for bearings.

The remainder of this paper is organized as follows: Section 2 describes the framework of the bearing remaining life prediction method proposed in this paper. The related theories of CNN and transfer learning networks are introduced, and the framework of the SPP-CNNTL neural network is proposed. In Section 3, the experimental analysis based on the full life data set of the bearing shows the effectiveness of the method. The comparison with other model methods highlights the superiority of this method. Finally, conclusions are given, and some future research directions are proposed in Section 4.

2. Proposed framework

2.1. Overall overview

In engineering applications, due to bearing processing and manufacturing errors, assembly errors, and material defects of the bearing itself, the entire degradation process of the bearing from the initial use to the final failure shows different trends. This leads to the problem of differences in the data distribution between the degradation data of each bearing. This violates the assumption that deep learning requires the training set and test set to have the same data distribution, so it reduces the RUL prediction accuracy. Therefore, this paper proposes a framework for predicting the remaining life of bearings based on a multi-scale convolutional transfer learning model. The flow of the framework is shown in Figure 1. It can be seen from Figure 1 that the method in this article is mainly divided into two parts: the first part is the degradation stage division. This part uses the normal stage data and the severe stage data of the bearing to construct a data set, trains the two-class neural network and realizes the degradation stage Automatic division. This method avoids the human error caused by the trouble of manually setting the fault threshold in the traditional method and makes the recognition effect more objective. When the bearing enters the degradation stage, the second part starts to predict the RUL of the bearing based on the SPP-CNNTL model. The model adds an SPP pool to solve the problem of the poor generalization ability of single-scale input. The domain adaptation technology in transfer learning is used to measure the difference between degraded data distributions in different directions and use the difference as a constraint condition of the prediction model so that the network model can learn the invariance between different failed bearing data.

2.2. Transfer learning

As a branch of machine learning, transfer learning can transfer learned knowledge in a different area, and its main idea is to find similarities between different datasets. Two basic concepts are mainly included in transfer learning, which are domain and task. The domain is the subject of learning, which is mainly composed of data and the probability distribution which can generate these data; Task is the goal of learning, which is mainly composed of tag and tag's corresponding function group. Thus, transfer learning can be expressed as follows: a labeled source domain $D_s = \{x_i, y_i\}_{i=1}^{n}$ and an unlabeled target domain $D_t = \{x_i\}_{i=1}^{n}$. They have different data distribution, $P_s(X_s) \neq P_t(X_t)$. The goal of transfer learning is to use labeled data D_s to learn the knowledge of the target domain D_t.

Domain adaptation is one of the research contents of transfer learning, which focuses on solving the problem of consistent feature space, consistent category space, and only inconsistent feature distribution. Domain adaptation mainly includes two strategies: One is to introduce the measurement function, minimizing its value to make the source domain and target domain obey the same distribution. Some measurement functions, such as Maximum Mean Discrepancy (MMD), KL divergence and CORAL, are often used. The other is to draw on the experience of the strategy of Generative Adversarial Network (GAN) --- adding domain classification module [4, 33].

Domain adaptive technology is proposed to solve the problem of different failure types of bearing RUL prediction, because domain

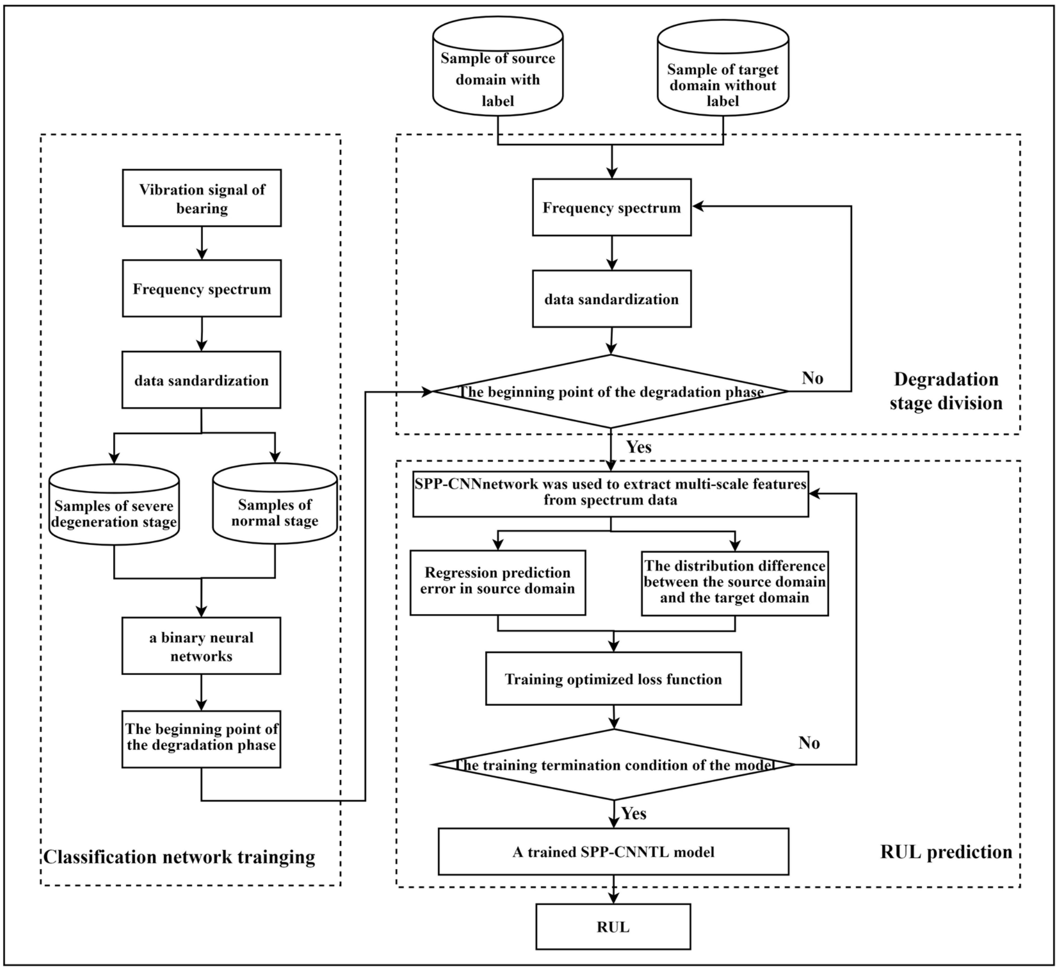
Sample of source domain with label
Sample of target domain without label
Vibration signal of bearing
Frequency spectrum
data sandardization
Samples of severe degeneration stage
Samples of normal stage
a binary neural networks
The beginning point of the degradation phase
Classification network trainging
Frequency spectrum
data sandardization
The beginning point of the degradation phase
No
Degradation stage division
Yes
SPP-CNNnetwork was used to extract multi-scale features from spectrum data
Regression prediction error in source domain
The distribution difference between the source domain and the target domain
Training optimized loss function
The training termination condition of the model
No
Yes
A trained SPP-CNNTL model
RUL prediction
RUL

Fig. 1. Flow chart of the method proposed in this article

adaptive technology can perform classification and prediction when the data distribution of the training set and the test set are similar. Questions in this article is described in transfer learning language as follows:

1. To get some labeled degenerative data and to be used as training set, $D_s = \{\chi_s, P_s(X)\}$ and get some unlabeled degenerative data as test set, $D_t = \{\chi_t, P_t(X)\}$.
2. Assuming the feature space of the source of domain and the target domain is the same, $\chi_s = \chi_t$. But the marginal distribution of two domains is different, $P_s(X_s) \neq P_t(X_t)$.
3. A classifier $f : x_t \to y_t$ is adopted to improve the accuracy of prediction by using the auxiliary data that are composed of labelled data- D_s and partial unlabeled data- D_t .

2.3. CNN

CNN is a kind of feedforward neural network, which was first proposed by LeCun in 1989 and used for image processing [10]. The CNN network mainly consists of convolution layers, pooling layers, and full connection layers. The convolutional layer reduces the parameter amount of the model by capturing the local regional connection feature of input information and applying the weight sharing principle, and further reduces the amount of training data by combing the similar features through the pooling layer. In order to extract features from the data, the CNN model usually alternately stacks convolutional layers and pooling layers, and configures the output layer as a fully connected layer.

1. Convolutional layer

The convolution layer consists of a set of convolution kernels, which are the core of feature extraction. The convolution kernel performs a convolution operation on the feature map output by the previous layer to achieve feature extraction of the local area. In addition, the convolutional layer also has the characteristics of weight distribution, which greatly reduces network parameters and avoids over-fitting. The specific convolutional layer operation is shown in the formula (1):

$$x_c^l = \sigma\left(\sum_{i=1}^{c^{l-1}} W_{i,c}^l * x_i^{l-1} + b_c^l\right) \tag{1}$$

where x_i^{l-1} is the output of channel i of l-1 layer, $W_{i,c}^l$ is the convolution kernel for layer l, b_c^l is bias, * is convolution operation, x_c^l is the output of channel c of layer l. $\sigma(\cdot)$ is the activation function. In this paper, the ReLU function is used as the activation function of the CNN network because it has the ability to accelerate the convergence and alleviate the vanishing gradient problem. The calculation is as follows:

$$ReLU(x) = \max(0, x) \tag{2}$$

2. Pooling layer

The main purpose of the pooling layer is to reduce the parameters of the neural network. It is usually added between two convolutional layers, and the input of the convolutional layer at a specific connection position is summarized in the form of non-linear sampling to improve the computational efficiency of the network and keep the feature translation unchanged. Common pooling layers include average pooling, maximum pooling, etc. And maximum pooling is used in this paper partially. The equation (3) is as follows:

$$p_c^l = \max\left\{x_{c\times k:(c+1)\times k}^l\right\} \tag{3}$$

where k is the length of pooling, p_c^l is the output of channel c layer1.

3. Spatial pyramid pooling

In order to solve the problem of inconsistent input image size, a spatial pyramid pool for target detection task is first proposed. SPP can extract features of different dimensions from the feature map by using pool kernels of various sizes, and stitch them to obtain multi-dimensional features. Therefore, this article adds SPP to the last layer of the CNN network model for multi-feature extraction to improve the generalization of the network.

4. Fully connected layer

The purpose of the fully connected layer is to perform regression or prediction tasks on the extracted features. After executing the SPP-CNN model in this article, the network will output multiple feature values and then pave them. The mapping between features and bearing RUL uses fully connected layers. The calculation formula (4) between complete connections is as follows:

$$h^l = \sigma^l\left(\left(W^l\right)^T \times v^{l-1} + b^l\right) \tag{4}$$

where σ^l is the activation function of the layer l, v^{l-1} is the output vector of layer l-1, W^l is the connection weight of the neurons in the l-th layer and the neurons in the l-1th layer, b^l is the bias, h^l is the output feature of the l-th hidden layer. The activation function of the output layer is the SoftMax function, and the other layers are the ReLU function.

2.4. SPP-CNNTL Learning model

The Figure 2 shows the framework of the SPP-CNNTL network model proposed in this paper. The network model mainly includes three parts: Multi-scale feature extraction module, regression prediction module, domain adaptive module. Among them, multi-scale feature extraction mainly uses the SPP-CNN model for feature extraction. The features that can represent bearing degradation information are extracted layer by layer by convolution and pooling operations from the input source domain and target domain. The regression prediction module is to predict the RUL of the bearing. The module uses the extracted multi-scale features as the judgment basis, and realizes the RUL prediction of the source domain samples through the fully connected layer. The domain adaptation module is based on the data distribution difference between the source domain and the target domain in the specified layer, and uses the MMD function value as a measure to constrain the RUL prediction part to minimize the difference between the data distribution. The specific network model structure is shown in the Table 1.

Table 1. SPP-CNNTL Network Model diagram

Layer	Module	Symbol	Operation	Parameter
1	Feature extraction	Input	Input signal	1×2048
2		C1	Convolution	5×1×3
3		P1	Pooling	2
4		C2	Convolution	5×3×6
5		P2	Pooling	2
6		SPP	Multi-Pooling	/
7		Flatten	/	126
8	Domain adaptive	FC1	Fully-connected	50
9		FC2	Fully-connected	10
10	RUL prediction	FO	Sigmoid	/

2.4.1. Domain adaptive model

Domain adaptive model is mainly to describe the difference among the data distribution of data set in some measures. Maximum mean difference is taken as the measurement function in this paper. This method measures the distance between two reproducing Hilbert space, which is a kernel learning method. The equation (5) is as follows:

$$MMD\left(h^s,h^t\right) = \left\|\frac{1}{n^s}\sum_{i=1}^{n^s}\phi\left(h_i^s\right) - \frac{1}{n^t}\sum_{i=1}^{n^t}\phi\left(h_j^t\right)\right\|_H^2 \tag{5}$$

where n^s the number of samples from the source domain, n^t is the number of samples from the target domain, $\phi(\cdot)$ is mapping which maps the original variable to the regenerative nuclear Hilbert space, $\|\cdot\|_H$ is the regenerative nuclear Hilbert space.

2.4.2. Target of optimization

The loss function of the proposed method are two parts:

1. Root mean square error term of the minimized regression task.

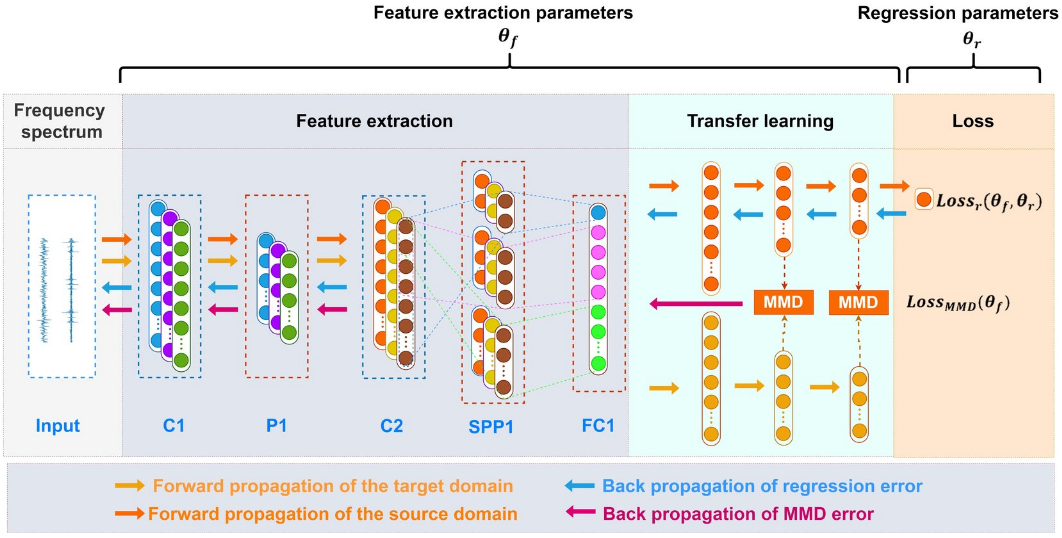
Feature extraction parameters
θ_f
Regression parameters
θ_r
Frequency spectrum
Feature extraction
Transfer learning
Loss
$Loss_r(\theta_f,\theta_r)$
MMD
MMD
$Loss_{MMD}(\theta_f)$
Input
C1
P1
C2
SPP1
FC1
Forward propagation of the target domain
Forward propagation of the source domain
Back propagation of regression error
Back propagation of MMD error

Fig. 2. SPP-CNNTL Network Model diagram

2. Minimized MMD term between the source domain and the target domain.

Loss function 1: The accuracy of RUL prediction of bearing is improved by minimizing differences in values. In other words, the main loss function is the difference between the predicted value and real labelled value. For regression tasks, the Mean Square Error (MSE) is the most commonly used as loss function. The equation is as follows:

$$Loss_r = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2 \tag{6}$$

where m is the size of batch of training set, y_i is the real label, $\hat{y}_i$ is the label of prediction.

Loss function 2: The migration of the last two layers is selected after analysis: for the RUL prediction of bearing after the full connection layer, the difference among different domains is minimized after MMD is added into different layers. The equation is as follows:

$$Loss_{MMD1} = \frac{1}{m_s^2}\sum_{i=1}^{m_s}\sum_{j=1}^{m_s}k\left(f1_i^s, f1_j^s\right) + \frac{1}{m_t^2}\sum_{i=1}^{m_t}\sum_{j=1}^{m_t}k\left(f1_i^t, f1_j^t\right) - \frac{1}{m_s m_t}\sum_{i=1}^{m_s}\sum_{j=1}^{m_t}k\left(f1_i^s, f1_j^t\right)$$

$$Loss_{MMD2} = \frac{1}{m_s^2}\sum_{i=1}^{m_s}\sum_{j=1}^{m_s}k\left(f2_i^s, f2_j^s\right) + \frac{1}{m_t^2}\sum_{i=1}^{m_t}\sum_{j=1}^{m_t}k\left(f2_i^t, f2_j^t\right) - \frac{1}{m_s m_t}\sum_{i=1}^{m_s}\sum_{j=1}^{m_t}k\left(f2_i^s, f2_j^t\right)$$

$$Loss_{MMD} = Loss_{MMD1} + Loss_{MMD2} \tag{7}$$

where $Loss_{MMD1}$ is the value of the last layer, $Loss_{MMD2}$ is the inverted second layer, $k(\cdot)$ is the kernel function, m_s is the number of source domain samples, m_t is the number of the target domain samples.

The final total loss function is as follows:

$$Loss = Loss_r + \lambda Loss_{MMD} \tag{8}$$

where hyperparameter λ decide the effect of MMD differences on prediction.

And set the parameter of feature extractor as θ_f ,and set the parameter of regression prediction of bearing RUL as θ_r .The equation 8 can be rewritten as follows:

$$Loss\left(\theta_f, \theta_r\right) = Loss_r\left(\theta_f, \theta_r\right) + \lambda Loss_{MMD}\left(\theta_f\right) \tag{9}$$

Adam optimizer is used to minimize the loss function and to find the saddle point of the loss function. The equation is as follows:

$$\theta_f \leftarrow \theta_f - \eta\left(\frac{\partial Loss_r}{\partial \theta_f} + \lambda\frac{\partial Loss_{mmd}}{\partial \theta_f}\right)$$

$$\theta_r \leftarrow \theta_r - \eta\left(\frac{\partial Loss_r}{\partial \theta_r}\right) \tag{10}$$

where η is learning rate.

3. Application of the proposed method

3.1. Introduction of data set

IEEE PHM 2012 Challenge [16] is adopted to verify the effectiveness of the method proposed in this paper. Experiment platform of PRONSTIA is constructed as the Figure 3. The test-bed consists of two parts: part of experimental simulation and part of measurement. The power of the experimental simulation is output by a motor with a power of 250 W. And the load simulation is applied to the bearing to accelerate the degradation of the bearing by applying a radial force load. The measurement portion adopts an acceleration sensor whose sampling frequency is 25.6 kHz and the acquisition channel is two channels in the horizontal and vertical direction. A signal sample is collected every 10s, and the length of the collected time is 0.1 s.

The data set contains bearing work data under three different loads. Working-condition 1: under 1800 rpm and 4000 N; Working-condition 2: 1650 rpm and 4200 N; Working-condition 3: 1500 rpm, 5000 N. Total 17 data sets of bearing are acquired which are working to failure. In condition 1, there are 7 bearings numbered from 1-1 to 1-7; In condition 2, there are 7 bearings numbered from 2-1 to 2-7; In condition 3, there 3 bearings numbered from 3-1 to 3-3. This paper selects the bearing in condition 1 for testing, and its partial degradation data is shown in Figure 4. Although the bearings are in the same working condition, they behave differently in degradation process. As pointed out in literature 3, under the working-condition 1, the bearings, 1-1 1-3 1-4, belongs to the same type of progressive degradation failure; the bearings, 1-2 1-5 1-6 1-7, belongs to the same type of sudden burst degenerate failure.

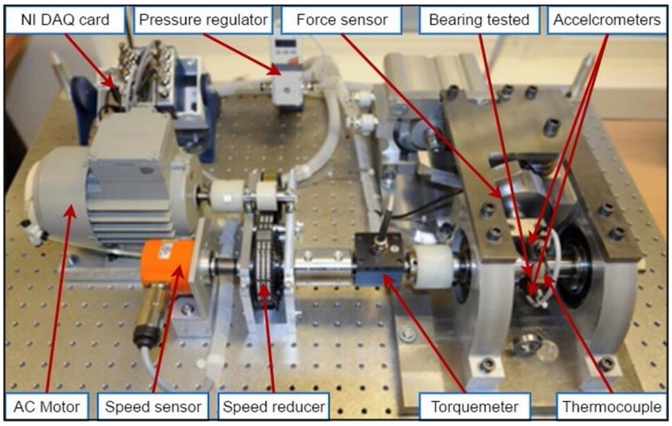

Fig. 3. The experimental platform

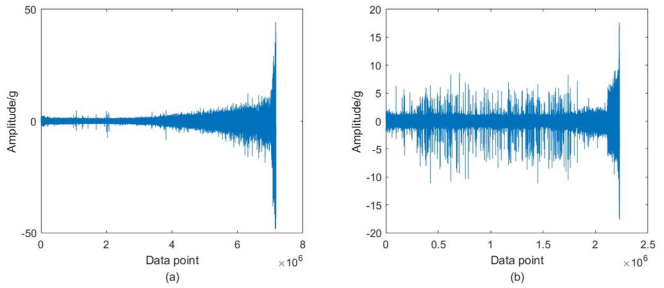

Fig. 4. Bearing degradation data under working-condition 1, (a) bearing1-1; (b) bearing1-2.

3.2. Starting point identification of degradation stage

The bearing 1-1 and 1-2 are selected as training set and the rest of them are used for testing. The full life diagram of raw signal is shown as Figure 5. The 500th-1000th collected data of bearing 1-1 and the 320th-400th collected data of bearing 1-2 are used as the normal stage data; the 2400th-2700th collected data of bearing 1-1 and the 831th-861th collected data are used as the data of severe fault stage.

Spectrum data is used as training data of binary classification neural network. Hardware of the experiment is a computer with i5-1035G1 CPU @ 1.00 GHz 1.19 GHz, 16 GB memory and software are MATLAB 2016a and PYTHON3.8.

After many attempts, the four-layer neural network is selected as the classifier, the number of the network nodes is 2048-10000-500-2 and the activation function of the front three-layer is the RELU function, the last layer use SoftMax function as activation function to implement the binary classification. The loss function is set as a cross-entropy function, train the network 20 times and the batch size is 8. In order to avoid false alarms, three consecutive predictions into the degradation stage mean that the stage is into degradation. Figure 6 shows some test bearing results. It can be seen from Figure 6 that the two-classification network can more accurately identify samples in the normal phase and samples in the degraded phase. Therefore, it can accurately determine the starting point of the degradation stage. The overall test results are shown in Table 2.

Table 2. Recognition of starting point during degradation phase

Bearing	Failure time/s	Failure point/s
1-1	2803	1517
1-2	871	821
1-3	2375	1332
1-4	1428	1090
1-5	2463	2444
1-6	2448	2100
1-7	2259	2241

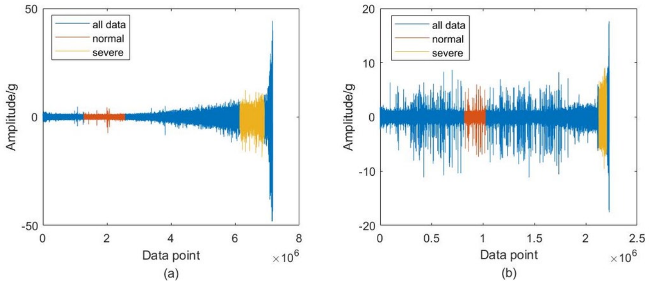

Fig. 5. The original vibration waveform of the bearing, (a) bearing1-1, (b) bearing1-2

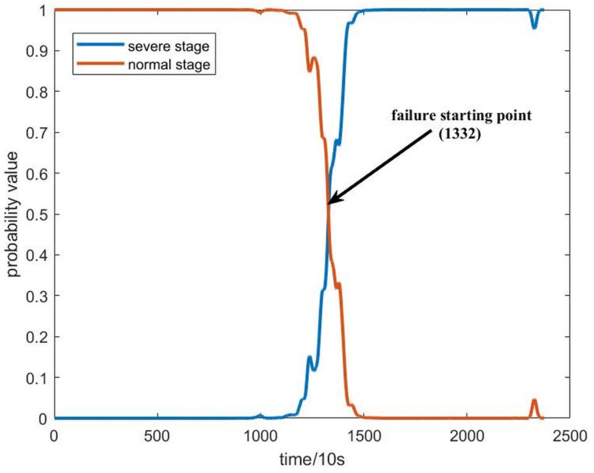

Fig. 6. Stage identification effect diagram of bearings 1-3

3.3. Prediction of RUL

3.3.1. Evaluation index and sample label

In order to quantitatively evaluate the effectiveness of the predictive RUL method proposed in this paper, this paper uses Root-Mean-Square-Error (RMSE) and Mean-Absolute-Error (MAE) as evaluation indicators. The calculation formula is shown in formula (11):

$$\begin{cases} MAE = \frac{1}{m}\sum_{i=1}^{m}\left|(y_i - \hat{y}_i)\right| \\ RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2} \end{cases} \tag{11}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and m is the number of samples.

Since the prediction model of RUL used in this paper is a supervised learning model, it is necessary to label the source domain samples. This article uses the remaining life percentage of the bearing as the label for these samples. This tag can control the amount of data used for network training not to be too large, and improve computational efficiency. (For example, assuming failure time of bearing is 2500 s and time of degradation is 500 s, when the bearing running at 1500 s, the label for that point is ($\frac{(1500-500)}{(2500-500)} = 50\%$).

3.3.2. Hyperparameters of the network

In order to obtain the best model prediction effect, this section discusses the important hyperparameters and network structure of the network. Since the setting of the learning rate will affect the convergence of the network model, which in turn affects the training effect of the model, the learning rate is an indicator that must be considered. Secondly, this paper uses the MMD function value as a scale function to measure the data of different failed bearings, and uses it as a part of the loss function, so it is of great significance to choose the MMD term trade-off coefficient. Therefore, this paper chooses the learning rate and the trade-off coefficient for experiments, and the selection range of hyperparameters is shown in Table 3.

Table 3. Value range of Hyper-parameters

Hyperparameters	Range
Learning rate	0.1, 0.01, 0.005, 0.001, 0.0005, 0.0001
Trade-off value	0, 0.1, 0.2, 0.3, 0.5, 0.7, 10, 50, 100

When the fixed trade-off coefficient is 0.2, try to experiment with different learning rate values. The prediction results are shown in Table 4. The values in Table 4 are the average values of multiple prediction results of all training set bearings. It can be seen from Table 4 that when the learning rate is large, the effect of the model is the worst. The possible reason is that a higher learning rate will prevent the network from converging to an optimal value. Because the gradient descent step is too large, it can only make the model hover around the optimal value, resulting in lower prediction accuracy. As the learning rate decreases, the prediction accuracy continues to improve. However, too small a learning rate will reduce the convergence speed. Under the same number of iterations, too small a learning rate may not achieve convergence. Therefore, considering the prediction accuracy and time-consuming considerations, this paper chooses the learning rate to be 0.001.

Table 5 shows the prediction effect of the compromise coefficient under different values. It can be seen from Table 5 that when the trade-off coefficient is selected as 0.2, the performance of the network model is the best. If the trade-off coefficient is too small, the constraint in-

Table 4. Influence of different learning rates on the prediction model

Learning rate	MAE	RMSE
0.1	0.2518	0.2909
0.01	0.1803	0.2219
0.005	0.1870	0.2313
0.001	**0.1702**	**0.2085**
0.0005	0.1930	0.2340
0.0001	0.1905	0.2280

Table 5. Influence of different trade-off coefficients on model prediction

Trade-off value	MAE	RMSE
0	0.2121	0.2595
0.1	0.1923	0.2378
0.2	**0.1702**	**0.2085**
0.3	0.1858	0.2263
0.5	0.1876	0.2247
0.7	0.1992	0.2410
10	0.1999	0.2400
50	0.1909	0.2310
100	0.1999	0.2361

formation between different data sets will be reduced, and the model will not be able to learn domain-invariant features. When the trade-off coefficient is greater than 0.5, because the weight of the MMD term is too large, the loss of the prediction model cannot be trained well. In summary, the compromise factor of 0.2 in this article is reasonable.

In order to determine the influence of the architecture of the network model, this paper adds the MMD function to the last layer of the network model (MMD1), adds the MMD function to the penultimate layer (MMD2), and adds MMD function to the last two layers (MMD12). The experimental results are shown in Table 6. Since the network model extracts the shallow information of the network model in the first few layers, the features extracted by the network model are more abstract in the subsequent layers. It can be seen from Table 6 that the effect of the single-layer MMD function is not as good as that of the double-layer MMD function. This is mainly because the single-layer MMD function is not enough to represent the difference in data distribution between the training set and the test. Therefore, it is reasonable to choose MMD12 as the network model architecture of this article.

Table 6. Influence of different locations of MMD on prediction

Trade-off value	MAE	RMSE
MMD1	0.1723	0.2159
MMD2	0.1860	0.2254
MMD12	**0.1702**	**0.2085**

3.3.3. Prediction of RUL

The PHM data set is used as the analysis data to verify the effectiveness and feasibility of the method in this paper. The original data of bearing 1-1 is used as the training set, and the lifetime percentage is used as the sample label, which belongs to the source domain. Unmarked data for bearings 1-5 and 1-7 are used as auxiliary data. The test sets are Bearing 1-2, 1-3, 1-4, 1-6.

Through theoretical analysis and experimental verification, the hyperparameters of the experimental model are set that Optimizer is Adam, Learning rate=0.001, Trade-off=0.2, Epoch=400, Batchsize=32. The network adopts two convolution and pooling layers for feature exaction, the kernel size is 5 in convolution and 2 in max pooling. In the transfer part of the full connection layer, the RBF function is selected as the kernel function for calculation of MMD distance and the width of the kernel is 1000. When the MMD measurement loss function accounts for 0.20 total loss, the network reaches the optimal effect. The batch size is 32, and half the data comes from the source domain, the rest is from the target domain. The epoch is set as 400. The loss function of the training process is shown in Figure 7. It can be seen that as the number of epochs increases, the loss of the training model does not decrease, indicating that the model has reached the effect of convergence. The prediction effect of the training set direction is shown in Figure 8. It can be seen from Figure 8 that this method shows a good fitting effect and good monotonicity for the bearings of the training set, and the failure time of the bearing can be almost perfectly predicted in the final stage. At the same time, it shows that the network architecture and hyperparameters selected in this paper are reasonable, and the network model can learn bearing degradation information from the training set.

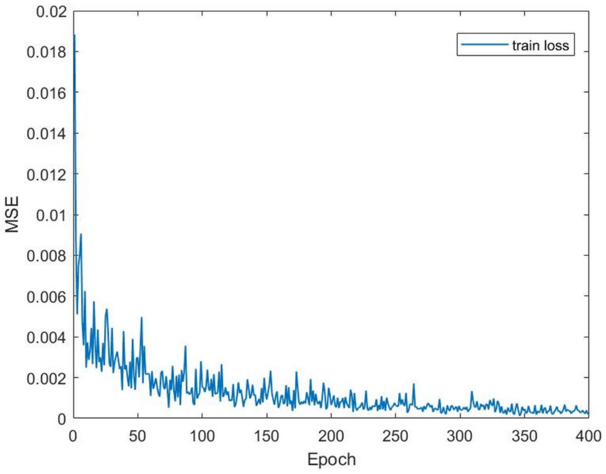

Fig. 7. Training loss diagram of network model

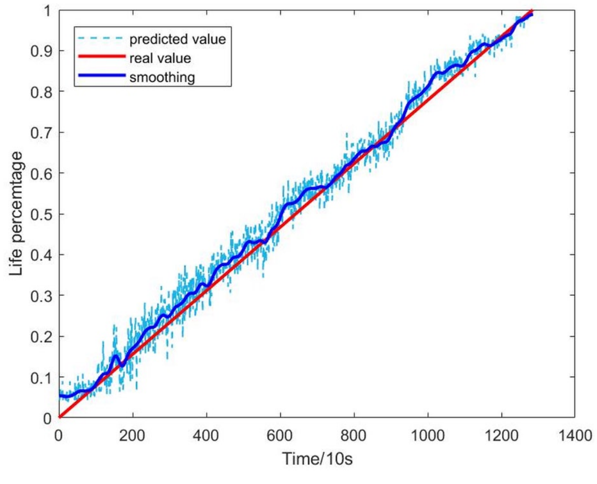

Fig. 8. The prediction effect of bearing in train set(bearing1-1)

As shown in Figure 9, it can be seen that the method in this paper shows high prediction accuracy for both the suddenly failed bearing 1-2 and the gradually failed bearing 1-3, and the fluctuation of the predicted value is significantly reduced after sliding average processing. Although in the process of predicting the degradation trend of the network model, the monotonicity of the bearing 1-2 is not satisfactory. However, in actual engineering, people pay more attention to the degradation trend and final RUL value of the bearing in the later period of

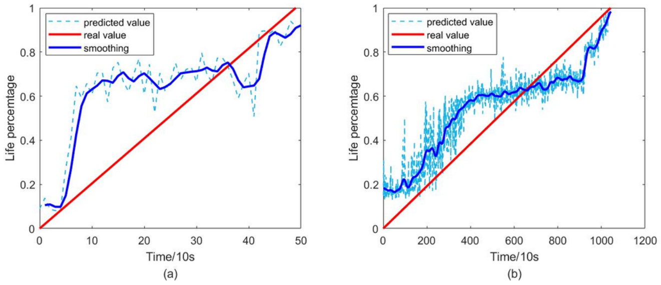

Fig. 9. The prediction effect of bearing in test set, (a) bearing1-2; (b) bearing1-3

operation. It can be seen from Figure 9 that both bearing 1-2 and bearing 1-3 have good monotonicity and higher prediction accuracy at the final moment. Even bearings 1-3 can predict the failure time almost without error at the last moment. In summary, the method proposed in this paper can meet the requirements of the RUL prediction of the bearing in actual engineering

3.4. Comparison analysis of model advantage

In order to verify the superiority of this method, this paper chooses the CNN model and the SPP-CNN model as the comparison model to verify the effectiveness of the improved strategy. Secondly, in order to verify the effectiveness of the migration strategy in this article, the current advanced migration learning models Transfer Component Analysis (TCA) and Domain-Adversarial Training of Neural Networks (DANN) are used as comparison models. The comparison model introduction is shown in Table 7.

Table 7. Comparison model

Model	Input	Transfer method
CNN	frequency spectrum	None
SPP-CNN	frequency spectrum	None
TCA [17]	traditional feature	MMD
DANN [7]	frequency spectrum	adversarial strategy
SPP-CNNTL (Proposed method)	frequency spectrum	MMD

In order to ensure the accuracy of the comparison effect, the architecture and hyperparameter settings of the comparison model are consistent with the selection of the proposed method. The experimental prediction results of different models are shown in Table 8.

Table 8. The MAE value of different models

Model	bearing1-1	bearing 1-2	bearing 1-3	bearing 1-4	bearing 1-6
CNN	**0.0160**	0.2828	0.2595	0.2083	0.3062
SPP-CNN	0.419	0.2580	0.1454	0.1651	0.3062
TCA	0.5023	0.2543	0.2034	0.1823	0.3124
DANN	0.0432	0.2392	0.1224	0.1523	0.3034
SPP-CNNTL	0.0201	**0.1802**	**0.1115**	**0.1332**	**0.2477**

From Table 8, compared with other models, there are three kinds of advantages in the proposed method in this paper.

1. The SPP-CNN model improves the accuracy of bearing RUL prediction. Although the traditional CNN model has higher prediction accuracy on the training set, its prediction effect on the test set is worse than that of the SPP-CNN model. The main reason is that SPP can improve the generalization ability, thereby improving the RUL prediction effect of the bearing under different failure degradation.
2. Transfer learning improves the accuracy of bearing RUL prediction. After using transfer learning, the model prediction ability of the training set and test set has been improved. It also has a better predictive effect for bearings that suddenly fail.
3. In order to demonstrate the superiority of the transfer strategy, this paper chooses TCA and DANN as the comparison model. The TCA model maps the features of the source domain and the target domain to the high-dimensional replicable kernel Hilbert space to minimize the distance between the source domain and the target domain. The input of the TCA model is 24 traditional statistical features, including time-domain features and wavelet packet energy. It selects the RBF function as the kernel function. The DANN model uses domain confrontation strategies to solve the problem of data distribution differences. The prediction effect of each model is shown in Table 7. It can be seen from the evaluation indicators in Table 7 that this paper has a higher RUL prediction accuracy for the tested bearing. Compared with other transfer learning models, the proposed method has higher prediction accuracy. The main reason is the use of adaptive technology to solve the problem of inconsistent allocation between training data and test data. And use the SPP-CNN layer to improve the generalization ability of the network to obtain a better transmission effect.

4. Conclusion

This paper proposes a RUL prediction model of bearing based on multi-feature deep convolution transfer learning. First of all, this paper uses the SPP layer to avoid the problems of poor prediction accuracy and poor generalization ability of a single feature. Then, based on the MMD migration mechanism, the SPP-CNN model was improved, and the problem of inconsistent data distribution of the degradation trend of each bearing caused by the failure of each bearing was solved. Finally, by using the PHM2012 bearing public data set, and comparing the results with the prediction effect of the transfer learning model, the following conclusions are drawn: 1. The method proposed in this paper has good monotonicity in the final stage of various types of failed bearings. Higher prediction accuracy can meet the actual needs of engineering applications. 2. The domain adaptive module can reduce the data distribution difference between different failure trends, so that the model in this paper has a wider application range. From the above content, it can be seen that compared with the current advanced RUL prediction, the method in this paper has obvious advantages.

Considering the great potential of deep learning models in RUL prediction, future work shows that the RUL prediction of bearings under different working conditions should be considered, so that the RUL prediction model has stronger practicability.

Acknowledgement

This research is subsidized by the Natural Science Foundation of China, 'Research on reliability theory and method of total fatigue life for large complex mechanical structures' (Grant No. U1708255).

References

1. Ali J B, Chebel-Morello B, Saidi L, Malinowski S, Fnaiech F. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mechanical Systems and Signal Processing 2015; 56-57:150-172, https://doi.org/10.1016/j.ymssp.2014.10.014.
2. Burns J E, Yao J, Chalhoub D, Chen J J, Summers R M. A Machine Learning Algorithm to Estimate Sarcopenia on Abdominal CT. Original Investigation 2020; 27(3):311-320, https://doi.org/10.1016/j.acra.2019.03.011.
3. Cheng H, Kong X, Chen G, Wang Q, Wang R. Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors. Measurement 2020; 168:108286, https://doi.org/10.1016/j.measurement.2020.108286.
4. Costa P, Akcay A, Zhang Y, Kaymak U. Remaining Useful Lifetime prediction via deep domain adaptation. Reliability Engineering & System Safety 2020; 195:106682, https://doi.org/10.1016/j.ress.2019.106682.
5. Dong S, Luo T. Bearing degradation process prediction based on the PCA and optimized LS-SVM model. Measurement 2013; 46(9):3143–3152, https://doi.org/10.1016/j.measurement.2013.06.038.
6. Dong S, Wen G, Lei Z, Zhang Z. Transfer learning for bearing performance degradation assessment based on deep hierarchical features. ISA Transactions 2020; 108(9):343-355, https://doi.org/10.1016/j.isatra.2020.09.004.
7. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. The journal of machine learning research 2016; 17(1): 2096-2030, https://dl.acm.org/doi/abs/10.5555/2946645.2946704.
8. Guo L, Lei Y, Xing S, Yan T, Li N. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. IEEE Transactions on Industrial Electronics 2018; 66(9): 7316-7325, https://doi.org/10.1109/TIE.2018.2877090.
9. Hinchi A Z, Tkiouat M. Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. Procedia Computer Science 2018; 127:123-132, https://doi.org/10.1016/j.procs.2018.01.106.
10 LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, Jackel L D. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation 1989; 1(4): 541–551, https://doi.org/10.1162/neco.1989.1.4.541.
11. Lei Y, Li N, Gontarz S, Jing L, Radkowski S, Dybala J. A model-based method for remaining useful life prediction of machinery. IEEE Transactions on reliability 2016; 65(3): 1314-1326, https://doi.org/10.1109/TR.2016.2570568.
12. Lei Y, Li N, Guo L, Li N, Yan T, Jing L. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing 2018; 104:799-834, https://doi.org/10.1016/j.ymssp.2017.11.016.
13. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Matti P. Deep Learning for Generic Object Detection: A Survey. International Journal of Computer Vision 2020; 128: 261–318, https://doi.org/10.1007/s11263-019-01247-4.
14. Lu W, Liang B, Cheng Y, Meng D, Yang J, Zhang T. Deep model based domain adaptation for fault diagnosis. IEEE Transactions on Industrial Electronics 2016; 64(3):2296-2305, https://doi.org/10.1109/TIE.2016.2627020.
15. Mao W, He J, Ming J Z. Predicting Remaining Useful Life of Rolling Bearing Based on Deep Feature Representation and Transfer Learning. IEEE Transactions on Instrumentation and Measurement 2019; 69(4):1594-1608, https://doi.org/10.1109/TIM.2019.2917735.
16. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Varnier C. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management. Denver, CO, USA, 1-8, 2012, https://hal.archives-ouvertes.fr/hal-00719503.
17. Pan S J, Tsang I W, Kwok J T, Yang Q. Domain adaptation via transfer component analysis. IEEE Transactions on Neural Networks 2011; 22(2):199–210, https://doi.org/10.1109/TNN.2010.2091281.
18. Mao W, He J, Ming J Z. Predicting Remaining Useful Life of Rolling Bearing Based on Deep Feature Representation and Transfer Learning. IEEE Transactions on Instrumentation and Measurement 2019; 69(4):1594-1608, https://doi.org/10.1109/TIM.2019.2917735.
19. Rai A, Upadhyay S H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International 2016; 96:289-306, https://doi.org/10.1016/j.triboint.2015.12.037.
20. Rai A, Upadhyay S H. Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 2018; 232(6):1118-1132, https://doi.org/10.1177/0954406217700180.
21. Salakhutdinov R, Hinton G. An Efficient Learning Procedure for Deep Boltzmann Machines. Neural Computation 2012; 24(8): 1967–2006, https://doi.org/10.1162/NECO_a_00311.
22. Su C, Li L, Wen Z. Remaining useful life prediction via a variational autoencoder and a time-window-based sequence neural network. Quality and Reliability Engineering International 2020; 36(5): 1639-1656, https://doi.org/10.1002/qre.2651.
23. Tang Y, Chen M, Wang C, Luo L, Zou X. Recognition and Localization Methods for Vision-Based Fruit Picking Robots: A Review. Frontiers in Plant Science 2020; 11:510, https://doi.org/10.3389/fpls.2020.00510.
24. Wang B, Lei Y, Li N, Li N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. IEEE Transactions on Reliability 2020; 69(1):401-412, https://doi.org/10.1109/TR.2018.2882682.
25. Wang B, Lei Y, Yan T, Li N, Guo L. Recurrent convolutional neural network: A new framework for remaining useful prediction of machinery. Neurocomputing 2020; 379:117-129, https://doi.org/10.1016/j.neucom.2019.10.064.
26. Wang F K, Mamo T. Hybrid approach for remaining useful life prediction of ball bearings. Quality and Reliability Engineering International 2019; 35(7): 2494-2505, https://doi.org/10.1002/qre.2538.
27. Wang Q, Zhao B, Ma H, Chang J, Mao G. A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion. Journal of Mechanical Science and Technology 2019; 33:2561–2571, https://doi.org/10.1007/s12206-019-0504-x.
28. Wang Y, Peng Y, Zi Y, Jin X, Tsui K L. A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem. IEEE Transactions on Industrial Informatics 2016; 12(3): 924-932, https://doi.org/10.1109/TII.2016.2535368.
29. Ye Z S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products. Applied Stochastic Models in Business and Industry 2015; 31(1):16-32, https://doi.org/10.1002/asmb.2063.
30. Zhang J, Wang P, Yan R, Gao R X. Long short-term memory for machine remaining life prediction. Journal of Manufacturing Systems 2018; 48(C): 78-86, https://doi.org/10.1016/j.jmsy.2018.05.011.
31. Zhang Y, Yang S, Li P, Hu X, Wang H. Marginalized Stacked Denoising Autoencoder with Adaptive Noise Probability for Cross Domain

Classification. IEEE Access 2019; 7:2169-3536, https://doi.org/10.1109/ACCESS.2019.2925811.

32. Zhu J, Chen N, Peng W. Estimation of Bearing Remaining Useful Life based on Multiscale Convolutional Neural Network. IEEE Transactions on Industrial Electronics 2019; 66(4):3208-3216, https://doi.org/10.1109/TIE.2018.2844856.
33. Zhu J, Chen N, Shen C. A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. Mechanical Systems and Signal Processing 2020; 139: 106602, https://doi.org/10.1016/j.ymssp.2019.106602.