lubricants

MDPI

Article

AsdinNorm: A Single-Source Domain Generalization Method for the Remaining Useful Life Prediction of Bearings

Juan Xu [1,*], Bin Ma [1], Weiwei Chen [2] and Chengwei Shan [3]

[1] School of Computer and Information, Hefei University of Technology, Hefei 230601, China; 2021171247@mail.hfut.edu.cn
[2] Shanghai Aerospace Control Technology Institute, Shanghai 201109, China; youthjiang@126.com
[3] CCTEG Changzhou Research Institute Tiandi (Changzhou) Automation Co., Ltd., Changzhou 213000, China; shanchengwei@cari.com.cn
* Correspondence: xujuan@hfut.edu.cn

Abstract: The remaining useful life (RUL) of bearings is vital for the manipulation and maintenance of industrial machines. The existing domain adaptive methods have achieved major achievements in predicting RUL to tackle the problem of data distribution discrepancy between training and testing sets. However, they are powerless when the target bearing data are not available or unknown for model training. To address this issue, we propose a single-source domain generalization method for RUL prediction of unknown bearings, termed as the adaptive stage division and parallel reversible instance normalization model. First, we develop the instance normalization of the vibration data from bearings to increase data distribution diversity. Then, we propose an adaptive threshold-based degradation point identification method to divide the healthy and degradation stages of the run-to-failure vibration data. Next, the data from degradation stages are selected as training sets to facilitate the RUL prediction of the model. Finally, we combine instance normalization and instance denormalization of the bearing data into a unified GRU-based RUL prediction network for the purpose of leveraging the distribution bias in instance normalization and improving the generalization performance of the model. We use two public datasets to verify the proposed method. The experimental results demonstrate that, in the IEEE PHM Challenge 2012 dataset experiments, the prediction accuracy of our model with the average RMSE value is 1.44, which is 11% superior to that of the suboptimal comparison model (Transformer model). It proves that our model trained on one-bearing data achieves state-of-the-art performance in terms of prediction accuracy on multiple bearings.

Keywords: adaptive threshold; RUL prediction; stage division; reversible instance normalization

check for updates

Citation: Xu, J.; Ma, B.; Chen, W.; Shan, C. AsdinNorm: A Single-Source Domain Generalization Method for the Remaining Useful Life Prediction of Bearings. *Lubricants* **2024**, *12*, 175. https://doi.org/10.3390/lubricants12050175

Received: 1 April 2024
Revised: 29 April 2024
Accepted: 12 May 2024
Published: 14 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bearings are essential in rotating machinery, serving as a vital part in maintaining the smooth functioning of industrial equipment. The failure of bearings during the equipment's operation not only poses a direct threat to the safety of the machinery but also leads to significant economic losses. Hence, the accurate prediction of the remaining useful life (RUL) of bearings is essential for the effective prognosis and health management of the equipment [1,2].

The current deep learning models have shown impressive performance in predicting the RUL of bearings [3]. However, these models rely on the assumption that the training and testing sets are independent identically distributed (I.I.D.) [4]. In realistic industrial scenarios, different bearings usually work under different conditions, and it is uncertain whether the vibration data collected from the unknown bearing are I.I.D. with the training bearing. Furthermore, obtaining the full life cycle, i.e., run-to-failure vibration data for every type of bearing, for the model training is unrealistic. Therefore, exploring the effective learning model trained on one bearing that can be generalized to predict the RUL of unknown bearings is of great application.

Domain adaptive approaches provide effective solutions to address the issue of data distribution discrepancies between training and testing sets [5–7]. These methods focus on closing the gap between the source and target domains, allowing the model to carry over knowledge from the source domain (such as a source bearing) to the target domain (like a target bearing) to predict the RUL, as shown in Figure 1a. However, domain adaptive methods cannot be applicable if the target bearing data are not accessible for model training.

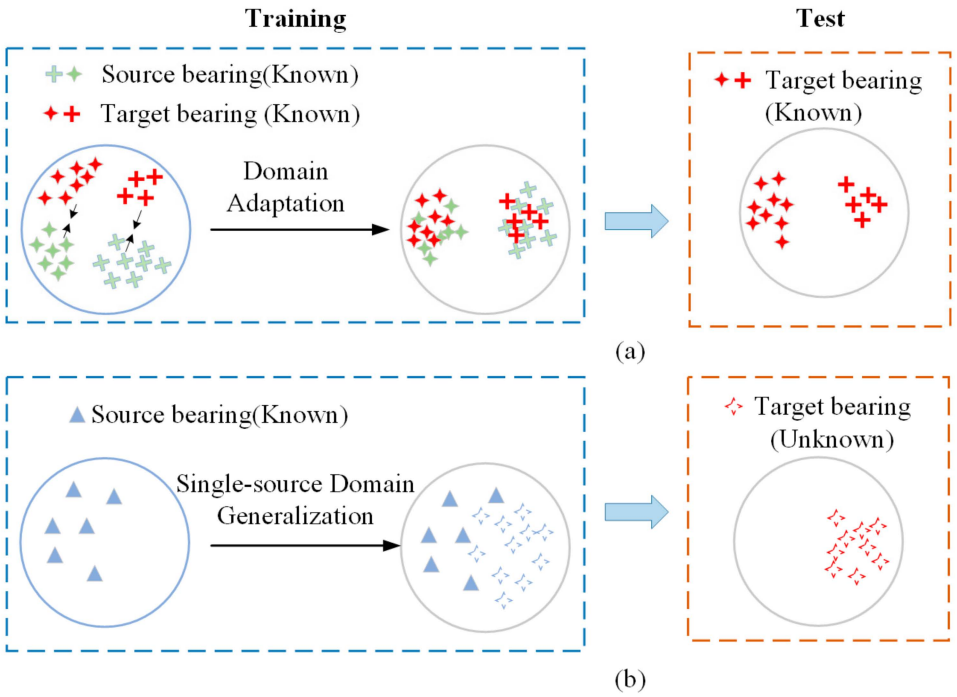


Figure 1. Schema of the two approaches. (**a**) The domain adaptive strategy necessitates that the target bearing's data be accessible for inclusion in model training; (**b**) the single-source domain generalization approach only uses source bearing data for model training without the target bearing data.

To address the aforementioned issues, the single-source domain generalization approaches provide a promising solution for predicting the RUL of unknown bearings. The approaches focus on data enhancement to expand the data distribution of the single-source domain and improve the generalization performance of the model to the unknown target domain, as shown in Figure 1b. Recently, single-source domain generalization approaches have blossomed in the computer vision (CV) field [8,9]. However, they have not yet been applied to the prediction RUL of bearings. This can be attributed to two main reasons as follows.

Firstly, the existing single-source domain generalization approaches do not consider the stage discrepancy of time series data. Different from the image data in the CV field, the vibration data of the full life cycle of bearings have significant time domain stage discrepancies. In the early health stage of the bearing, the fluctuation of its vibration time domain data is very tiny. Such vibration data of the health stage have a subtle, sometimes even negative, influence on describing the degradation trend of the bearing. Therefore, it is very important to divide the vibration data of the full life cycle to eliminate the effects of the early health stage data. In addition, the current single-source domain generalization methods strive to enhance the data distribution diversity using image-oriented augmentation techniques. However, such techniques may produce "counterfactual data" that completely deviate from the authentic vibration data distribution of bearings, which will deteriorate the generalization ability of the model.

To address the issues mentioned earlier, we suggest a more universal RUL prediction model, trained on a single bearing, that can be applied directly to other bearings, termed as the adaptive stage division and parallel reversible instance normalization (AsdinNorm)

model. The problem of predicting the RUL for unknown bearings has been approached as a single-source domain generalization learning challenge through two key steps: firstly, determining the degradation point locations of different bearings using an adaptive threshold stage division method to adaptively and iteratively lock the degradation trend and finding the final degradation point locations after many iterations to divide the health and degradation stages of the run-to-failure vibration data of bearings, and combining instance normalization and instance denormalization of the bearing data into a unified GRU-based RUL prediction network for the purpose of leveraging the distribution bias in instance normalization, as well as to obtain a better overall prediction accuracy and improve the generalization performance of the model.

The main contributions of this paper can be summarized as follows:

1. The proposed AsdinNorm model comprises three modules: instance normalization, adaptive threshold stage division, and parallel reversible normalization RUL prediction, respectively, used for enhancing the diversity of data distribution, degradation stage division, and leveraging the distribution bias in instance normalization of the vibration data of the source bearing, which improve the prediction accuracy and generalization ability of the model.
2. We designed an adaptive threshold-based degradation point identification method to effectively divide the health and degradation stages of the full life cycle vibration data of bearings. The designed adaptive threshold algorithm iteratively updates the degradation point locations to quickly and efficiently obtain the final degradation point locations of different bearings. Correspondingly, the vibration data of the degradation stage are selected for the model training for the purpose of reducing the interference of early fluctuations of the health stage, as well as eliminating the influence of the data distribution discrepancy between the training bearings and unknown testing bearings on the model performance.
3. We explored the parallel instance normalization and denormalization algorithm of the source bearing data and then combined it into a unified GRU-based RUL prediction network, which avoids the generation of "counterfactual" data, as well as the distribution bias in data enhancement, and achieves a better prediction accuracy while improving the generalization performance of the model.

2. Related Works

The single-source domain generalization approaches [10,11] use data augmentation techniques to expand the data distribution of the training set (single-source domain) to cover the data distribution of the unknown target domain as much as possible and, further, to improve the generalization performance of the model. They are usually categorized as GAN-based, meta-learning-based, and scaling-based approaches.

GAN-based [10,12,13] methods can create additional data resembling the source domain by using generative and discriminative models. HSIGAN [14] allows the discriminator to perform classification in addition to distinguishing between real and synthetic data, i.e., it learns to generate overall real samples and also encourages the generator to learn the representation of different classes of samples. DAGAN [15] learns a large number of data-enhanced transformations by training the autoencoder. BAGAN [16] trains the autoencoder to learn the multivariate normal-terrestrial distribution to the image, which represents the distribution of the overall dataset.

The meta-learning [17] approaches use the training data as a meta-training set, while the generated data serve as a test set to learn robust feature representations using a meta-learning strategy. Qiao et al. [18] increased the source sample size in the input and label space and evaluated the guidance based on uncertainty. This method is used for data enhancement, domain generalization, and the effective training of models in a Bayesian meta-learning framework. The findings indicate that the proposed approach is effective and outperforms others in various tasks.

The scaling-based [19] approaches are the general technology for single-source domain generalization. ASR-Norm [20] uses neural networks to adaptively normalize and scale statistics to match various domains. SORAG [21] uses the manual synthesis of new samples to improve the robustness of the model to tackle the problem of sample imbalance. SamplePairing [22] performs basic data enhancement (e.g., random flipping) and then superimposes the data by pixels in the form of averaging to synthesize new samples, which can expand the diversity of the samples and enhance the generalization ability of model.

In short, the single-source domain generalization methods are currently used mainly for classification tasks of images. However, when dealing with the prediction problem for bearing vibration data, the interference of vibration data at different stages and the distribution bias in data enhancement must be considered. Therefore, it is crucial to develop a novel single-source domain generalization method that is more suitable to the characteristics of vibration data for RUL predictions of unknown bearings.

3. Proposed Method

To briefly describe the RUL prediction problem for bearings, given two sets of bearing vibration signals in different working conditions: the source bearing dataset $H_s = \left\{h_s^1, h_s^2, \cdots h_s^R\right\}$ and the unknown target bearing dataset $H_t = \left\{h_t^1, h_t^2, \cdots h_t^M\right\}$, where R, M represents the total count of the samples.

Our model captures the degradation feature from the vibration data of the source bearing $h^i, h^{i+1}, \cdots h^{i+t}$; then, the model predicts the vibration data h^{i+t+1} [23]:

$$P\left(h_s^i, h_s^{i+1}, \cdots h_s^{i+t}, \theta\right) \rightarrow h_s^{i+t+1}, \tag{1}$$

where θ is the parameter of the model, and h_s^i denotes the i_{th} vibration data of the source bearing.

The model parameters are optimized via the iterative training, expressed as shown:

$$\theta = arg \max_{\hat{\theta}} p\left(\hat{\theta} / H_s\right). \tag{2}$$

Finally, inputting the target bearing H_t to the trained model, the model predicts the $i + t + 1$ vibration data of the testing bearing $h_t^{i+t+1} = \mathrm{P}\left(h_t^i, h_t^{i+1}, \cdots h_t^{i+t}\right)$.

The AsdinNorm model's architecture is depicted in Figure 2 and consists of three main components: an instance normalization module, an adaptive threshold stage division module, and a parallel reversible normalization RUL prediction module.

3.1. Instance Normalization

The instance normalization module is designed to preserve the non-stationary information from vibration data while reducing the difference in data distribution from target bearings. Firstly, we obtain the peak-to-peak values $\mathrm{S} = \left\{s^1, s^2, \cdots, s^R\right\}$ from the vibration signal data of the source bearing H_s, which can alleviate the interference of noise and facilitate a clear representation of the degradation trend, expressed as follows:

$$s^i = max\left(h_s^i\right) - min\left(h_s^i\right). \tag{3}$$

Next, we convert the peak-to-peak values S into a time series $\mathrm{X} = \left\{x^1, x^2, \cdots x^n\right\}$ using a sliding window. At last, we normalize the input time series x^i by applying the instance mean and standard deviation. The variable L_x represents the length of the input sequence, and the mean μ^i and standard deviation ρ^i of each instance of the input sequence are calculated as follows:

$$\mu^i = \frac{1}{L_x} \sum_{j=1}^{L_x} x_j^i, \tag{4}$$

where x_j^i denotes the jth sample of the ith sliding window.

$$\rho^i = \frac{1}{L_x}\sum_{j=1}^{L_x}\left(x_j^i - \mu^i\right)^2. \tag{5}$$

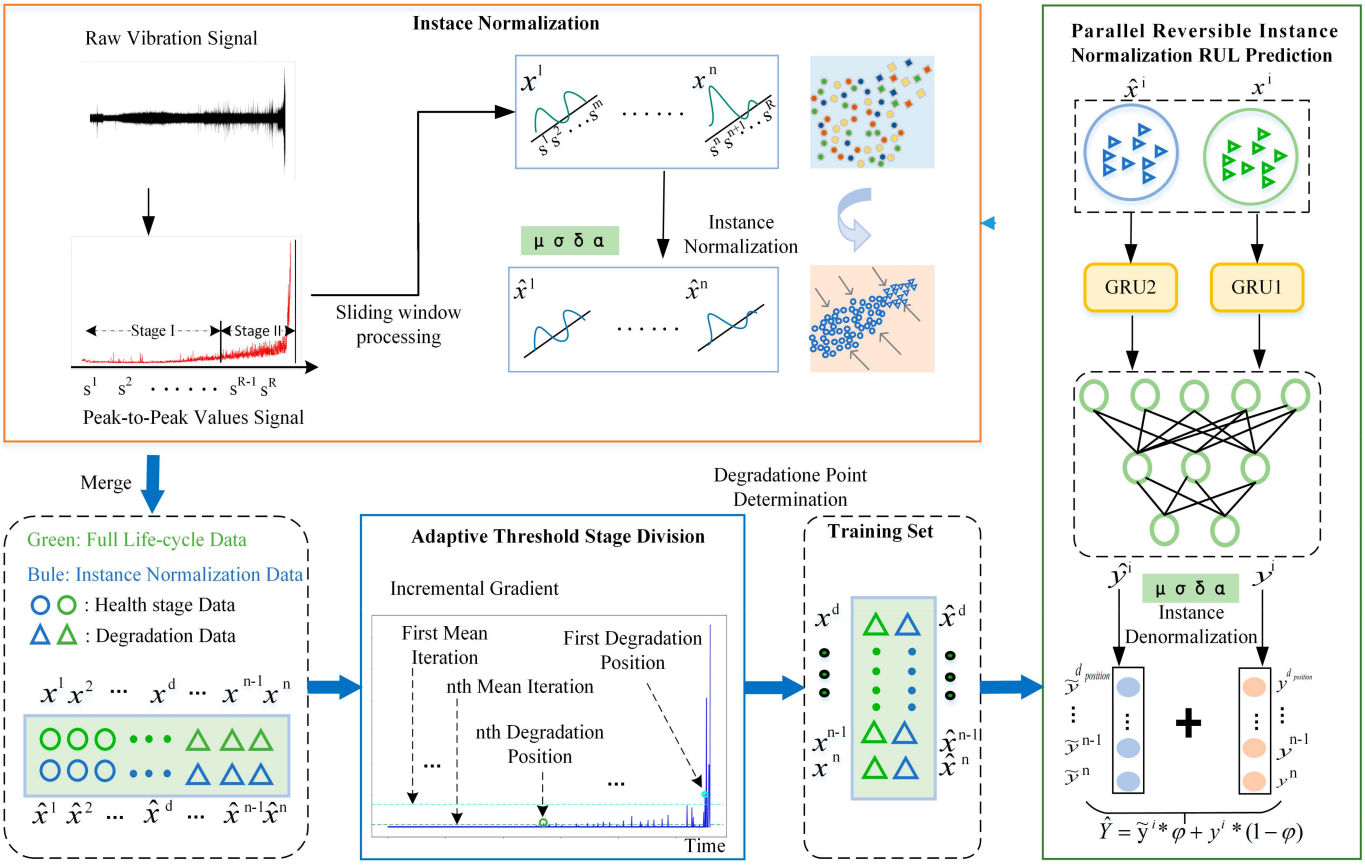


Figure 2. The structure of the proposed AsdinNorm model.

With these statistics, we derive the normalized [24] input sequence data $\hat{x}^i$ from the input sequence data x^i.

$$\hat{x}^i = \alpha\left(\frac{x^i - \mu^i}{\sqrt{\rho^i + \varepsilon}}\right) + \delta, \tag{6}$$

where α, δ, $\in$, and R^1 are the learnable affine parameter vectors used in the instance normalization method to equalize the effective information across the bearings.

Importantly, we merge the normalized data with the time series data x^i to form a new time series input $\hat{X} = \{x^i, \hat{x}^i\}_{i=1}^n$, which serves as the input to the adaptive threshold stage division module.

3.2. Adaptive Threshold Stage Division

In this section, an adaptive threshold stage division method is proposed to determine the location of the bearing degradation points for the purpose of the segregation of the health stage and the fast degradation stage. This critical step facilitates the accurate prediction RUL of unknown target bearings via using the bearing data of the degradation stage. The specific procedure involves the following steps, as shown in Figure 3.

Calculate to obtain the degradation path: Firstly, we adopt the isotonic regression algorithm to transform the irregular bearing data into segmented incremental step data. Suppose that function F is the mapping function of the isotonic regression algorithm, the input of the algorithm is the peak-to-peak value of the vibration data s, and the

output is $D^*_{1\to R} = F(S) = \left\{d^*_1, d^*_{2,} \cdots d^*_R\right\}$ with a monotonically increasing trend. This transformation ensures that the degradation trend of the bearing data is monotonically increasing and eliminates the noise interference from the original data. Figure 3a illustrates the original degradation trend of the bearing data via the isotonic regression algorithm. Several jump points can be found in the figure, which make it difficult to determine the proper degradation point positions.

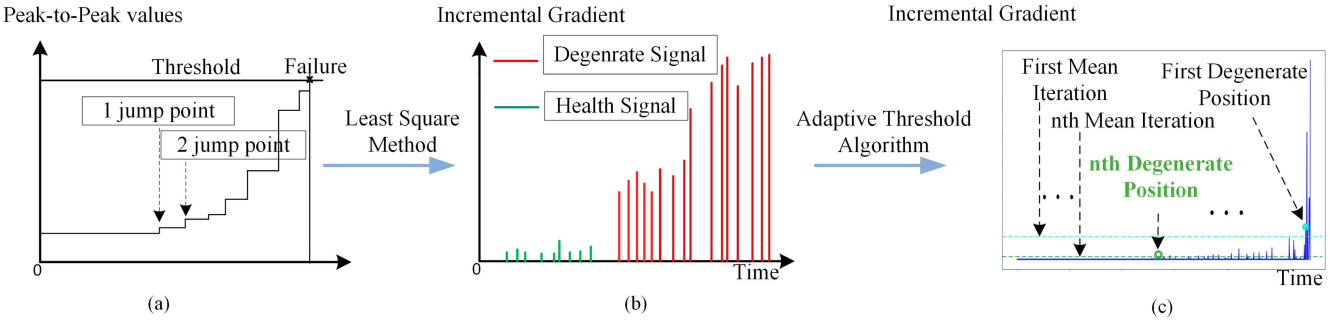


Figure 3. Adaptive threshold stage division algorithm: (**a**) degradation path; (**b**) gradient path; (**c**) adaptive threshold iteration process.

The algorithm generates the gradient path: Correspondingly, we use the least square method [25] with a sliding window to calculate the gradient Δ_i with the window size m, as shown in Figure 3b. The specific formula is as follows:

$$\Delta_i = \frac{\sum_{j=i}^{i-m+1} q_j d^*_j - \frac{1}{m}\sum_{j=i}^{i-m+1} q_j \sum_{j=i}^{i-m+1} d^*_j}{\sum_{j=i}^{i-m+1} q_j^2 - \frac{1}{m}\left(\sum_{j=i}^{i-m+1} q_j\right)^2}, \tag{7}$$

where q_j is the subscript of the corresponding peak-to-peak value S, and $q_j = 1, 2 \cdots R - m + 1$.

Adaptive threshold iteration process: In order to determine the proper degradation point from a number of jump points and determine the position of the stage division, we further propose an adaptive threshold algorithm to compare the incremental gradients of degradation over multiple iterations. Specifically, we use the initial position point d^*_1 as the first point d_{start} for the initialization of this iterative algorithm and the final point d^*_R position as the tail d_{end}. We calculate the average value of the gradient, denoted as $Av = mean\{\Delta_{start}, \Delta_{start+1}, \cdots, \Delta_{end}\}$.

This algorithm requires two key conditions: the first is the gradient increment of degradation with a continuous fluctuation; that is, for any Δ_i greater than zero, $\{\Delta_i, \Delta_{i+1}, \cdots, \Delta_{i+C-1}\} \in \{\Delta_{start}, \Delta_{start+1}, \cdots, \Delta_{end}\}$, there are C in total. The second is that there is a gradient increment of Q, which is greater than Av. C and Q are set by human experience.

Update the position of the degradation points $d_{position}$ by iterating repeatedly until the optimal degradation point is chosen at the end of the iteration. The algorithm flow is illustrated in Figure 4, and the algorithm process is described as follows:

1. Proceed to the second step only if the count of incremental gradients with consecutive jumps is at least C (where $(C > 0)$); otherwise, halt the process.
2. If the number of incremental gradient points meeting condition (1) is at least Q (where $(0 < Q \leq C)$), proceed to step 3; otherwise, halt the process.
3. If condition (2) is met, the starting point of the selected gradient is set as the new degradation point, and the final point of the gradient with continuous jumps becomes the new gradient's end. The updated gradient average Av is then calculated for round 4.
4. The algorithm converges to the final degradation point and stops; if not, return to step 1.

As illustrated in Figure 3c, the adaptive threshold stage division algorithm determines the proper degradation point position to eliminate the interference caused by multiple

jump points. Subsequently, the final degradation point is used to divide the bearing data into two stages: the health and fast degradation stages. We use the bearing data from the degradation stage as the input $\hat{X} = \{x^i, \hat{x}^i\}_{i=d_{position}}^{n}$ of the prediction module in the later section.

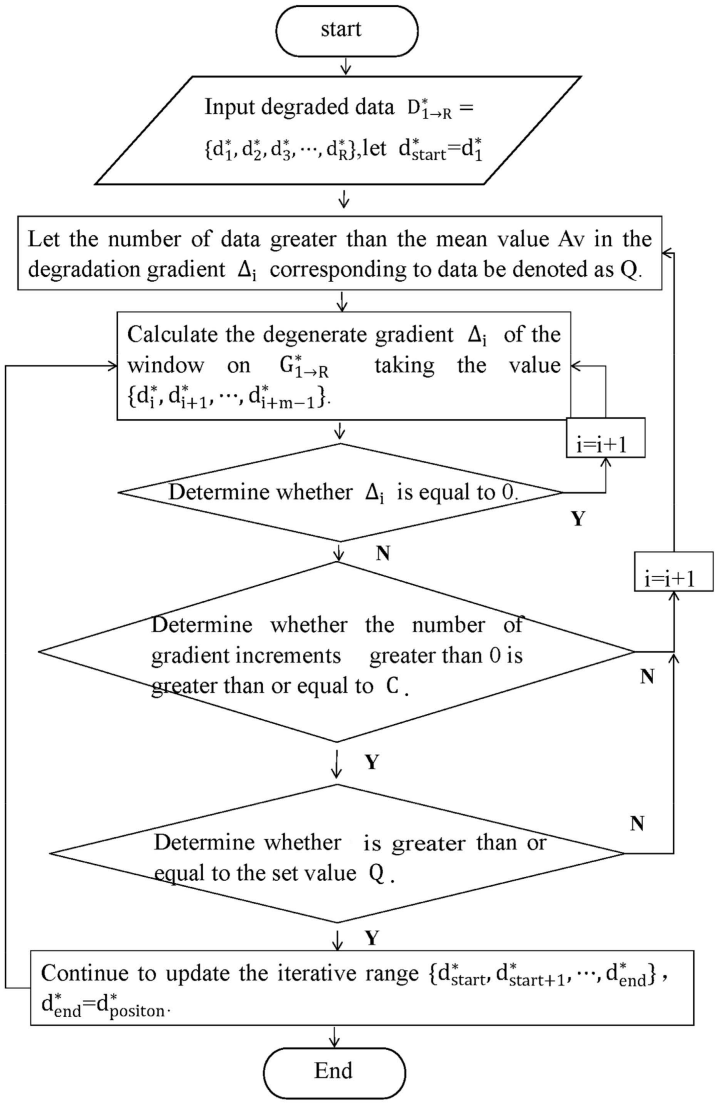


Figure 4. Flow chart of the adaptive threshold algorithm.

3.3. Parallel Reversible Instance Normalization RUL Prediction

As previously mentioned, the training set input to this module consists of two branches: the normalized data $\hat{X}$ and the time series data x^i. Therefore, the parallel reversible instance normalization RUL prediction module is accordingly designed with two branches to process the input data of these two parts. Firstly, $\hat{X} = \{x^i, \hat{x}^i\}_{i=d_{position}}^{n}$ is the input of the prediction module, and the output is $Y = P(\hat{X}) = \{y^i, \hat{y}^i\}_{i=d_{position}}^{n}$. Considering that the direct RUL prediction results $\hat{y}^i$ from the normalized data $\hat{x}^i$ may result in a distribution bias from the actual data, we further calculate the reverse normalized [5] predicted value $\tilde{y}^i$ from the the predicted value $\hat{y}^i$, expressed as follows:

$$\tilde{y}^i = \sqrt{\rho^i + \varepsilon} * \left(\frac{\hat{y}^i - \delta}{\alpha}\right) + \mu^i. \tag{8}$$

Importantly, the weights of the two prediction values are simultaneously optimized via model training to achieve a better overall prediction accuracy. The representation is as follows:

$$\hat{Y} = \tilde{y}^{i} * \varphi + \ y^{i} * (1 - \varphi). \tag{9}$$

Among these, $\varphi \in [0, 1]$ are the elements of the learnable affine parameter vector.

In terms of the design of the RUL predictor, it is necessary to use a network model that extracts deep information features well and has a lighter neural network structure. Therefore, we construct the GRU-based predictor P to predict the RUL $\hat{Y}$. The RUL predictor is composed of two single-layer gated cyclic unit GRUs and three fully connected layers. The GRU-based predictor's parameters are listed in Table 1.

Table 1. Model structural parameters.

No.	Layer	Operator	Dimensions
1	Input	Input samples	(12,1)
2	GRU1	Prediction	(None,12,30)
3	GRU2	Prediction	(None,12,30)
4	FC1	Fully connected	(None,60)
5	FC2	Fully connected	(None,120)
6	FC3	Fully connected	(None,1)

The pseudocode of the proposed method is shown in Algorithm 1.

Algorithm 1: AsdinNorm for RUL prediction.

1: **Input:** (Training stage) Source domain: $H_s = \left\{ h_s^i \right\}_{i=1}^{R}$, where h_s^i shows the i_{th} sample, and R shows the number of samples.
2: Data preprocessing: peak-to-peak values extract.
3: for I_1 epochs, do
4: Randomly initialize the weight of the AsdinNorm model θ.
5: Instance normalization from Equations (3) to (5).
6: Use the adaptive threshold stage division module to select the degradation stage data.
7: Use Equation (10) to calculate the margin loss.
8: Use Equations (8) and (9) to obtain the RUL prediction values and update the affine parameters δ, α, and φ.
9: end for
10: **Output:** The AsdinNorm model with optimal θ.
11: **Input:** (Test stage) Unseen target domain $H_t = \left\{ h_t^i \right\}_{i=1}^{M}$, where h_t^i shows the i_{th} sample, and M shows the number of samples.
12: peak-to-peak values extract.
13: Use the adaptive threshold stage division module to select the degradation stage data.
14: Use Equations (8)–(10) to obtain the RUL prediction values and calculate the evaluation indicators.
15: **Output:** RMSE of the target bearings.

4. Experiment and Discussion

4.1. Experiment Description

We conducted experiments using two public datasets: the IEEE PHM Challenge 2012 bearing dataset and the XJTU-SY bearing dataset. The IEEE PHM Challenge 2012 bearing dataset is provided by the bearing degradation experiments on the PRONOSTIA test stand. The PRONOSTIA experimental setup includes three primary components: rotational components, load components, and data measurement components, as illustrated in Figure 5a. The load is 4000 N. Vibration signals are captured every 10 s, with each recording lasting 0.1 s. Table 2 displays the dataset description under three operating conditions. We use the vibration signal of bearing 1_1 as the training set and test the other 12 bearings.

The XJTU-SY bearing datasets are provided by the Institute of Design Science and Basic Component at Xi'an Jiaotong University (XJTU), encompassing vibration signals from 15 rolling bearings operating under three distinct conditions. The vibration signals depict

the operational-to-failure transitions of the 15 rolling bearings across these three conditions. The dataset is sampled at a frequency of 25.6 kHz with a sampling period of 1 min, as shown in Figure 5b. Comprehensive details of the two datasets are presented in Table 2. In the PHM 2012 dataset, bearing 1_1 is utilized as the training set, with the remaining bearings serving as the test set. For the XJTU-SY dataset, bearing 3_1 is employed as the training set, while the remaining bearings constitute the test set to evaluate the model's generalization performance.

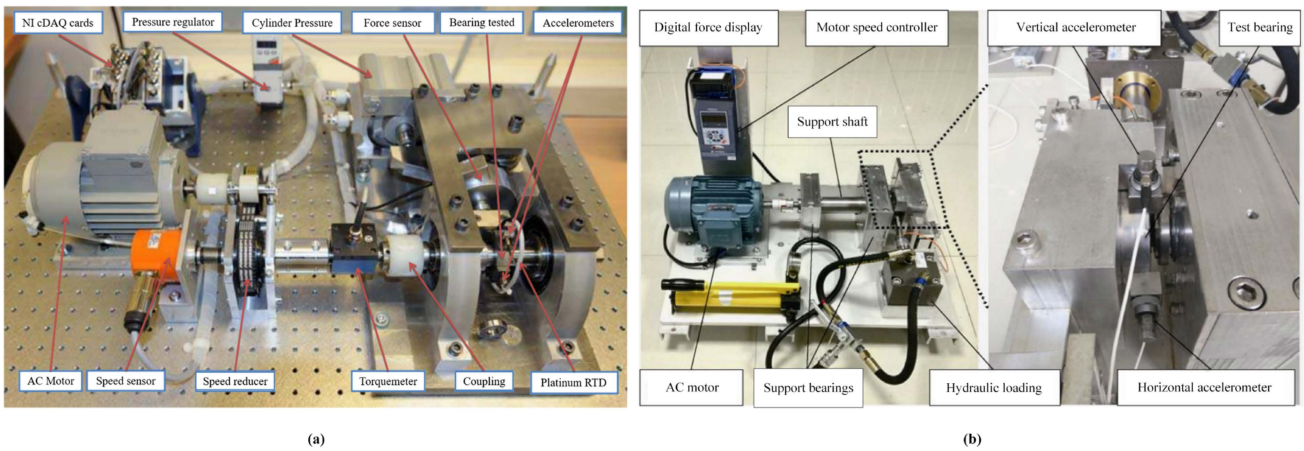


Figure 5. Test platforms: (**a**) PRONOSTIA experimental platform; (**b**) XJTU-SY test platform.

Table 2. Description of the datasets.

Data Set	Rotation Speed (rpm)	Load	Component	Dimension	Division
PHM2012	800	4000	Bearing1_1	(2803,2560)	training
			Bearing1_2	(871,2560)	testing
			Bearing1_3	(2375,2560)	testing
			Bearing1_5	(2463,2560)	testing
			Bearing1_6	(2448,2560)	testing
	1650	4200	Bearing2_1	(911,2560)	testing
			Bearing2_2	(797,2560)	testing
			Bearing2_3	(1955,2560)	testing
			Bearing2_4	(751,2560)	testing
			Bearing2_5	(2311,2560)	testing
			Bearing2_6	(701,2560)	testing
	1500	5000	Bearing3_1	(515,2560)	testing
			Bearing3_2	(1637,2560)	testing
XJTU-SY	2100	12,000	Bearing1_1	(123,2560)	testing
			Bearing1_2	(161,2560)	testing
			Bearing1_3	(158,2560)	testing
	2250	11,000	Bearing2_1	(491,2560)	testing
			Bearing2_2	(161,2560)	testing
			Bearing2_3	(533,2560)	testing
	2400	10,000	Bearing3_1	(2538,2560)	training
			Bearing3_2	(2496,2560)	testing
			Bearing3_3	(371,2560)	testing
			Bearing3_4	(1515,2560)	testing

We designed experiments in three aspects (Comparison Results Using Full Life cycle Data and Fast Degradation Data; Ablation Experiments; Comparison Experiments) to validate our method. Figure 6 illustrates the experimental flow chart of the verified procedure of the proposed method.

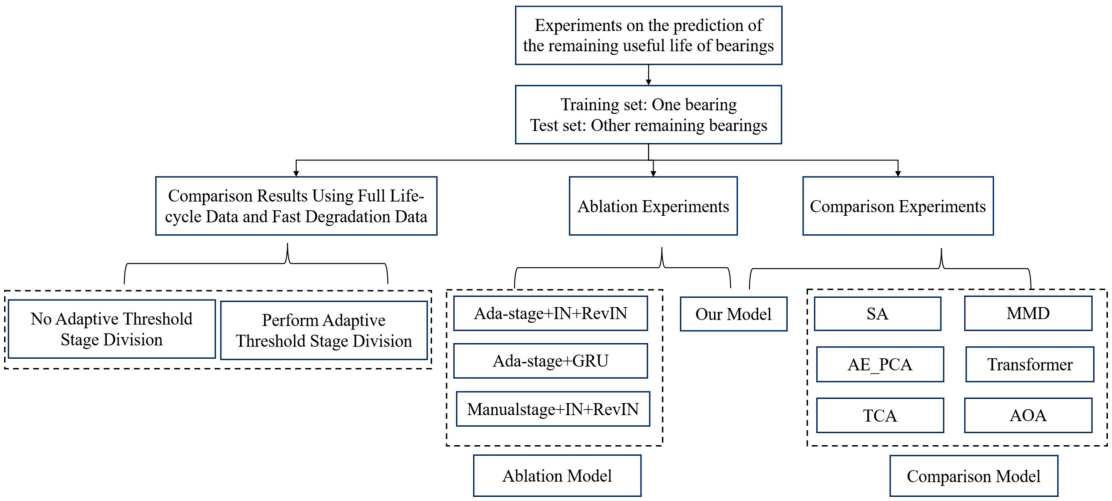


Figure 6. Experimental flow chart.

4.2. Adaptive Threshold Stage Division Experiment

Figure 7 presents the iterative process of the adaptive threshold stage division algorithm applied on bearing 1_3 and bearing 2_2 of the PHM 2012 dataset, along with the final degradation point position. It can be seen that the method can variously determine the positions of the degradation points for each iteration and continuously update the mean degradation gradient value. The iterative convergence of the algorithm identifies the final degradation points by filtering out several interference jump points, which accurately captures the subtle variations in the bearing data.

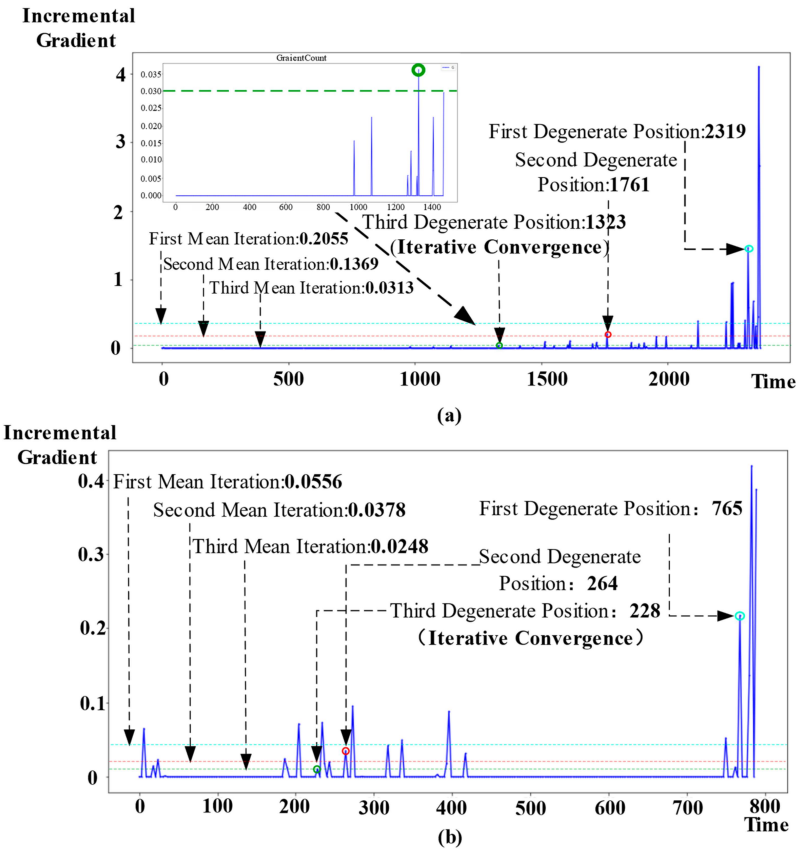


Figure 7. The iterative process of the adaptive threshold stage division algorithm: (**a**) bearing 1_3; (**b**) bearing2_2.

As depicted in Figure 7a, it can be observed that bearing 1_3 converges to the degradation point at 1323 after three iterations. Similarly, Figure 7b illustrates that bearing 2_2 converges to the degradation point at 228 after three iterations. The specific iteration counts and degradation point positions for each bearing can be referenced in Table 3.

Table 3. Adaptive threshold stage division results.

Data Set	Number of Iterations	Sample Size	Degenerate Point
Bearing 1_1	3	2803	1350
Bearing 1_2	3	871	720
Bearing 1_3	3	2375	1323
Bearing 1_5	2	2463	2240
Bearing 1_6	3	2448	1620
Bearing 2_1	6	911	130
Bearing 2_2	3	797	228
Bearing 2_3	1	1955	1840
Bearing 2_4	2	751	635
Bearing 2_5	2	2311	2165
Bearing 2_6	2	701	585
Bearing 3_1	3	515	390
Bearing 3_2	4	1637	1501

Using the proposed algorithm, all bearings are stage-divided, and Table 3 lists the number of iterations and final degradation point positions. The vibration data of the degradation stage are selected for the RUL predictions.

4.3. Comparison Results Using Full Life Cycle Data and Fast Degradation Data

To verify the effectiveness of stage division in the proposed method, we, respectively, use full life cycle data and fast degradation data of the bearings to predict the RUL. We choose the root mean square error ($RMSE$) [26] as the metric to evaluate the model performance, expressed as follows:

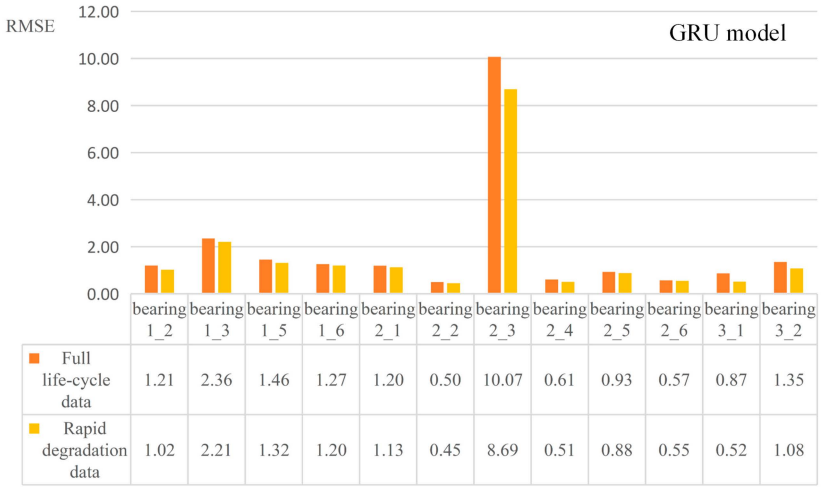
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}, \tag{10}$$

where Y_i represents the actual RUL value, $\hat{Y}_i$ is the estimated RUL value, and N indicates the total number of samples. The smaller the value of the $RMSE$, the superior the prediction performance of the model.

Figure 8 illustrates the prediction results of the GRU model and our proposed model for the full life cycle and fast degradation data of all test bearings, respectively. It is observed that the $RMSE$ of two models for the fast degradation stage data are smaller than those for the full life cycle data. This is because the data distribution discrepancy between the health stage and the fast degradation stage is significant. The distribution of data in the health stage can bias the model learning and affect the prediction performance in the fast degradation stage of the bearing data.

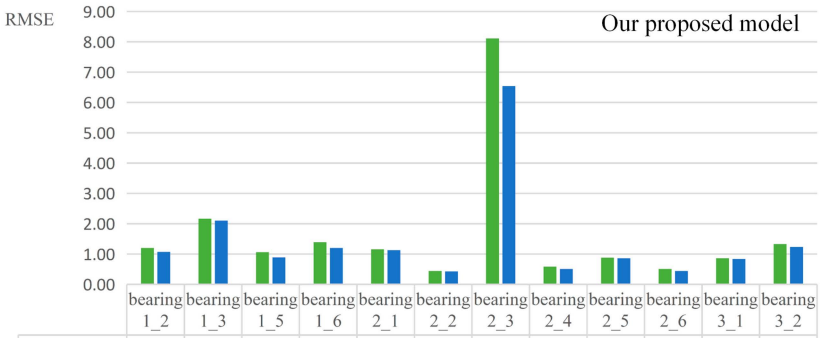
Further, we train our proposed model with the full life cycle data of bearing 1_1 of the PHM2012 dataset and the data of the fast degradation stage, respectively. Figure 9 shows the prediction results of bearing 1_2. The blue curve is the real values of bearing 1_2. The red curve is the prediction results of bearing 1_2 trained on the data of the fast degradation stage of bearing 1_1. The green curve is the one trained on the full life cycle data of bearing 1_1. It can be seen that the model trained with the data of the fast rapid degradation stage has a superior fitting to the vibration data with the rapid variation.

Two sets of experimental results demonstrate that dividing the bearing data into two stages and using fast degradation stage data for the RUL prediction results in a superior performance than using the full life cycle data for the prediction.



	bearing 1_2	bearing 1_3	bearing 1_5	bearing 1_6	bearing 2_1	bearing 2_2	bearing 2_3	bearing 2_4	bearing 2_5	bearing 2_6	bearing 3_1	bearing 3_2
Full life-cycle data	1.21	2.36	1.46	1.27	1.20	0.50	10.07	0.61	0.93	0.57	0.87	1.35
Rapid degradation data	1.02	2.21	1.32	1.20	1.13	0.45	8.69	0.51	0.88	0.55	0.52	1.08

(a)



	bearing 1_2	bearing 1_3	bearing 1_5	bearing 1_6	bearing 2_1	bearing 2_2	bearing 2_3	bearing 2_4	bearing 2_5	bearing 2_6	bearing 3_1	bearing 3_2
Full life-cycle data	1.20	2.17	1.06	1.39	1.16	0.44	8.11	0.59	0.88	0.51	0.86	1.33
Rapid degradation data	1.07	2.10	0.89	1.20	1.13	0.43	6.54	0.51	0.86	0.44	0.84	1.24

(b)

Figure 8. Prediction results for the full life cycle and fast degradation data of all test bearings: (**a**) GRU model; (**b**) our proposed model.

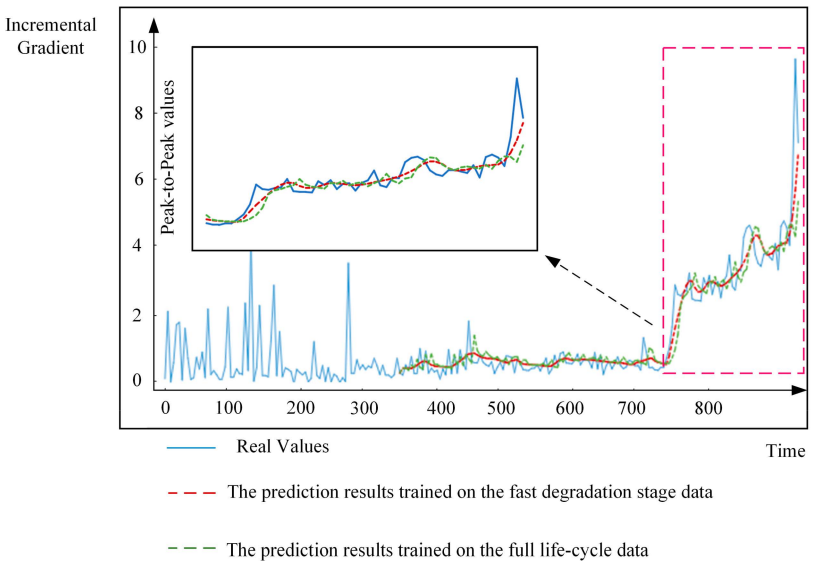
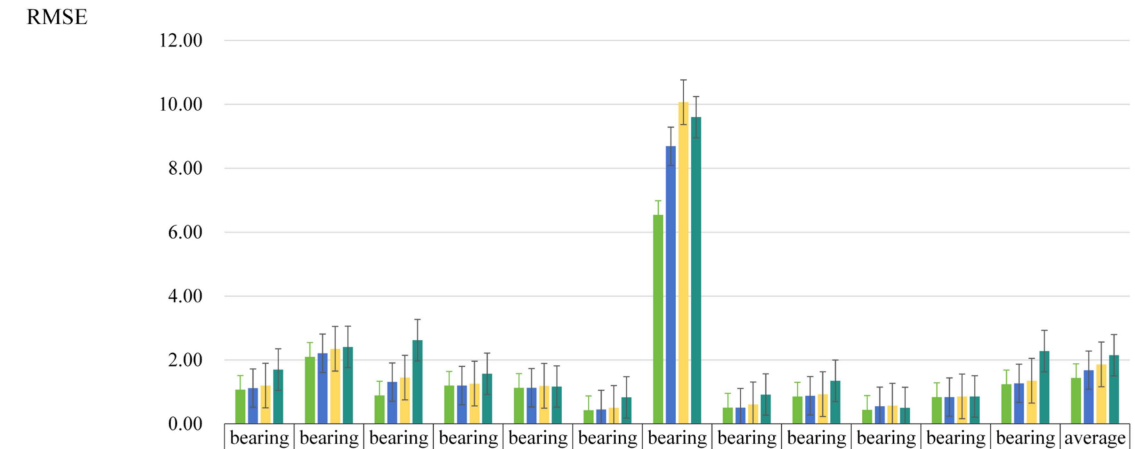


Figure 9. Prediction results of bearing 1_2 using different training data of bearing 1_1.

4.4. Ablation Study

In this section, we conduct three ablation experiments to verify the effectiveness of each module of our proposed method. Specifically, the model with the adaptive threshold stage division module and GRU module (termed as Adapstage+GRU); the model with the adaptive threshold stage division module, instance normalization module, and reversible normalization-based RUL prediction module (termed as Adapstage+IN+RevIN); and the model with the manual threshold stage division module, instance normalization module, and parallel reversible normalization-based RUL prediction module (termed as Manualstage+IN+RevIN) are constructed, respectively.

For the PHM 2012 dataset, we train three models on bearing 1_1. The predictive results for the 12 test bearings are depicted in Figure 10. Similarly, for the XJTU-SY dataset, the training bearing is 3_1, and the predictive results for nine test bearings are illustrated in Figure 11.



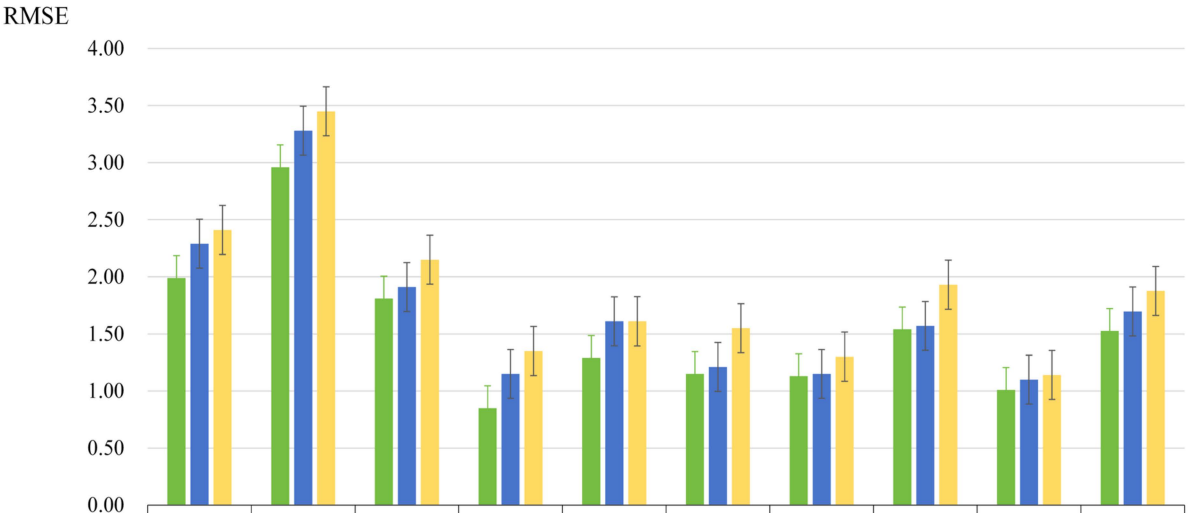
	bearing 1_2	bearing 1_3	bearing 1_5	bearing 1_6	bearing 2_1	bearing 2_2	bearing 2_3	bearing 2_4	bearing 2_5	bearing 2_6	bearing 3_1	bearing 3_2	average value
Ours	**1.07**	**2.10**	**0.89**	**1.20**	**1.13**	**0.43**	**6.54**	**0.51**	**0.86**	**0.44**	**0.84**	**1.24**	**1.44**
Adap-stage+IN+RevIN	1.12	2.21	1.31	**1.20**	**1.13**	0.45	8.69	**0.51**	0.88	0.55	**0.84**	1.27	1.68
Adap-stage+GRU	1.20	2.35	1.45	1.26	1.19	0.50	10.07	0.61	0.93	0.57	0.86	1.35	1.86
Manualstage+IN+RevIN	1.7	2.41	2.62	1.57	1.17	0.83	9.6	0.92	1.35	0.5	0.86	2.28	2.15

Figure 10. Ablation experimental results on the PHM 2012 dataset.

As shown in Figure 10, the above bolded data is the best result for all the ablation models. In the experiments on the PHM2012 dataset, the Manualstage+IN+RevIN model exhibits the worst prediction performance, with an average *RMSE* value of 2.15. The next-worst prediction performance is the Adap-stage+GRU model, with an average *RMSE* value of 1.86. Then comes the Adap-stage+IN+RevIN model, with an average *RMSE* value of 1.68, and finally, the best prediction is our proposed method, with an average *RMSE* value of 1.44, and it can be seen that both the average *RMSE* value and the *RMSE* values of individual bearing predictions are the smallest for our proposed method, which proves that the prediction performance of our proposed method is better than the other models. This is because the degradation trends of different bearings under different operating conditions differ obviously. Using fixed thresholds to perform stage division for all bearings will result in non-negligible bias in selecting degradation points for some bearings with significantly different data distributions. Meanwhile, the Adapstage+IN+RevIN model has a suboptimal prediction performance. Although the normalization and inverse normalization methods are used in that model, the lack of learnable parallel processes affects the generalization ability of the model.

As shown in Figure 11, the above bolded data is the best result for all the ablation models. In the experiments on the XJTU-SY dataset, it can be seen that the GRU+FC model exhibits the worst prediction performance, with an average *RMSE* value of 1.88. The next-worst prediction performance is the GRU+RevIN model, with an average *RMSE* value of 1.70. Finally, it was our proposed method, which predicts an average *RMSE* value

of 1.53. From the above results, the effectiveness and generalization ability of the parallel reversible normalized RUL prediction module proposed in our method can be seen.



	bearing1_1	bearing1_2	bearing1_3	bearing2_1	bearing2_2	bearing2_3	bearing3_2	bearing3_3	bearing3_4	average value
Ours	**1.99**	**2.96**	**1.81**	**0.85**	**1.29**	**1.15**	**1.13**	**1.54**	**1.01**	**1.53**
GRU+RevIN	2.29	3.28	1.91	1.15	1.61	1.21	1.15	1.57	1.10	1.70
GRU+FC	2.41	3.45	2.15	1.35	1.61	1.55	1.30	1.93	1.14	1.88

Figure 11. Ablation experimental results on the XJTU-SY dataset.

However, as seen in Figures 10 and 11, it can be seen that the prediction accuracy of the proposed model for some bearings (e.g., bearing 1_6 and bearing 3_1) is not much different or even the same compared to the prediction accuracy of other ablation models. This is because we use bearing 1_1 for training and other bearings as test bearings, the peak-to-peak height of bearing 1_1 reaches more than 50, the degradation trend is smoother in the early stage, and the performance is more dramatic in the later stage. The maximum peak-to-peak values in bearing 1_6 and bearing 3_1 are not more than 10, and the data distribution of their degradation trend is more different from that of bearing 1_1. Our model performs very well in the prediction of other bearings that have peak-to-peak heights and degradation trends closer to the training bearings, and the data distribution differences between them are smaller, so the model shows a better generalization ability and prediction accuracy.

Summarizing the experimental results shows that our proposed model can obtain more accurate degradation positions, as well as a better prediction accuracy, than any other model.

4.5. Comparison with State-of-the-Art Methods

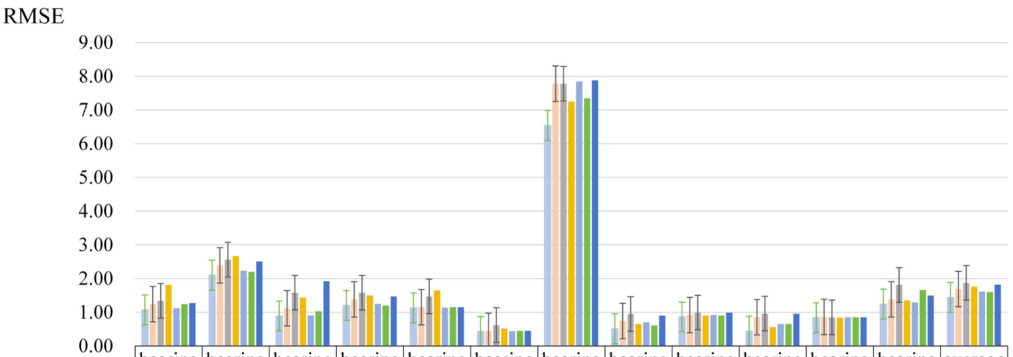
In this section, we compare our model with six state-of-the-art methods on the PHM 2012 dataset to verify its superiority. The comparison models include the AE (Autoencoder) [27], SA (Self-Attention) [28], MMD (Maximum Mean Discrepancy) [29], TCA [30], Transformer [31], and AOA models [32] listed in Table 4. Among them, SA, AE, and Transformer are the prevalent learning models; TCA and MMD belong to the domain adaptive models; and AOA belongs to the domain generalization model.

Figure 12 shows the prediction results of different comparison models for all tested bearings, the above bolded data is the best result for all the comparison models. As well as the average *RMSE* values. From the average *RMSE* values, it can be seen that the AE model predicts an average *RMSE* value of 1.87, which is the worst prediction accuracy among all the comparison models, followed by the AOA model, which predicts an average *RMSE* of 1.82, then the TCA model, which predicts an average *RMSE* of 1.76, and the SA model, which predicts an average *RMSE* of 1.69; it is worth noting that the MMD

model and Transformer model predict an average *RMSE* value of 1.61 and 1.6, respectively. Finally, our proposed model predicts an average *RMSE* value of 1.44.

Table 4. Description of the comparison models.

Method	Description
SA	The Self-Attention model obtains more information by globally associating weights and then performs weighted sum of inputs, i.e., using information from other regions. We use the Self-Attention model to transform the features of different time series as the input parameter matrix, gain the weights by the similarity measure, and then weighted sum them.
AE_PCA	The AE_PCA model extracts the features from multi-class bearing data and then maps the high-dimensional features to low-dimensional features by the principal component analysis method to retain the effective features. Hence, we input the input information into the AE model for feature extraction, and then, the downscaled features by the PCA method are used for the prediction.
TCA	When the source and target domains have different data distributions, the TCA model maps the two domain bearing data together into a high-dimensional regenerated Hilbert space, preserving the respective internal properties to the maximum extent and improving the prediction performance of the model.
MMD	The MMD model is to minimize the distributional distance between latent features, followed by inputting these latent features into a predictor for the RUL prediction.
Transformer	The Transformer model uses the idea of an attention mechanism to process time series data. We take all training bearing data as the input word vector matrix and select the important information to improve the model performance by globally associating the weight factors and weighted summation.
AOA	The GAN consists of downsampled convolution and upsampled transposed convolution to form the generator. Resnet18-1d is used to form the discriminator for generating pseudo samples, and then, the pseudo samples, as well as the source domain samples, are trained together with the predictor to predict the RUL.



	bearing 1_2	bearing 1_3	bearing 1_5	bearing 1_6	bearing 2_1	bearing 2_2	bearing 2_3	bearing 2_4	bearing 2_5	bearing 2_6	bearing 3_1	bearing 3_2	average value
Ours	**1.07**	**2.10**	**0.89**	**1.20**	**1.13**	**0.43**	**6.54**	**0.51**	**0.86**	**0.44**	**0.84**	**1.24**	**1.44**
SA	1.24	2.39	1.12	1.38	1.15	0.45	7.78	0.74	0.92	0.85	0.86	1.38	1.69
AE_PCA	1.34	2.56	1.58	1.58	1.47	0.62	7.78	0.95	0.99	0.96	0.85	1.81	1.87
TCA	1.81	2.67	1.44	1.50	1.65	0.52	7.25	0.65	0.90	0.56	**0.84**	1.35	1.76
MMD	1.12	2.23	0.91	1.25	1.14	0.44	7.85	0.70	0.92	0.65	0.85	1.29	1.61
Transformer	1.24	2.20	1.03	**1.20**	1.15	0.45	7.35	0.61	0.90	0.65	0.85	1.66	1.6
AOA	1.27	2.51	1.92	1.47	1.15	0.45	7.88	0.90	0.99	0.96	0.85	1.50	1.82

Figure 12. Experimental results of the comparison models on the PHM 2012 dataset.

It is observed that all the *RMSE* values of our model are lower than the other compared models. Among these, the AE and SA only extract the significant features of the training bearings and ignore the variations in data distribution across the different test bearings; therefore, the two models fail to obtain satisfactory prediction results. The TCA and the MMD models enhance the prediction accuracy by reducing the distance between the source domain bearing and the known target domain bearing. However, since the vibration data of different bearings have the different time series lengths with respect to the full life cycle, the distance metric-based domain adaptive approaches produce a certain distance bias, which has an obvious impact on the prediction performance. The AOA model uses

GAN to generate pseudo samples, which expands the data distribution of the samples, but the expansion range is uncontrollable, which will affect the prediction progress. As the conditions of generalization are strict, the prediction effect will be affected.

However, as can be seen in Figure 12, in the comparison experiments, the difference in prediction accuracy between our proposed model and the comparison model on some bearings is not large, or the results are even the same. Comparing the learning conditions of the models, it can be seen that our proposed single-source domain generalization model relies on only one bearing training and makes predictions under the condition that the target domain bearing is unknown. It can be seen from the figure that the prediction accuracy of the domain adaptive model is also better; the reason is that this kind of model can use the target bearing to perform some adaptive methods, thus bringing the distance between the source domain and the target domain closer, which makes the domain adaptive model perform very well for predictions in cross-domain scenarios. Learning models and domain generalization models, on the other hand, do not perform as well. If the bearings of the target domain are not visible to the domain adaptive model, then its prediction accuracy drops drastically.

It should be pointed out that bearing 2_3 is the worst case among all the tested bearings. This is because the degradation stage of bearing 2_3 lasts for a short time. The data vary widely, and correspondingly, the data distribution of bearing 2_3 differs considerably from the other eleven bearings; therefore, the prediction results of bearing 2_3 perform more poorly than any other test bearings on each comparison model.

4.6. Generalization Error Bound Analysis

It should be pointed out that, in Figure 12, the $RMSE$ of bearing 1_3 and bearing 2_3 are obviously larger and even exceed several times those of the other bearings. Therefore, we analyze the reason from the perspective of the generalization error bound.

The generalization error usually indicates the generalization performance of the model for unknown target data, which are obtained by subtracting the training error from the error expectation over the entire input space. The generalization error bound [33] is the maximum allowed value of the generalization error, beyond which the feasibility of model is problematic, defined as follows: when the space is assumed to be a finite function set $F = \{f_1, f_2, \cdots f_d\}$, the following inequality holds for any function $f \in F$, with a probability of $1 - \delta$ at least.

$$R(f) \leq \hat{R}(f) + \epsilon(d, N, \delta), \tag{11}$$

$$\epsilon(d, N, \delta) = \sqrt{\frac{1}{2\mathrm{N}}\left(1\log \mathrm{d} + \log\frac{1}{\delta}\right)}, \tag{12}$$

where the left-hand side of the inequality $R(f)$ denotes the generalization error, the right-hand side denotes the generalization error bound, $\hat{R}(f)$ denotes the empirical risk, and $\epsilon(d, N, \delta)$ corresponds to a correction quantity, which is a monotonically decreasing function of the corresponding sample N. d denotes the number of functions, and the more functions, the larger the correction. Correspondingly, the empirical risk $\hat{R}(f)$ is defined as follows:

$$\hat{R}(f) = \frac{1}{N}\sum_{i=1}^{N} RMSE(Y_i, \hat{Y}_i), \tag{13}$$

where Y_i represents the true values, and $\hat{Y}_i$ represents the predicted values.

In the PHM 2012 experiment dataset, the amount of training sample N is 1440, and the number of functions d is 150. The range of values of the probability δ is set to [0, 1], and according to Equation (12), the larger δ is, the smaller ϵ is. When the value of δ is set to 1, the minimum value of ϵ is 0.0274. According to Equation (13), $\hat{R}_f$ is 1.422. Following Equation (11), we obtain the value of the generalization error bound as 1.44. The specific results are listed in Table 5.

Table 5. Generalization error bound analysis.

Empirical Risk $\hat{R}(f)$	Sample Size N	Assume Functions d	Probability δ	Correction Quantity ϵ	Generalization Error Bound
1.422	1440	150	1.0	0.024	1.44
			0.00001	0.049	1.47

As can be seen from Figure 13, except for bearing 1_3 and bearing 2_3, all other bearings meet the above generalization error bound inequality on the PHM dataset. This indicates that the model does not have generalization ability for bearings 1_3 and 2_3; therefore, the two prediction result *RMSE* values are larger than that of the other test bearings. The experimental results of our model are in accordance with the theoretical calculations.

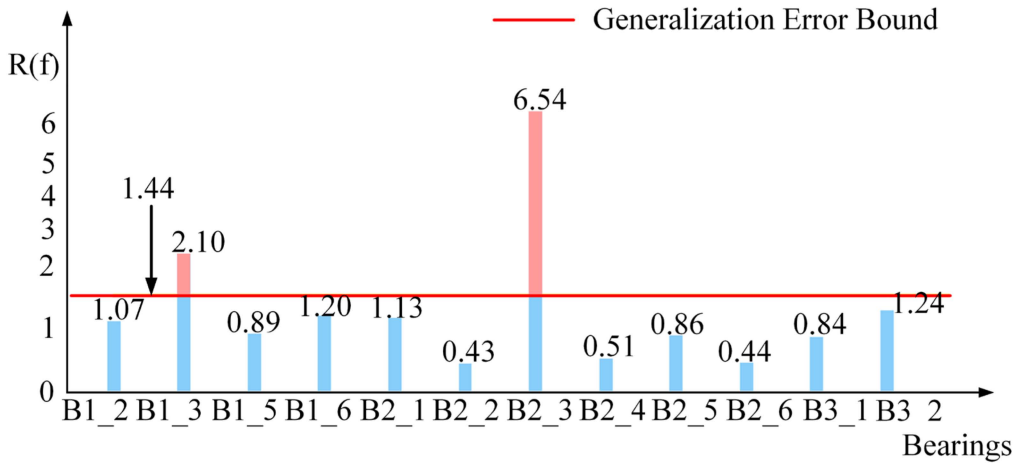


Figure 13. Generalization error analysis of the experimental results.

5. Conclusions

In this paper, to tackle the problem that the unknown target bearing data are unavailable or unknown for model training, we propose a novel single-source domain generalization method for the RUL predictions of bearings, termed as the adaptive stage division and parallel reversible instance normalization model. Firstly, we raise an adaptive threshold stage division approach to determine the degradation point in the full life cycle vibration data of bearings. Further, we explore the instance normalization and denormalization algorithms of the source bearing data and then combine them into a unified GRU-based RUL prediction network, avoiding the distribution bias in data enhancement and concurrently enhancing the generalization performance of the model for unknown bearings. In the two ablation experiments of the PHM2012 dataset and XJTU-SY bearing dataset, it can be seen that the average prediction accuracy (*RMSE* value) of the proposed method is 1.44 and 1.53, respectively, which are 17% and 11% higher than that of the second-best models, i.e., the Adap-stage+IN+RevIN and GRU+RevIN models. In the comparison experiment of the PHM2012 dataset, the average prediction accuracy of the proposed model is 1.44, which is 11% superior to that of the suboptimal comparison model (Transformer model). Comparison of the experimental results shows that the model offers a good generalization performance for predicting the RUL of unknown bearings. This method introduces a novel approach to single-source domain generalization for RUL predictions.

It is noted that the generalization ability of our model on bearings 1_3 and 2_3 is still unsatisfactory. In future works, we can attempt to increase the training samples and explore advanced data augmentation techniques to expand the data distribution, such that the model has a wider generalization error bound and generalization ability.

Author Contributions: Conceptualization, J.X.; writing—original draft preparation, B.M.; formal analysis, W.C.; validation, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 52375089 and JiangHuai Advance Technology Center Dream Fund Project 2023-ZM01J003.

Data Availability Statement: Public datasets used in our paper: https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset (accessed on 3 July 2018) and https://biaowang.tech/xjtu-sy-bearing-datasets (accessed on 29 July 2021).

Conflicts of Interest: Chengwei Shan was employed by CCTEG Changzhou Research Institute Tiandi (Changzhou) Automation Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Camerini, V.; Coppotelli, G.; Bendisch, S.; Kiehn, D. Impact of pulse time uncertainty on synchronous average: Statistical analysis and relevance to rotating machinery diagnosis. *Mech. Syst. Signal Process.* **2019**, *129*, 308–336. [CrossRef]
2. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
3. Xu, J.; Ma, B.; Fan, Y.Q.; Ding, X. ATPRINPM: A single-source domain generalization method for the remaining useful life prediction of unknown bearings. In Proceedings of the 2022 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), Harbin, China, 22–24 December 2022; pp. 1–6.
4. Ding, P.; Jia, M.; Wang, H. A dynamic structure-adaptive symbolic approach for slewing bearings' life prediction under variable working conditions. *Struct. Health Monit.* **2021**, *20*, 273–302. [CrossRef]
5. Kim, T.; Kim, J.; Tae, Y.; Park, C.; Choi, J.H.; Choo, J. Reversible instance normalization for accurate time-series forecasting against distribution shift. In Proceedings of the International Conference on Learning Representations, Beijing, China, 19–21 February 2021.
6. Jiang, Y.; Xia, T.; Wang, D.; Fang, X.; Xi, L. Adversarial Regressive Domain Adaptation Approach for Infrared Thermography-Based Unsupervised Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7219–7229. [CrossRef]
7. Ragab, M.; Chen, Z.; Wu, M.; Foo, C.S.; Kwoh, C.K.; Yan, R.; Li, X. Contrastive Adversarial Domain Adaptation for Machine Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2022**, *17*, 5239–5249. [CrossRef]
8. Chen, H.; Jin, M.; Li, Z.; Fan, C.; Li, J.; He, H. Ms-mda: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition. *Front. Neurosci.* **2021**, *15*, 778488. [CrossRef] [PubMed]
9. Yuan, Y.; Li, Y.; Zhu, Z.; Li, R.; Gu, X. Joint domain adaptation based on adversarial dynamic parameter learning. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *5*, 714–723. [CrossRef]
10. Yang, F.E.; Cheng, Y.C.; Shiau, Z.Y. Adversarial teacher-student representation learning for domain generalization. In Proceedings of the Advances in Neural Information Processing Systems, Beijing, China, 19–23 April 2021; Volume 34, pp. 19448–19460.
11. Zhao, C.; Shen, W. Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis. *IEEE Trans. Ind. Inform.* **2022**, *19*, 2909–2918. [CrossRef]
12. Li, L.; Gao, K.; Cao, J.; Huang, Z.; Weng, Y.; Mi, X.; Yu, Z.; Li, X.; Xia, B. Progressive domain expansion network for single domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Kuala Lumpur, Malaysia, 18–20 December 2021; pp. 224–233.
13. Zhao, K.; Jiang, H.; Wang, K.; Pei, Z. Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis. *Knowl. -Based Syst.* **2021**, *222*, 106974. [CrossRef]
14. Liu, W.; You, J.; Lee, J. Hsigan: A conditional hyperspectral image synthesis method with auxiliary classifier. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3330–3344. [CrossRef]
15. Katsuma, D.; Kawanaka, H.; Surya Prasath, V.B.; Aronow, B.J. Data augmentation using generative adversarial networks for multi-class segmentation of lung confocal if images. *J. Adv. Comput. Intell. Intell. Inform.* **2022**, *26*, 138–146. [CrossRef]
16. Bird, J.J.; Barnes, C.M.; Manso, L.J.; Ekárt, A.; Faria, D.R. Fruit quality and defect image classification with conditional gan data augmentation. *Sci. Hortic.* **2022**, *293*, 110684. [CrossRef]
17. Verma, V.K.; Brahma, D.; Rai, P. Meta-learning for generalized zero-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6062–6069.
18. Wu, L.; Xie, P.; Zhou, J.; Zhang, M.; Chunping, M.; Xu, G.; Zhang, M. Robust self-augmentation for named entity recognition with meta reweighting. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Beijing, China, 16–18 December 2022; pp. 4049–4060.
19. Qiao, F.; Peng, X. Uncertainty-guided model generalization to unseen domains. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Beijing, China, 29–31 October 2021; pp. 6790–6800.

20. Fan, X.; Wang, Q.; Ke, J.; Yang, F.; Gong, B.; Zhou, M. Adversarially adaptive normalization for single domain generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online, China, 19–25 June 2021; pp. 8208–8217.
21. Duan, Y.; Liu, X.; Jatowt, A.; Yu, H.T.; Lynden, S.; Kim, K.S.; Matono, A. Sorag: Synthetic data over-sampling strategy on multi-label graphs. *Remote Sens.* **2022**, *14*, 4479. [CrossRef]
22. Isaksson, L.J.; Summers, P.; Raimondi, S.; Gandini, S.; Bhalerao, A.; Marvaso, G. Mixup (sample pairing) can improve the performance of deep segmentation networks. *J. Artif. Intell. Soft Comput. Res.* **2022**, *12*, 29–39. [CrossRef]
23. Xu, J.; Duan, S.; Chen, W.; Wang, X.; Fan, Y. SACGNet: A remaining useful life prediction of bearing with self-attention augmented convolution GRU network. *Lubricants* **2022**, *10*, 21. [CrossRef]
24. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision (CVPR), Hawaii, HI, USA, 21–26 July 2017; pp. 1501–1510.
25. Wang, H.; Liao, H.; Ma, X.; Bao, R. Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model. *Reliab. Eng. Syst. Saf.* **2021**, *210*, 107504. [CrossRef]
26. Xu, J.; Qian, L.; Chen, W.; Ding, X. Hard negative samples contrastive learning for remaining useful-life prediction of bearings. *Lubricants* **2022**, *10*, 102. [CrossRef]
27. Ren, L.; Sun, Y.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **2018**, *48*, 71–77. [CrossRef]
28. Zhang, Z.; Song, W.; Li, Q. Dual-aspect self-attention based on transformer for remaining useful life prediction. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [CrossRef]
29. Mao, W.; He, J.; Zuo, M.J. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1594–1608. [CrossRef]
30. 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* **2018**, *168*, 108286. [CrossRef]
31. Zou, W.; Lu, Z.; Hu, Z.; Mao, L. Remaining useful life estimation of bearing using deep multi-scale window-based transformer. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3514211. [CrossRef]
32. Ding, Y.; Jia, M.; Cao, Y.; Ding, P.; Zhao, X.; Lee, C.G. Domain generalization via adversarial out-domain augmentation for remaining useful life prediction of bearings under unseen conditions. *Knowl. -Based Syst.* **2023**, *261*, 110199. [CrossRef]
33. Rigollet, P. Generalization error bounds in semi-supervised classification under the cluster assumption. *J. Mach. Learn. Res.* **2007**, *8*, 1369–1392.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.