IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 March 2024, accepted 12 May 2024, date of publication 20 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403476

RESEARCH ARTICLE

Rolling Element Bearing Degradation Prediction Using Dynamic Model and an Improved Adversarial Domain Adaptation Approach

SIMENG XU[1], CHENXING JIANG[1,2], CANGJIE YANG[1], HAIFENG CAO[1], ZHENGKAI SONG[1,2], RULIN ZHENG[1], SHIYI LU[1], HANZHANG XU[2], AND HENGCHENG ZHANG[2]

[1]College of Energy, Xiamen University, Xiamen 361102, China
[2]Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), Xiamen 361102, China

Corresponding authors: Chenxing Jiang (jiangchx@xmu.edu.cn) and Hengcheng Zhang (hc.zhang@xmu.edu.cn)

This work was supported by the Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM) under Grant HRTP-2022-51.

ABSTRACT Rolling element bearing degradation prediction is an important issue in rotating machinery. With the rapid development of artificial intelligence (AI), AI-based bearing degradation prediction has aroused extensive attention. However, current methods rely on whole life cycle data, which is quite difficult to acquire in real industrial scenarios. To solve this problem, a rotor-bearing dynamic model is built to generate simulation signals for a range of spall sizes, and an improved domain adversarial neural network is proposed to transfer degradation knowledge from simulation data to experimental data. To be specific, complete simulation data is used to pre-train a network for learning comprehensive degradation knowledge, and guides the extracted high-level features in the adversarial domain adaptation stage to align with it as an additional optimization item. The proposed approach is verified on bearing degradation datasets under different working conditions, and results show that the proposed approach can successfully predict bearing degradation progress only with some early stage experimental data.

INDEX TERMS Bearing degradation prediction, domain adversarial neural network, rotor-bearing dynamic model, prognostics and health management.

I. INTRODUCTION

Rolling element bearing (REB) is an important component of rotating machinery, which is widely used in transportation, manufacturing and electric power industry [1]. Due to contamination, corrosion, poor lubrication, long-time operation, etc., about 45∼55% rotating machinery failures are caused by REB [2]. Among them, race damage accounts for 90%, whereas rolling element and cage damage accounts for only 10% [3]. The damage is first initiated by sub-surface fatigue crack, with continuous load and impact between the rolling element and the race, sub-surface fatigue crack extends to the surface, resulting in a spall. Over time, the spall grows bigger in size, and eventually leading to catastrophic failure [4]. Premature replacement of REB may be quite expensive, on the other hand, the safety of personnel and machines cannot be risked. Therefore, degradation assessment of current fault severity and prognostics of future revolution is of great practical significance [5], [6].

At present, REB degradation prognostics methods can be divided into three main categories, which are physics model-based approaches, statistical model-based approaches and artificial intelligence-based (AI-based) approaches [7]. Physics model-based approaches describe the degradation process by establishing a comprehensive mathematical model. The most common ones are Paris-Erdogan (PE) model [8], [9] and Forman model [10], [11]. However, actual physical system is usually complex, and there are random

The associate editor coordinating the review of this manuscript and approving it for publication was Yongquan Sun.

factors such as background noise, speed fluctuations, etc., which make it difficult to be built accurately. Statistical model-based approaches achieve degradation prediction by fitting the monitoring parameters into statistical models such as autoregressive model [12], [13], wiener process model [14], [15], [16] and gamma process model [17], [18], which can describe the uncertainties of degradation process, but they usually require a great deal of expert knowledge. On the other hand, AI-based approaches attempt to find the potential degradation trends from a large amount of historical data, thereby reducing the reliance on physical and expert knowledge. With the rapid development of AI, deep learning has aroused extensive attention because they can extract effective features from raw vibration data automatically. Typical deep learning methods, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and their variants, have been used in predicting REB degradation. According to literature research, most of current REB degradation prediction study take remaining useful life (RUL) as the evaluation standard for degradation state. Li et al. [19] established a deep CNN to predict the RUL of REB, and the input data are prepared by time window approach. Wang et al. [20] proposed a framework with convolutional attention mechanism and temporal convolutional network (CAMTCN) to predict the RUL of REB, and Efficient Adaptive Shrinkage (EAS) model is utilized to eliminate the noise of vibration signal. Zheng et al. [21] proposed a multi-layer Long Short Term Memory (LSTM) network to predict RUL. As the LSTM processes time sequence only in forward direction, feature extraction ability is relatively poor, Jin et al. [22] employed bidirectional-LSTM (Bi-LSTM) to predict RUL, which learned features of time sequence in both forward and backward direction, and obtained better capability in time series modelling.

Although the above AI-based approaches can predict the remaining useful life of REB successfully, they have two prerequisites: (1) sufficient labelled degradation data; (2) training dataset and testing dataset meet the same distribution. Unfortunately, these assumptions are often unrealistic. Firstly, due to safety concerns, machines in real industrial scenarios couldn't be allowed to run with a severely degraded REB, vibration data of severe degradation stage could not be obtained. Moreover, REBs operating under different working conditions (different rotating speed, load, etc.) may occur data distribution differences.

To tackle these problems, domain adaptation has been introduced [23], [24], whose principle is to learn mutual features between the labelled source domain and the unlabelled target domain, as a result, the degradation knowledge can be transferred from one domain to another. Domain adaptation include discrepancy metric-based domain adaptation and adversarial training-based domain adaptation [25]. Discrepancy metric-based domain adaptation adopts distribution metrics such as Maximum Mean Discrepancy (MMD), Correlation Alignment (CORAL), Wasserstein Distance, etc. between source and target domains and minimizes them during training [26]. Xu et al. [27] established a domain adaptation network based on Bi-LSTM, and MMD was embedded to minimize the distribution difference between the source domain and the target domain. Results showed that domain adaptation based on MMD is effective for predicting RUL of defect bearings under different working conditions. In order to improve the robustness of the network, Rathore and Harsha [28] used multi-kernel MMD (MK-MMD) to realize domain adaptation, and verified the superiority of this method by comparing it with existing methods. In addition, adversarial training-based domain adaptation has also received widespread attention, which consists of three sub-networks, namely, feature extractor, RUL predictor and domain discriminator. Its main concept is to extract domain invariant features that confuses the domain discriminator, so RUL predictor can achieve good prediction results in both source and target domains [29]. Costa et al. [30] established an adversarial domain adaptation network, in which the feature extractor was LSTM, RUL predictor and domain discriminator were both simple fully connected networks, and has predicted the RUL of aero engine under different working conditions successfully.

Although the above domain adaptation methods achieved good prediction results on RUL cross different working conditions, complete run-to-failure experimental data is still required as the source domain. In addition, most of the current AI-based REB degradation studies use RUL to evaluate the degree of degradation, but there is no standard definition of the end of bearing useful life, which makes it upon individual explanation case by case.

To address these issues, a novel approach based on rotor-bearing dynamic model and an improved domain adaptation neural network (improved DANN) is proposed. The overview of the proposed approach is shown in **Fig. 1**. Main contributions are listed as follows:

1) To overcome the drawback of incomplete experimental data, a rotor-bearing dynamic model is built to obtain simulation signals of the whole life cycle.
2) To make neural network better learn the degradation characteristics of vibration signals, signal processing and segmentation is utilized to prepare appropriate input.
3) Instead of RUL, spall size is used to represent the degradation state more directly.
4) An improved DANN is proposed to address the issue of data distribution difference between the simulation data and the experimental data. Firstly, a Bi-LSTM network is employed as a pre-trained model to learn the whole life cycle degradation knowledge of the source simulation data. Then, pre-trained feature serves as a reference to guide the extracted high-level feature in adversarial domain adaptation stage to align with it, thus retaining the degradation knowledge and alleviating the negative effect of incomplete experimental data. The effectiveness of the proposed approach is

verified on bearing degradation datasets under different working conditions.

The rest of this article are organized as follows. In **Section II**, dynamic modelling of rotor-bearing is introduced, and simulation signal and experiment signal are contrasted from time domain and squared envelope spectrum. **Section III** introduces the signal processing and segmentation method to obtain appropriate input data for the subsequent neural network. **Section IV** introduces the improved DANN, and **Section V** presents the experimental results and discussion. Finally, conclusions are given in **Section VI**.

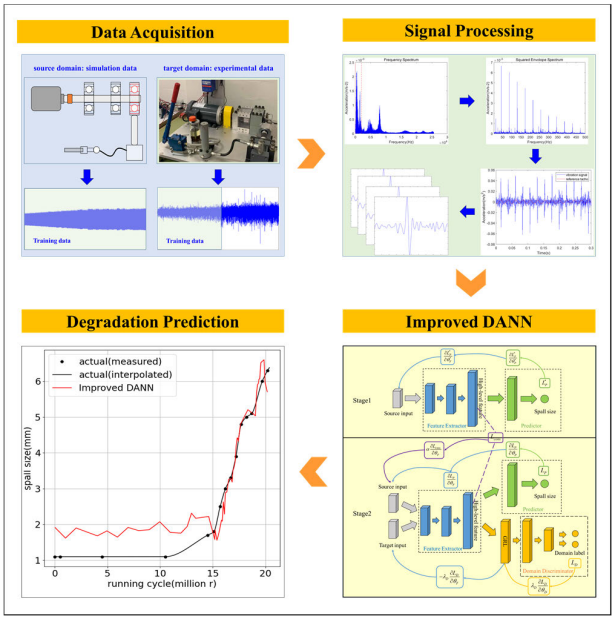


FIGURE 1. Overview of the proposed approach.

II. DYNAMIC MODELING OF ROTOR-BEARING SYSTEM

To simulate the vibration response of the defective bearing on test rotor, rotor-bearing system dynamic model is established, as is shown in **Fig. 2, 3**, where the former is the diagram of rotor system, containing motor, rotor shaft, coupling, two support bearings and a test bearing; the latter is the diagram of the test bearing with an outer race defect.

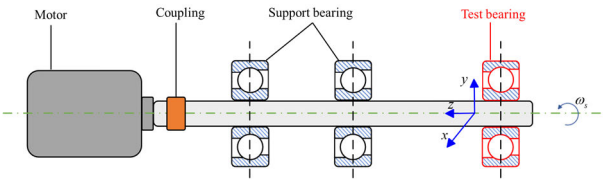


FIGURE 2. Diagram of rotor system.

A. ROTOR SYSTEM MODEL

Based on finite element method, the rotor is divided into several beam elements, each has two nodes. Each node

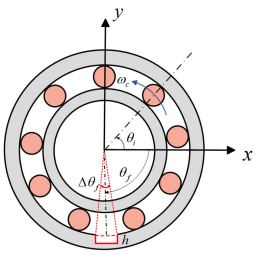


FIGURE 3. Diagram of test bearing with an outer race defect.

considers 6 degrees of freedom (6-DOF), which includes 3 directions of translations and 3 directions of rotations. The motion equation of beam element can be expressed as:

$$m_s \ddot{q}_s + c_s \dot{q}_s + k_s q_s = 0 \tag{1}$$

where q_s is the generalized displacement vector of beam element, m_s, k_s and, c_s are the mass, damping and stiffness of the beam element, respectively.

$$q_s = \{x_s, y_s, z_s, \theta_{x_s}, \theta_{y_s}, \theta_{z_s}, x_{s+1}, y_{s+1}, z_{s+1}, \theta_{x_{s+1}}, \theta_{y_{s+1}}, \theta_{z_{s+1}}\} \tag{2}$$

Supporting bearings and coupling are considered as springs with certain stiffness and damping, and their mass are ignored. By assembling all elements, the overall equation of motion can be expressed as:

$$M\ddot{q} + C\dot{q} + Kq = F \tag{3}$$

where M, C, K are the mass, damping and stiffness matrix of the system, respectively, details can be found in [31]. The excitation of the system, namely, the bearing nonlinear contact force, will be described in the next subsection.

B. BEARING MODEL

The test bearing is a deep groove ball bearing, simplified to a 3-DOF model, namely, horizontal DOF, vertical DOF and rotational DOF. The outer race is fixed on the bearing housing, and the inner race is rotating with the shaft. Nonlinear factors, including nonlinear contact force between the rolling elements and race, time-varying stiffness and radial clearance are considered [32].

Angular speed of the shaft is $\omega_s = 2\pi f_s$, then angular speed of the cage is obtained according to the geometric relationship:

$$\omega_c = \frac{\omega_s}{2}\left(1 - \frac{D_b}{D_p}\right) \tag{4}$$

where D_p is the pitch diameter, D_b is the rolling element diameter.

N_b rolling elements are uniformly arranged in the cage. Considering the slippage of rolling elements, the angular position of the ith rolling element at time t can be expressed

as [33]:

$$\theta_i = \frac{2\pi\ (i-1)}{N_b} + \omega_c t + (0.5 - rand) \times 0.01 \quad (i = 1, 2, \ldots, N_b) \quad (5)$$

For the ith rolling element, the contact deformation between the rolling element and race can be expressed as:

$$\delta_i = x_s \cos\theta_i + y_s \sin\theta_i - c - h \quad (6)$$

where, x_s and y_s are the relative displacement of the inner and outer race in x and y direction, respectively. c is the radial clearance. h is the additional deformation introduced by outer race defect, which changes with the angle position of the rolling element [34]:

$$h = \begin{cases} \min(h_0, r_b^2 - \sqrt{r_b^2 - 0.25R_o^2(\theta_i - \theta_f + 0.5\Delta\theta_f)^2}) \\ \quad \text{if } \theta_f - 0.5\Delta\theta_f < \theta_i \le \theta_f \\ \min(h_0, r_b^2 - \sqrt{r_b^2 - 0.25R_o^2(\theta_i - \theta_f - 0.5\Delta\theta_f)^2}) \\ \quad \text{if } \theta_f < \theta_i \le \theta_f + 0.5\Delta\theta_f \\ 0 \\ \quad \text{if } \theta_f + 0.5\Delta\theta_f \le \theta_i \le \theta_f - 0.5\Delta\theta_f \end{cases} \quad (7)$$

where h_0 is the depth of the defect, r_b is the radius of the rolling element, R_o is the radius of the outer race, θ_f is the angle position of the defect centre, $\Delta\theta_f$ is the span angle of the defect.

Therefore, according to the Hertzian contact theory, the contact force between the ith rolling element and the race can be expressed as [35]:

$$F_i = K\delta_i^{1.5} \quad (8)$$

where K is the load-deflection factor, which is related to the bearing geometry and material properties.

By projecting and summing up the Hertzian contact force of each rolling element in x and y direction, the total bearing Hertzian contact force in x and y direction can be expressed as:

$$\begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N_b} F_i \cos\theta_i \\ \sum_{i=1}^{N_b} F_i \sin\theta_i \end{bmatrix} \quad (9)$$

In addition, to simulate the randomness of the vibration signal, Gaussian white noise is also added to the simulation signal [32]. Simulation signals of different sizes can be obtained by specifying the defect span angle $\Delta\theta_f$. The rotor-bearing dynamic model is established by Matlab and solved by the ordinary differential equation solver in Simulink.

C. SIMULATION RESULTS AND DISCUSSION

The acceleration responses and the squared envelope spectrum of the simulation signal and the experiment signal are shown in **Fig. 4, 5**. It can be seen from **Fig. 4** that both experiment signal and simulation signal have periodic impact components, whose interval is the reciprocal of the Ball Pass Frequency Outer race (BPFO). However, the impact characteristics of the simulation signal are more significant than that of the experiment signal, because the experiment signal has a naturally extended fault, the entry/exit event are weak, which is easily masked by the background noise. As is shown in **Fig. 5**, for the simulation signal, the main frequency component is the BPFO (21.5 Hz in theoretical) and its harmonics, but for the experiment signal, besides BPFO, there are other frequency components that may be aroused by other components of the rotor system, which are neglected in dynamic modelling. This also leads to a certain distribution difference between simulation data and experimental data.

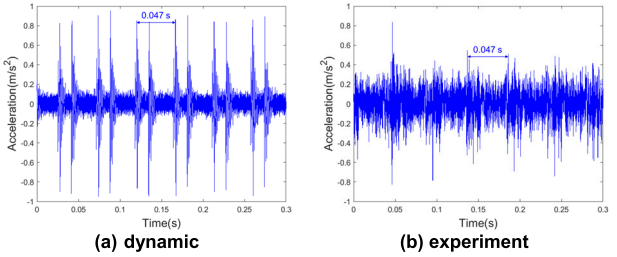


FIGURE 4. Acceleration responses.

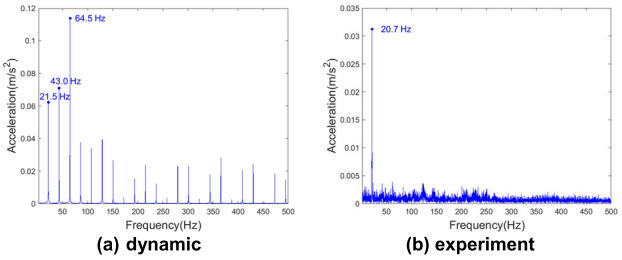


FIGURE 5. Squared envelope spectrum.

III. SIGNAL PROCESSING AND SEGMENTATION

Before introducing the domain adaptation approach, time-domain signal is pre-processed and truncated by ball-passing-defect period to make sure that each signal section is of the same length and contains spall size information to be used as the input of the following neural network. This is different from most previous studies, which directly cut raw vibration signals into equal length segments without considering the characteristics of the fault signals, resulting in an unclear physical relationship between such input data and the severity of degradation.

Since the experiment signal is a naturally extended fault, the defect induced impulses are not clear, so it is necessary to be filtered by the most impulsive frequency band firstly. Then order tracking is employed to remove the speed fluctuation of the shaft and the slippage of rolling elements by resampling time-domain signal into angular-domain. Specifically, squared envelope spectrum is obtained and peak-search is employed to identify $BPFO_{actual}$ near the theoretical value of $BPFO$, then it is band-pass filtered around $BPFO_{actual}$ with a tolerance of $\pm 10\%$ (to consider

speed fluctuation) to construct a reference tacho signal, the reference tacho signal is unwrapped to obtain the phase angle, and the vibration signal with constant time interval is interpolated and resampled by constant phase steps into angular domain [36]. Moreover, to ensure proper angular domain resolution, the angular signal is resampled (the sampling number of each ball-passing-defect period is 200), and is segmented by this period as input data for the following neural network.

In addition, in order to make the input data with different degradation degree to have the same vibration scale, so that the neural network can converge quickly, the signal is normalized to [−1,1]:

$$x' = \frac{2(x - x_{\min})}{(x_{\max} - x_{\min})} - 1 \tag{10}$$

The whole procedure of signal processing and segmentation is shown in **Fig. 6**:

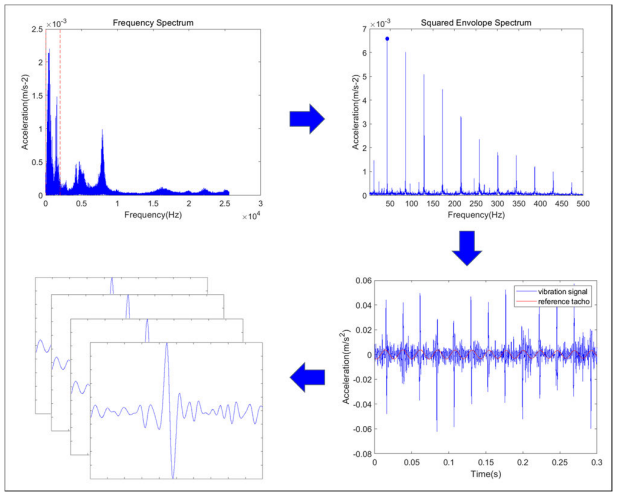


FIGURE 6. Procedure of signal processing and truncating.

IV. THE IMPROVED DANN

A. PROBLEM DEFINITION

Since the dynamic model is a simplification of the actual rotor-bearing system, inevitably, distribution differences exist between the simulation data and the experimental data. To eliminate its influence by the most extent, domain adaptation is employed, which generalizes the knowledge from simulation data to experimental data by learning domain invariant features. To facilitate the narration, relevant definition and symbolic expression are introduced firstly.

Let $D_s = \{\mathbf{x}_s^i, y_s^i\}_{i=1}^{n_s}$ denotes the source domain, which is the simulation dataset, $\mathbf{x}_s^i$ and y_s^i represent the ith input data and the corresponding spall size label, n_s represents the number of training samples of the source domain. Let $D_t = \{\mathbf{x}_t^i\}_{i=1}^{n_t}$ denotes the target domain, which is the experimental dataset, $\mathbf{x}_t^i$ represents the ith input data, n_t represents the number of training samples of the target domain. There exist distribution differences between source domain and target domain, i.e., $P(\mathbf{x}_s) \neq P(\mathbf{x}_t)$. It should be noted that the source domain contains whole life cycle simulation data, whereas the target domain only contains early degradation stage experimental data, making the target domain incomplete.

B. DOMAIN ADVERSARIAL NEURAL NETWORK

Domain Adversarial Neural Network (DANN) is first proposed by Ganin et al. [29]. The basic idea is to find task-discriminative and domain-invariant features through adversarial training, thereby achieving good prediction results in target domain without knowing the label of target data. DANN consists of three sub-networks, namely feature extractor $G_F(\cdot;\theta_F)$, with parameter θ_F, spall size predictor $G_P(\cdot;\theta_P)$, with parameter θ_P, and domain discriminator $G_D(\cdot;\theta_D)$, with parameter θ_D. For source data and target data, high-level features are firstly extracted by the feature extractor, then spall size predictor is used to predict the corresponding spall size from the source feature. On the other hand, source features and target features are also mapped to the domain label through the domain discriminator.

There are two goals in the training process of DANN. The first one is that the spall size should be predicted precisely. By optimizing the parameters of the feature extractor and the spall size predictor, the regression loss is minimized to ensure that the high-level features extracted by the feature extractor can reflect the spall size, in other words, the extracted high-level features are task-discriminative. The other goal is to find domain-invariant features, which means that the high-level features extracted by the feature extractor are the mutual degradation features that the domain discriminator cannot distinguish whether they are from the source domain or the target domain. Thus, the feature extractor updates its parameters by maximizing the domain discriminator loss, in the opposite direction to the parameter update of the domain discriminator. In order to simplify the model training process, Gradient Reversal Layer (GRL) is introduced between the feature extractor and the domain discriminator. During backpropagation, the gradient is multiplied by a negative constant through GRL before the feature extractor. Whereas in the forward propagation, the parameters remain unchanged. Mathematically, GRL can be treated as a pseudo-function $R_\lambda(\mathbf{x})$, its forward and backward propagation can be described as:

$$R_\lambda(\mathbf{x}) = \mathbf{x} \tag{11}$$

$$\frac{dR_\lambda}{d\mathbf{x}} = -\lambda\mathbf{I} \tag{12}$$

where $\mathbf{I}$ is an identity matrix.

Therefore, the total loss function of DANN can be expressed as:

$$\begin{aligned} L(\theta_F, \theta_P, \theta_D) = &\sum_{\mathbf{x}_i \in D_s} L_P(G_P(G_F(\mathbf{x}_i;\theta_F);\theta_P), y_i) \\ -\lambda &\sum_{\mathbf{x}_i \in (D_s \cup D_t)} L_D(G_D(G_F(\mathbf{x}_i;\theta_F);\theta_D), d_i) \end{aligned} \tag{13}$$

where, L_P and L_D represent the regression loss of the spall size predictor and the classification loss of the domain discriminator, respectively. λ is the weight factor that balances these two losses. $\mathbf{x}_i$ and y_i represent the ith input sample and the spall size label. d_i represents the domain label, where source domain is 0 and target domain is 1.

During training, the loss function L is optimized by seeking a saddle point solution $\hat{\theta}_F, \hat{\theta}_P, \hat{\theta}_D$ of the min-max problem,

$$(\hat{\theta}_F, \hat{\theta}_P) = \arg\min_{\theta_F, \theta_P} L(\theta_F, \theta_P, \hat{\theta}_D) \tag{14}$$

$$\hat{\theta}_D = \arg\max_{\theta_D} L(\hat{\theta}_F, \hat{\theta}_P, \theta_D) \tag{15}$$

At the saddle point, the parameters of the spall size predictor, $\hat{\theta}_P$, minimize the regression loss; the parameters of the domain discriminator, $\hat{\theta}_D$, minimize the domain classification loss; and the parameters of the feature extractor, $\hat{\theta}_F$, minimize the regression loss and maximize the domain classification loss simultaneously.

The saddle point can be found using gradient descent method with a learning rate μ, which can be described as:

$$\theta_F \leftarrow \theta_F - \mu\left(\frac{\partial L_P^i}{\partial \theta_F} - \lambda \frac{\partial L_D^i}{\partial \theta_F}\right) \tag{16}$$

$$\theta_P \leftarrow \theta_P - \mu \frac{\partial L_P^i}{\partial \theta_P} \tag{17}$$

$$\theta_D \leftarrow \theta_D - \mu \frac{\partial L_D^i}{\partial \theta_D} \tag{18}$$

C. THE IMPROVED DANN

In real industrial scenarios, it is often unrealistic to collect degradation data throughout its life time, especially for safety-critical REB, where complete failure will lead to serious consequences. Therefore, to predict failure progression only with some early degradation stage experimental data is of vital significance. DANN directly aligns all training data of the source domain and the target domain, the lack of severe degradation stage target data may lead to wrong alignment, as is shown in **Fig. 7**.

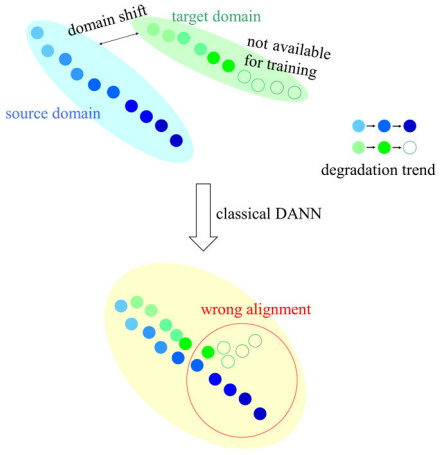


FIGURE 7. Diagram of wrong alignment.

To deal with this problem, an improved DANN is proposed, which consists of two stages. Stage 1 is pre-training on source domain, which is the whole life cycle simulation dataset. A regression network is established by combing the feature extractor and spall size predictor to learn the mapping relationship between the input data and their corresponding spall sizes. Thus, the pre-trained features cover the degradation knowledge of the whole life cycle. Stage 2 is the adversarial domain adaptation stage. Based on classical DANN, a consistency-based regularization term is added to force the high-level features to align with the source pre-trained features. In this way, the domain adversarial neural network not only learns domain-invariant features, but also retains the bearing degradation knowledge of the whole life cycle, thus reducing wrong alignment. The overview of the improved DANN is shown in **Fig. 8**.

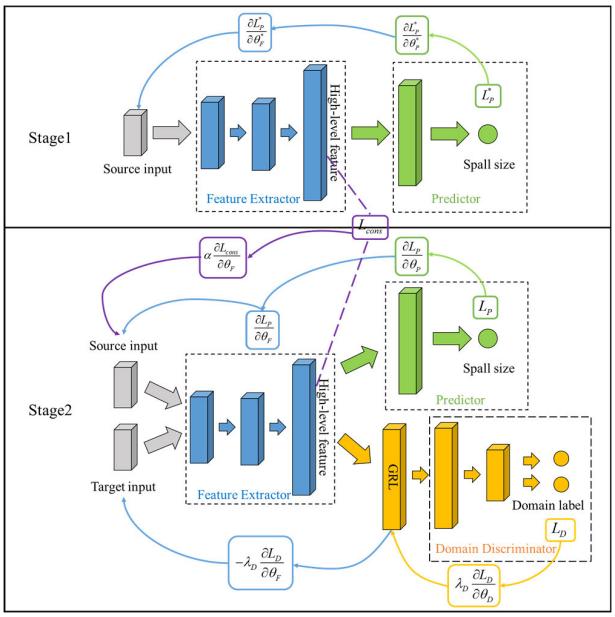


FIGURE 8. Overview of the improved DANN.

Stage 1 (Pre-Training): In Stage 1, whole life cycle simulation data is used for training, the pre-training network contains feature extractor $G_F^*(\cdot;\theta_F^*)$,with parameter θ_F^*, and spall size predictor $G_P^*(\cdot;\theta_P^*)$, with parameter θ_P^*. Obviously, this is a classical regression network, the loss function L_P^* is mean square error (MSE):

$$L_P^* = \frac{1}{n_s}\sum_{i=1}^{n_s}(y_i - \hat{y}_i)^2 \tag{19}$$

where, y_i and $\hat{y}_i$ are the spall size label and predicted spall size respectively.

After training on source simulation data, the pre-trained regression network learns the mapping relationship between the input data and the spall size of the whole running cycles, which can guide the feature alignment in the subsequent adversarial domain adaptation stage.

Stage 2 (Adversarial Domain Adaptation): In this stage, degradation knowledge will be generalized from the source simulation domain to the target experiment domain. It should be noted that the experimental data lacks severe degradation data, which may have a negative effect on the classical DANN. Luckily, the pre-trained regression network learns the regression relationship between the input data and the corresponding spall size of the whole life cycle, which means the pre-trained high-level features contain complete degradation knowledge. Therefore, in Stage 2, the extracted high-level features are forced to follow the pre-trained features as an additional optimization objective, which can be also described as consistency-based regularization. To achieve this purpose, L1 loss is introduced to measure the alignment between the extracted source features in Stage 2 and the pre-trained features:

$$L_{cons} = \frac{1}{n}\left(\sum_{i=1}^{n} \left|f(x_i) - f^*(x_i)\right|\right) \tag{20}$$

where $f(x_i)$ is the high-level features in Stage 2, $f^*(x_i)$ is the pre-trained feature, and n is the number of features.

Therefore, the total loss function of the adversarial domain adaptation stage is given as:

$$L_{total} = L_P - \lambda L_D + \alpha L_{cons} \tag{21}$$

$$L_p = \frac{1}{n_s}\sum_{i=1}^{n_s} (y_i - \hat{y}_i)^2 \tag{22}$$

$$L_D = \frac{1}{n_s + n_t}\sum_{i=1}^{n_s+n_t} \left(d_i \log \frac{1}{\hat{d}_i} + (1 - d_i)\log \frac{1}{1 - \hat{d}_i}\right) \tag{23}$$

where L_P is the spall size predictor loss, L_D is the domain adversarial loss, λ and α are the weight factors of domain adversarial training and consistency-based regularization, respectively.

V. EXPERIMENTAL STUDY

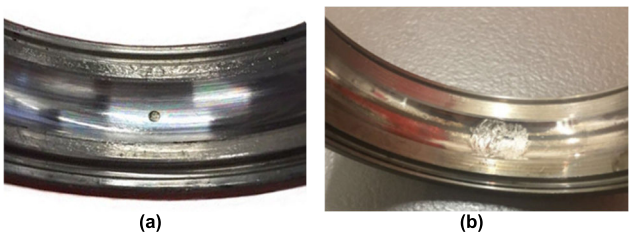
A. EXPERIMENTAL DESIGN AND SETUP

1) EXPERIMENTAL DESIGN

To validate the bearing degradation prediction performance of the proposed approach, UNSW naturally extended bearing defect dataset [37] is used in this study, since it is the few public dataset that provides exact spall size labels during the run-to-failure process. The test bench is shown in **Fig. 9**, which illustrates that the shaft rotor is driven by a phase AC motor through a jaw coupling, and two bearings are used to support it. The test bearing is installed at the free end of the shaft, and its outer race is held by a floating housing case, which is loaded horizontally by a hydraulic system push rod. Two accelerometers are mounted at the bearing housing case in both horizontal and vertical direction, they are sampled at 51.2 kHz for 12 s periodically.

To speed up the degradation process and control the position of the spall, a small round defect located at the centre of the loading area is seeded on the outer race by electrical discharge machining, as is shown in **Fig. 10 (a)**. The test bearing is periodically disassembled to measure the spall size and observe the surface morphology, and then reassembled to continue the run-to-failure process until the spall size is larger than 7 mm or other failure such as cage fracture occurs, the extended spall at the end of the run-to-failure experiment is shown in **Fig. 10 (b)**. Therefore, signals around the disassemble-reassemble point are with the spall label, whereas other signals in between are unlabelled.

FIGURE 9. UNSW dataset test bench.



(a) (b)

FIGURE 10. Outer race defect: (a) initial seeded spall; (b) extended spall at the end of the experiment.

TABLE 1. Domain adaptation tasks.

Task	Working condition
1	load:10.5 kN, shaft speed 6 Hz
2	load:10.5 kN, shaft speed 12 Hz
3	load:10.5 kN, shaft speed 15 Hz
4	load:10.5 kN, shaft speed 20 Hz
5	load:7.0 kN, shaft speed 6 Hz
6	load:7.0 kN, shaft speed 12 Hz
7	load:7.0 kN, shaft speed 15 Hz
8	load:7.0 kN, shaft speed 20 Hz

Eight groups of domain adaptation tasks with different working conditions are employed, as detailed in **Table 1**. Corresponding simulation data are generated for each task

as the source domain, with horizontal experimental data as the target domain (since the load is in the horizontal direction). Simulation dataset and experimental dataset are both divided into training set, validating set and testing set. It should be noted that the experimental data of the severe degradation stage (when the spall size is larger than 4.0mm) is not available in training, which coincides with the need for bearing degradation prediction in real industrial scenarios.

2) PARAMETER SETTING

For simulation and experiment signals, signal processing methods in **Section III** are firstly employed to generate input data for training. Each input data contains one ball-pass-defect period, and its length is 200.

To better learn the mapping relationship between the input data and the corresponding spall size, Bi-LSTM network is employed for feature extraction. LSTM is a variant of RNN. By introducing gating mechanisms (input gate, forget gate, output gate), it solves the problem that RNN is prone to incur gradient vanishing or exploding when processing long time series [38]. Furthermore, Bi-LSTM introduces backward LSTM to process the input time series in a reverse direction, which can capture both past and future information, thereby having stronger feature extraction ability for time sequences [39]. Relevant studies have proved that Bi-LSTM achieves good prediction in bearing degradation prediction [22], [40], [41].

As for the network structure, for the feature extractor, firstly two Bi-LSTM layers are adopted to perform feature extraction, and the number of hidden layer units is 32. To prevent overfitting, dropout rate is set as 0.3. Then, after flattening the output through a flatten layer, the dimension of high-level features is reduced to 128 through a fully connected layer (FC layer). The spall size predictor receives the output from feature extractor as the input and maps it to a single output value, that is the spall size. The domain discriminator consists of two FC layers, each with 128, 2 hidden units. All FC layers are followed by the ReLU activation function. The detailed network structure is shown in **Fig. 11**.

During training, Adam optimizer is utilized to find optimal parameters, and the learning rate is set to $1e^{-4}$. Batch size is set to 128, and the maximum number of iterations is 10000. For the weight factor, the domain adversarial weight factor adopts a strategy that increases with the number of iterations [29]:

$$\lambda = \frac{2}{1 + \exp(-10p)} - 1 \tag{24}$$

where p is the ratio of current iteration to maximum iteration.

The consistency regularization weight factor is determined by grid search. The experiments are conducted on a cloud server with NVIDIA Tesla A100 GPU, and the codes are implemented on the GPU-based PyTorch platform. All experiments are carried out 5 times and the average prediction is taken as the result.

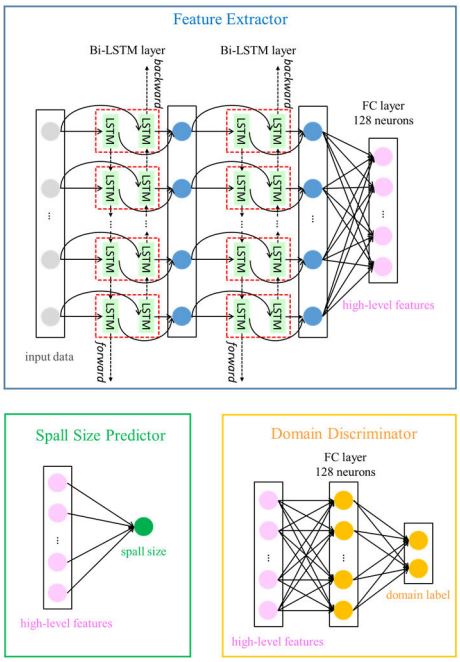


FIGURE 11. Detailed structure of the sub-networks.

3) EVALUATION METRICS

To evaluate the spall size prediction performance of different approaches, commonly used evaluation metric in regression network, Root Mean Squared Error (RMSE) is employed [42]. Moreover, since the key subject in bearing degradation prediction is to avoid security incidents, early prediction is preferable compared with late prediction. In other words, if the predicted spall size is smaller than the actual spall size, it will get a higher penalty score [43]. RMSE and Score function can be expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{25}$$

$$Score = \begin{cases} \sum_{i=1}^{N} e^{-\frac{(\hat{y}_i - y_i)}{10}} - 1 & \text{if } \hat{y}_i < y_i \\ \sum_{i=1}^{N} e^{\frac{(\hat{y}_i - y_i)}{13}} - 1 & \text{if } \hat{y}_i \geq y_i \end{cases} \tag{26}$$

where $\hat{y}_i$ is the predicted spall size, y_i is the spall size label.

B. EXPERIMENT RESULTS AND DISCUSSION

1) COMPARISON OF DIFFERENT FEATURE EXTRACTION ARCHITECTURES

Spall size prediction performances of different feature extraction architectures are firstly discussed. In our proposed method, Bi-LSTM network is used for feature extraction. Other popular feature extraction architectures are also considered, including: Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN). In other words, the feature extraction archi-

tecture is replaced by MLP, CNN, RNN layers respectively, the rest are consistent with the proposed one.

As for the spall size label, since only a limited number of experiment signals has been measured for spall size, the rest spall sizes are obtained by interpolation.

The comparison results of different architectures are shown in **Fig. 12**, and the average performance are shown in **Table 2, 3**, indicating that the simple MLP-DANN has the worst feature extraction ability in vibration data, CNN-DANN improves the prediction performance by learning local features of the vibration data, but is not enough to capture long-term dependence. On the other side, through direct modelling of the vibration data, in general, RNN-DANN provides better prediction, but RNN may incur gradient vanishing or exploding problems. Whereas Bi-LSTM-DANN introduces gating mechanisms and processes input data from both forward and backward directions, the long-term dependence can be captured more effectively, and it has the best performance in all given working conditions, which can prove that Bi-LSTM is a suitable feature extraction network in this study.

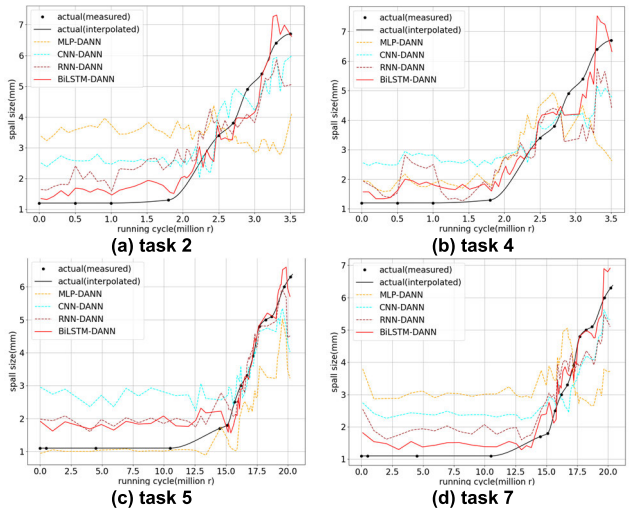


FIGURE 12. Spall size prediction results of different feature extraction architectures.

2) COMPARISON OF DIFFERENT DOMAIN ADAPTATION APPROACHES

Moreover, to verify the superiority of the proposed method, different domain adaptation approaches are also employed for comparison.

(a) Baseline: This method is training on the source domain (simulation dataset) only, and is directly tested on the target domain. The network structure is the same as the first stage of the proposed method, which is a typical regression network and the loss function is the spall size prediction loss L_P. Since no domain adaptation approach is employed, this method can be treated as a baseline.

TABLE 2. RMSE of different feature extraction architectures.

Task	MLP-DANN	CNN-DANN	RNN-DANN	Bi-LSTM-DANN
1	2.1791	1.0267	1.5722	**0.4676**
2	2.1472	1.0443	0.9145	**0.4586**
3	1.5651	1.3912	0.8785	**0.3799**
4	1.3811	1.2356	1.0773	**0.5378**
5	1.2155	1.0505	0.6161	**0.4899**
6	1.6528	1.2236	0.9654	**0.6532**
7	1.8491	0.9377	0.8075	**0.4037**
8	1.7064	0.8482	0.9640	**0.6833**
Average	1.7120	1.0947	0.9744	**0.5093**

TABLE 3. Score of different feature extraction architectures.

Task	MLP-DANN	CNN-DANN	RNN-DANN	Bi-LSTM-DANN
1	9.6226	4.0331	6.5033	**1.9298**
2	9.4834	4.2555	3.6781	**1.6582**
3	7.1081	5.7344	3.9179	**1.5413**
4	5.2246	4.8951	4.0043	**1.9851**
5	5.2768	3.6964	2.0566	**1.7220**
6	7.5339	5.3081	4.1344	**2.7694**
7	8.8580	4.2067	3.5259	**1.5199**
8	7.0186	3.0529	3.3706	**2.2768**
Average	7.5158	4.3978	3.8989	**1.9253**

(b) MK-MMD: This method is a discrepancy metric-based domain adaptation. MK-MMD is employed to calculate the distribution differences between the source and target domain. The network structure of feature extractor and spall size predictor are the same as the proposed method, and the loss function includes spall size prediction loss L_P and MMD loss.

(c) DANN: This method is the classical adversarial training-based domain adaptation, and the network structure of feature extractor, spall size predictor and domain discriminator are the same as that of the proposed one in the second stage, the only difference is that consistency-based regularization is not considered. The loss function includes prediction loss L_P and domain discriminator loss L_D.

(d) Improved DANN: This method is the proposed approach, which consists of two stages. Stage 1 is pre-training on the whole life cycle source domain. Stage 2 is the adversarial training-based domain adaptation, and consistency-based regularization is added as an additional optimization item to force the extracted high-level features to align with the pre-trained features.

Fig. 13 indicates that the prediction performance of the baseline is the worst, because there exits domain difference between the simulation and the experimental datasets, thus domain adaptation is necessary. Comparing MK-MMD and DANN, in general, the RMSE and Score of DANN is slightly smaller than that of MK-MMD, indicating that the domain adaptation based on adversarial training is better than domain adaptation based on MK-MMD in this study.

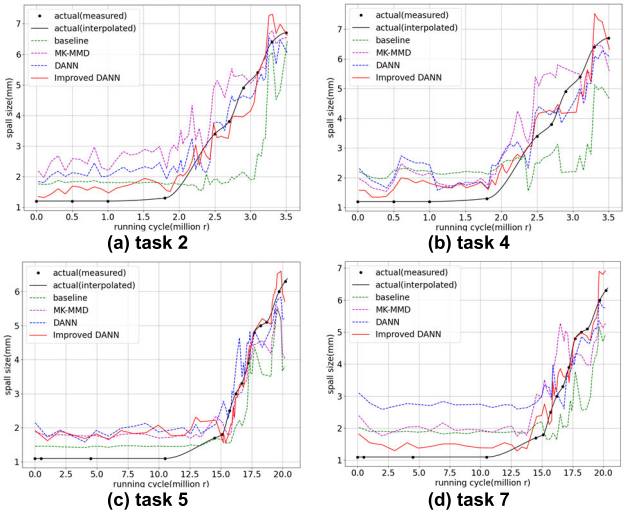


FIGURE 13. Spall size prediction results of different domain adaptation approaches.

TABLE 4. RMSE of different domain adaptation approaches.

Task	Baseline	MK-MMD	DANN	Improved DANN
1	0.9599	0.5586	0.8068	**0.4676**
2	1.4409	1.1577	0.7202	**0.4586**
3	1.3937	0.5543	0.4545	**0.3799**
4	1.4563	0.9863	0.7989	**0.5378**
5	0.9143	0.6681	0.6399	**0.4899**
6	1.6905	1.2679	0.7450	**0.6532**
7	1.2366	1.0265	1.0336	**0.4037**
8	1.1603	1.0251	0.9708	**0.6833**
Average	1.2816	0.9056	0.7712	**0.5093**

TABLE 5. Score of different domain adaptation approaches.

Task	Baseline	MK-MMD	DANN	Improved DANN
1	4.1391	2.4011	3.1560	**1.9298**
2	5.9742	4.1922	2.8243	**1.6582**
3	6.3882	2.3050	1.8435	**1.5413**
4	6.0814	3.5608	3.0396	**1.9851**
5	3.7798	2.5028	2.2638	**1.7220**
6	8.2896	5.2145	3.0184	**2.7694**
7	5.9971	4.3095	3.9792	**1.5199**
8	4.8991	4.1761	3.3930	**2.2768**
Average	5.6936	3.5828	2.9397	**1.9253**

However, the prediction error of the severe degradation stage is larger than that of the early degradation stage. It is because target domain training data only contains early degradation stage experimental data, the prediction effectiveness of the severe degradation stage cannot be guaranteed. Whereas the proposed improved DANN pre-train a regression network on simulation data to learn the whole life cycle degradation knowledge, and guides the feature extractor in adversarial domain adaptation stage to retain the degradation knowledge, thereby capturing the degradation trend of the severe degradation stage of the experimental data to the greatest extent. As seen from **Table 4, 5**, the proposed method has the best performance, comparing with DANN, the average RMSE and Score decreased by 33.96% and 34.51%, respectively.

VI. CONCLUSION

In this article, a rolling element bearing degradation prediction approach based on dynamic model and improved DANN is proposed, which is proved to be able to predict the spall size growth only with some early degradation stage experiment data.

Firstly, the rotor-bearing dynamic model is constructed to provide bearing simulation signals of a range of spall sizes. Secondly, to have uniform input for the neural networks, simulation and experiment signals are filtered and truncated into ball-pass-defect segments. What's more, since only early fault stage experimental data is available for training, which makes the target domain incomplete, a two-stage improved DANN is proposed. Degradation knowledge of the whole life cycle is learnt by training on labelled simulation data in the first stage of pre-training, and the high-level features are relatively aligned with the pre-trained features in the second stage of adversarial domain adaptation. The proposed approach is validated under different working conditions, and is compared with different feature extraction architectures and popular domain adaptation approaches, proving its superiority on bearing degradation prediction. In summary, this study gives an alternative option for bearing degradation prediction when only early-stage degradation data is available. It is applicable in real industrial scenarios to avoid sudden breakdown and provides a guidance for making maintenance plan.

Despite the promising results, there are still many shortcomings. From the perspective of method verification, the proposed method is verified on rotor-bearing test bench, fault impulses are relatively obvious, further verification on defect bearing inside the rotating machinery is urgently needed. In addition, from the perspective of the machine learning model, a more advanced and lightweight model can be considered, so as to balance the degradation prediction performance and the time and resource cost of model training. In addition, the alignment of the degradation features of the simulation data and experimental data is relatively rough, methods such as contrastive learning and local domain discriminator can be further introduced to realize the fine-grained alignment of the simulation data and experimental data.

REFERENCES

[1] T. A. Harris and W. J. Anderson, "Rolling bearing analysis," *J. Lubrication Technol.*, vol. 89, no. 4, p. 521, 1967.
[2] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribol. Int.*, vol. 96, pp. 289–306, Apr. 2016, doi: 10.1016/j.triboint.2015.12.037.
[3] D. Bently, "Predictive maintenance through the monitoring and diagnostics of rolling element bearings," *Appl. Note*, vol. 44, pp. 2–8, Jun. 1989.
[4] M. R. Hoeprich, "Rolling element bearing fatigue damage propagation," *J. Tribol.*, vol. 114, no. 2, pp. 328–333, Apr. 1992, doi: 10.1115/1.2920891.

[5] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, nos. 1–2, pp. 314–334, Jan. 2014, doi: 10.1016/j.ymssp.2013.06.004.

[6] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *IEEE Access*, vol. 7, pp. 122644–122662, 2019, doi: 10.1109/ACCESS.2019.2938227.

[7] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.

[8] Y. Qian, R. Yan, and R. X. Gao, "A multi-time scale approach to remaining useful life prediction in rolling bearing," *Mech. Syst. Signal Process.*, vol. 83, pp. 549–567, Jan. 2017, doi: 10.1016/j.ymssp.2016.06.031.

[9] Y. Sheng, H. Liu, and J. Li, "Bearing performance degradation assessment and remaining useful life prediction based on data-driven and physical model," *Meas. Sci. Technol.*, vol. 34, no. 5, May 2023, Art. no. 055002, doi: 10.1088/1361-6501/acb374.

[10] R. G. Forman, "Study of fatigue crack initiation from flaws using fracture mechanics theory," *Eng. Fract. Mech.*, vol. 4, no. 2, pp. 333–345, Jun. 1972.

[11] C. H. Oppenheimer and K. A. Loparo, "Physically based diagnosis and prognosis of cracked rotor shafts," *Proc. SPIE*, vol. 4733, pp. 122–132, Jun. 2002.

[12] Y. Qian, R. Yan, and S. Hu, "Bearing degradation evaluation using recurrence quantification analysis and Kalman filter," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2599–2610, Nov. 2014, doi: 10.1109/TIM.2014.2313034.

[13] P. Ding, M. Jia, and X. Yan, "Stationary subspaces-vector autoregressive with exogenous terms methodology for degradation trend estimation of rolling and slewing bearings," *Mech. Syst. Signal Process.*, vol. 150, Mar. 2021, Art. no. 107293, doi: 10.1016/j.ymssp.2020.107293.

[14] N. Li, Y. Lei, T. Yan, N. Li, and T. Han, "A Wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 2092–2101, Mar. 2019, doi: 10.1109/TIE.2018.2838078.

[15] Q. Liu, Y. Zhang, X. Si, and Z. Fan, "DLVR-NWP: A novel data-driven bearing degradation model for RUL estimation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023, doi: 10.1109/TIM.2023.3244839.

[16] Y. Song, S. Xu, and X. Lu, "A sliding sequence importance resample filtering method for rolling bearings remaining useful life prediction based on two wiener-process models," *Meas. Sci. Technol.*, vol. 35, no. 1, Jan. 2024, Art. no. 015019, doi: 10.1088/1361-6501/acffe3.

[17] J. Lawless and M. Crowder, "Covariates and random effects in a gamma process model with application to degradation and failure," *Lifetime Data Anal.*, vol. 10, no. 3, pp. 213–227, Sep. 2004, doi: 10.1023/b:lida.0000036389.14073.dd.

[18] H. Wang, H. Liao, X. Ma, and R. Bao, "Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model," *Rel. Eng. Syst. Saf.*, vol. 210, Jun. 2021, Art. no. 107504, doi: 10.1016/j.ress.2021.107504.

[19] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018, doi: 10.1016/j.ress.2017.11.021.

[20] H. Wang, J. Yang, R. Wang, and L. Shi, "Remaining useful life prediction of bearings based on convolution attention mechanism and temporal convolution network," *IEEE Access*, vol. 11, pp. 24407–24419, 2023, doi: 10.1109/ACCESS.2023.3255891.

[21] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2017, pp. 88–95.

[22] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, and R. Yan, "Bi-LSTM-based two-stream network for machine remaining useful life prediction," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[23] M. S. Azari, F. Flammini, S. Santini, and M. Caporuscio, "A systematic literature review on transfer learning for predictive maintenance in Industry 4.0," *IEEE Access*, vol. 11, pp. 12887–12910, 2023, doi: 10.1109/ACCESS.2023.3239784.

[24] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, and M. Xu, "Cross-domain fault diagnosis using knowledge transfer strategy: A review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019, doi: 10.1109/ACCESS.2019.2939876.

[25] Z. Wu, H. Tang, Y. Li, Y. Ren, L. Chen, and A. Kumar, "Simulation and experimental analysis of rotor-bearing system with rolling element bearing fault in axial piston pump under churning condition," *Proc. Inst. Mech. Eng., K, J. Multi-Body Dyn.*, vol. 237, no. 1, pp. 98–113, 2023, doi: 10.1177/14644193221142333.

[26] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schoelkopf, and G. R. G. Lanckriet, "Hilbert space embeddings and metrics on probability measures," *J. Mach. Learn. Res.*, vol. 11, pp. 1517–1561, Apr. 2010.

[27] J. Xu, M. Fang, W. Zhao, Y. Fan, and X. Ding, "Deep transfer learning remaining useful life prediction of different bearings," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8.

[28] M. S. Rathore and S. P. Harsha, "Rolling bearing prognostic analysis for domain adaptation under different operating conditions," *Eng. Failure Anal.*, vol. 139, Sep. 2022, Art. no. 106414, doi: 10.1016/j.engfailanal.2022.106414.

[29] Y. Ganin, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.

[30] P. R. D. O. da Costa, A. Akçay, Y. Zhang, and U. Kaymak, "Remaining useful lifetime prediction via deep domain adaptation," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106682, doi: 10.1016/j.ress.2019.106682.

[31] F. Cheli and G. Diana, *Advanced Dynamics of Mechanical Systems*. Cham, Switzerland: Springer, 2015.

[32] N. Sawalhi and R. B. Randall, "Simulating gear and bearing interactions in the presence of faults," *Mech. Syst. Signal Process.*, vol. 22, no. 8, pp. 1924–1951, Nov. 2008, doi: 10.1016/j.ymssp.2007.12.001.

[33] N. S. Feng, E. J. Hahn, and R. B. Randall, "Using transient analysis software to simulate vibration signals due to rolling element bearing defects," in *Proc. 3rd Austral. Congr. Appl. Mech.*, Sydney, NSW, Australia, 2002, pp. 689–694.

[34] D. Petersen, C. Howard, and Z. Prime, "Varying stiffness and load distributions in defective ball bearings: Analytical formulation and application to defect size estimation," *J. Sound Vibrat.*, vol. 337, pp. 284–300, Feb. 2015.

[35] T. A. Harris, *Essential Concepts of Bearing Technology*. Boca Raton, FL, USA: CRC Press, 2006.

[36] F. Bonnardot, M. El Badaoui, R. B. Randall, J. Danière, and F. Guillet, "Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)," *Mech. Syst. Signal Process.*, vol. 19, no. 4, pp. 766–785, Jul. 2005, doi: 10.1016/j.ymssp.2004.05.001.

[37] H. Zhang, P. Borghesani, S. Zhuang, and Z. Peng, "Bearing run-to-failure datasets of UNSW," Mendeley Data, V3, 2021. [Online]. Available: https://data.mendeley.com/datasets/h4df4mgrfb/3, doi: 10.17632/h4df4mgrfb.3.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[39] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[40] J. Luo and X. Zhang, "Convolutional neural network based on attention mechanism and bi-LSTM for bearing remaining life prediction," *Appl. Intell.*, vol. 52, no. 1, pp. 1076–1091, Jan. 2022, doi: 10.1007/s10489-021-02503-2.

[41] R. Guo and Y. Ji, "Remaining useful life prediction for bearing of an air turbine starter using a novel end-to-end network," *Meas. Sci. Technol.*, vol. 34, no. 6, Jun. 2023, Art. no. 065109, doi: 10.1088/1361-6501/acbed0.

[42] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," 2018, *arXiv:1809.03006*.

[43] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–9.

SIMENG XU received the B.S. degree in energy and power engineering from Northeastern University, Shenyang, China, in 2021. She is currently pursuing the M.S. degree with the College of Energy, Xiamen University, Xiamen, China. Her research interests include rolling element bearing fault diagnosis and degradation prediction based on dynamic modeling, signal processing, and deep-learning.

CHENXING JIANG received the Ph.D. degree in engineering thermophysics from Harbin Institute of Technology, Harbin, China, in 2015. He is currently an Associate Professor and the Ph.D. Supervisor of the College of Energy, Xiamen University, Xiamen, China. He has presided over the National Natural Science Foundation of China. His research interests include power equipment dynamics, fault diagnosis, and intelligent manufacturing.

CANGJIE YANG received the B.S. degree in energy and power engineering from East China University of Science and Technology, Shanghai, China, in 2022. He is currently pursuing the M.S. degree with the College of Energy, Xiamen University. His current research interests include fault diagnosis, transfer learning, and their applications on bearings of the permanent magnet synchronous motor.

HAIFENG CAO was born in Hefei, China. He received the B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree in advanced energy with Xiamen University, Xiamen, China. His research interests include machinery condition monitoring and intelligent fault diagnostics of rotating machinery.

ZHENGKAI SONG received the B.S. degree in energy and power engineering from Shandong University, in 2021. He is currently pursuing the M.S. degree with the School of Energy, Xiamen University. His research interests include the application of data-driven algorithms in rotating machinery and other scientific and engineering fields.

RULIN ZHENG received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 2022. She is currently pursuing the M.S. degree with the College of Energy, Xiamen University, Xiamen, China. Her research interest includes dynamic modeling of the gear transmission rotor systems.

SHIYI LU was born in Fujian, China. He received the B.S. degree in new energy science and engineering from Xiamen University, Xiamen, China, in 2021, where he is currently pursuing the master's degree in energy and power engineering. His research interests include fault diagnostics of rotating bearings and vibration control of rotating machines.

HANZHANG XU received the B.S. and Ph.D. degrees from Harbin Engineering University, in 2017 and 2023, respectively. He is currently with the Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), as a Postdoctoral Fellow. He has been doing research on shafting dynamics and bearing tribology, since 2017.

HENGCHENG ZHANG received the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2021. He was with the University of New South Wales, in 2022. He is currently a Researcher with the Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), Xiamen, China. His research interests include digital twins, vibration analysis, structural health monitoring, dynamics, tribology, signal processing, and machine learning.