A Cotraining-Based Semisupervised Approach for Remaining-Useful-Life Prediction of Bearings

Xuguo Yan 1,2,3,*, Xuhui Xia 1,2,3, Lei Wang 1,2,3 and Zelin Zhang 1,2,3

1 Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
2 Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
3 Precision Manufacturing Institute, Wuhan University of Science and Technology, Wuhan 430081, China
* Correspondence: yanxuguo@wust.edu.cn; Tel.: +86-156-2306-1079

Abstract: The failure of bearings can have a significant negative impact on the safe operation of equipment. Recently, deep learning has become one of the focuses of RUL prediction due to its potent scalability and nonlinear fitting ability. The supervised learning process in deep learning requires a significant quantity of labeled data, but data labeling can be expensive and time-consuming. Cotraining is a semisupervised learning method that reduces the quantity of required labeled data through exploiting available unlabeled data in supervised learning to boost accuracy. This paper innovatively proposes a cotraining-based approach for RUL prediction. A CNN and an LSTM were cotrained on large amounts of unlabeled data to obtain a health indicator (HI), then the monitoring data were entered into the HI and the RUL prediction was realized. The effectiveness of the proposed approach was compared and analyzed against individual CNN and LSTM and the stacking networks SAE+LSTM and CNN+LSTM in the existing literature using RMSE and MAPE values on a PHM 2012 dataset. The results demonstrate that the RMSE and MAPE value of the proposed approach are superior to individual CNN and LSTM, and the RMSE value of the proposed approach is 54.72, which is significantly lower than SAE+LSTM (137.12), and close to CNN+LSTM (49.36). The proposed approach has also been tested successfully on a real-world task and thus has strong application value.

Keywords: RUL prediction; cotraining; semisupervised learning; bearings

1. Introduction

The prognostics and health management (PHM) technique has recently been studied in terms of assessing and managing the health status of equipment or components with the help of statistical algorithms or models using large quantities of condition-monitoring data and information [1]. The PHM technique can predict potential failures in advance and combine various equipment or component information to make maintenance decisions, thereby improving the safety of production processes and reducing maintenance costs [2]. RUL prediction is the most challenging technique in PHM [3]. It can be defined as the time interval during which equipment or components can be properly used continuously [4]. In this paper, we study bearings, whose failure can have a significant negative impact on the safe operation of equipment. Therefore, it is urgent and necessary to develop an effective RUL prediction approach to estimate their RUL under multiple operational conditions. Existing RUL prediction techniques mainly include model-based and data-driven approaches [5]. Model-based approaches describe the degradation stages of equipment or components by determining an accurate physical or mathematical model [6]. However, a number of complex equipment or components render it arduous to define an accurate model due to multiple operating environments. With the rapid development of artificial intelligence and machine learning technologies, data-driven approaches have become a remarkable tool for RUL prediction. These approaches are based on data fusion and feature...
extraction of monitoring data and history sensor data from various degradation stages of equipment or components to establishing the mapping relationship between monitoring data and RUL [7]. The approaches do not require a priori knowledge, but are based on existing data and use individual analytical processing to mine implicit associations in the data to perform predictive operations [8]. The model construction of a health indicator (HI) that matches the degradation trend of equipment or components is central to data-driven RUL prediction approaches.

Model construction methods for RUL prediction mainly include statistical modeling and traditional machine-learning and deep-learning methods [9]. Recently, deep learning has become one of the focuses of RUL prediction due to its potent scalability and nonlinear fitting ability [10–12]. Early studies in this area relied on large quantities of training data for model learning. However, in real-world cases, the operation conditions of most complex equipment or components may change and the data distribution shows some variability. This will lead to a sharp drop in the performance of RUL prediction due to the poor robustness and generalization of the model [13]. The usual solution to this problem is to retrain or fine-tune the model parameters, which is a supervised learning process that requires a significant quantity of labeled data for training to improve learning performance. Unfortunately, in real-world cases, there are usually abundant unlabeled data available, but few labeled data [14]. The data labeling process of supervised learning, moreover, can be expensive and time-consuming.

Semisupervised learning can reduce the quantity of required labeled data through exploiting the available unlabeled data in supervised learning to boost the accuracy. It can address the poor generalization problem of supervised learning when there are abundant unlabeled data available, but only a few labeled samples [15]. Cotraining is one of the major semisupervised learning paradigms that iteratively trains two classifiers on two different views, and uses the predictions of either classifier on the unlabeled examples to augment the training set of the other [16]. Cotraining can obtain higher accuracy than traditional supervised learning methods due to the ability to combine multiple views of similar samples and prediction results on multiple classifiers [17]. Nowadays, cotraining and its improved approaches have been successfully applied to natural language processing [18], pattern recognition [19] and other fields. However, comparatively few publications are available about cotraining-based approaches for RUL prediction.

In this study, we innovatively adopt cotraining to RUL prediction of mechanical components. In order to better evaluate the improvement within the existing approaches, CNN and LSTM are selected as the initial network model. Individual CNN and LSTM, and the stacking networks SAE+LSTM and CNN+LSTM are selected for comparison with a PHM 2012 dataset. Since various indicators can evaluate the performance of RUL prediction, in this paper, RMSE and MAPE values are adopted to evaluate the prediction results and RMSE values are selected for comparison. We suppose that through cotraining, more accurate RUL prediction results can be achieved by fully using the deep feature extraction capability of CNN and the long-term memory capability of LSTM for time series data. The main contributions of this paper are as follows:

- We propose a semisupervised learning approach for RUL prediction to address the problem that there are usually abundant unlabeled data available, but few labeled data in the actual production scenario.
- We found little literature available on cotraining for RUL prediction. We innovatively propose a cotraining-based approach for RUL prediction of bearings. A CNN and an LSTM were cotrained on large quantities of unlabeled data to obtain the HI of the bearings, then the monitoring data can be input into the HI to realize the RUL prediction.
- We conducted experiments on open datasets (PHM 2012) and real datasets of gearbox bearings in Wuhan Iron and Steel Company, China. The experimental results verify the effectiveness of the proposed approach.

The remainder of this study is organized as follows. Related work is discussed in Section 2. The proposed cotraining-based RUL prediction approach is presented along
with a detailed description of each step in Section 3. The experimental setup and results are presented in Section 4. Finally, Section 5 presents the conclusions and suggests some possible avenues for further research.

2. Related Work

In this section, we focus on the current state of research on feature extraction and RUL prediction model construction in the data-driven RUL prediction approach. An overview of data-driven RUL prediction process and associated algorithms is given in Figure 1.

Figure 1. Processes and associated algorithms of data-driven RUL prediction approach.

One of the key steps in RUL prediction is to extract effective degradation features from original signals. The accuracy of RUL prediction is determined by the quality of feature extraction. Vibration signal extraction is the most widely used feature extraction technique to obtain the health status of equipment or components. It extracts the feature indicators from the vibration signal’s time domain, frequency domain, and time–frequency domain. Time-domain signal processing algorithms mainly include correlation analysis [20] and time-domain statistical analysis [21]. Frequency-domain signal processing algorithms mainly include spectrum analysis [22], cepstrum analysis [23], envelope analysis [24], order ratio spectrum analysis [25], and holographic spectrum analysis [26]. Time–frequency domain signal processing algorithms mainly include short-time Fourier transform [27], Wigner–Ville distribution [28], empirical mode decomposition methods [29] and wavelet transform [30]. These algorithms are based on manual feature extraction, which frequently require some prior knowledge and experience. The features extracted by these algorithms are mostly low-level features. In recent years, deep learning has shown its unique potential and advantages in feature extraction. Many scholars have applied deep learning to the field of signal feature extraction. Hinchin and Tkouat [31], for example, employed a CNN model to extract rolling bearing vibration signal features and achieved improved results in feature extraction. However, these CNN-based feature extraction algorithms require a large quantity of labeled data for model monitoring and adjustment, but the data labeling process in real-world scenarios can be expensive and time-consuming. In order to solve this problem, Liu [32] provided an unsupervised deep neural network through exploiting unlabeled data to extract high-level vibration signal features of rolling bearings and achieved certain results.

RUL prediction model construction in the data-driven RUL prediction approach mainly includes traditional machine-learning and deep-learning methods. Traditional machine learning-based RUL prediction methods estimate equipment or components’ RUL by identifying patterns of variation from a large quantity of monitoring data. For example, Fumeos [33] developed an online correlation vector regression model and optimized the model for RUL prediction of bearings using heuristic algorithms. The prediction model based on traditional machine learning can meet the needs of RUL prediction; however, it does not consider the deep-level mapping relationship between degradation features and
health status, resulting in a lack of generalization ability of the model. The RUL prediction model constructed by deep-learning methods can solve this problem. Many scholars have applied deep learning to the field of model construction for RUL prediction. Representative network models include: BP neural networks [34], extreme learning machine (ELM) [35], CNN [36], LSTM [37] and deep stacking network model of CNN and LSTM. In recent years, deep stacking network models of CNN and LSTM have received increasing attention from scholars due to their ability to handle chronological and spatial relationships of degradation signals. For example, Mao [38] used the Hilbert–Huang transform to extract time–frequency domain information in vibration signals and used this information as a label for whether the data were in a fault state and trained the information by CNN. After training, the monitoring data were input into the trained CNN, and then LSTM was trained with the output of the penultimate neural layer of the CNN as the input data and the RUL as the label.

In recent years, cotraining and its improved approaches have been successfully applied to various fields. However, comparatively few publications are available about cotraining-based approaches for RUL prediction. The cotraining-based approach for RUL prediction of mechanical components is a valuable area of research.

3. Methods

The detail of proposed cotraining-based RUL prediction approach is presented in this section. The abundant unlabeled data can be fully used in the training process to improve the accuracy of RUL prediction.

3.1. Brief Introduction of Cotraining, CNN and LSTM

Cotraining is an effective semisupervised learning method. It uses unlabeled samples to improve prediction accuracy. In the cotraining process, random sampling is used to gradually select unlabeled samples to train classifiers [39]. An algorithm flowchart of cotraining is shown in Figure 2.

![Figure 2. Algorithm flowchart of cotraining.](image)

First, the labeled data are divided into two views to obtain the data representation under two different views, and two different classifiers are trained using different views as the initial classifier. Then, the initial classifier is used to estimate the label confidence of unlabeled samples, and high-confidence samples are added to the labeled data to further iterative training and optimize the classifier. When all unlabeled samples are self-labeled by the classifier, the training model ends.

Define a sample space $x = x_1 \times x_2$, where $x_1$ and $x_2$ correspond to two different “views” of the same sample. The process of standard cotraining algorithm is shown in Algorithm 1:
Algorithm 1: The standard cotraining algorithm

**Input:** a set $L$ of labeled training samples
a set $U$ of unlabeled samples

**Process:**
1. Create a pool $U'$ of samples by choosing $u$ samples at random from $U$
2. Loop for $k$ iterations:
   - Use $L$ to train a classifier $h_1$ that considers only the $x_1$ portion of $x$
   - Use $L$ to train a classifier $h_2$ that considers only the $x_2$ portion of $x$
   - Allow $h_1$ to label $p$ positive and $n$ negative samples from $U'$
   - Allow $h_2$ to label $p$ positive and $n$ negative samples from $U'$
   - Add these self-labeled samples to $L$
3. Randomly choose $2p + 2n$ samples from $U$ to replenish $U'$

Step 1: Define the labeled training set $L$ and the unlabeled dataset $U$;
Step 2: Randomly select $u$ samples from $U$ to create sample buffer pool;
Step 3: Consider two views $x_1$ and $x_2$, and train the classifiers $h_1$ and $h_2$ using $L$;
Step 4: All samples in $U$ are labeled with $h_1$, from which $p$ positive and $n$ negative samples are selected with high confidence. $h_2$ is treated in the same way;
Step 5: Add these self-labeled samples to $L$, that is, choose $p + n$ samples from $h_1$ to $x_2$, and choose $p + n$ samples from $h_2$ to $x_1$. Then randomly choose $2p + 2n$ samples from $U$ to replenish $U'$;
Step 6: Iterate Step 3 to Step 5 $k$ times.

Cotraining starts with training both classifiers on the labeled training set, and then classifier A labels a portion of the unlabeled dataset, generating pseudolabels for the unlabeled samples. Then, samples with high labeling confidence are selected and added to the training set of classifier B. Similarly, classifier B also labels a portion of the unlabeled dataset, selects those with high labeling confidence and adds to the training set of classifier A. Labeling each other until the maximum number of iterations or no samples with high confidence are added. In this way, the training set of the classifier is expanded continuously, allowing the classifier to learn more knowledge.

In the cotraining algorithm, unlabeled sample with the highest labeling confidence is the sample that is most consistent with the labeled sample of the classifier after labeling. It is to maximize:

$$
\Delta_u = \frac{1}{|L|} \sum_{x_i \in L} (y_i - h(x_i))^2 - \frac{1}{|L|} \sum_{x_i \in L} (y_i - h'(x_i))^2
$$

in the sample set $U$, where $h$ denotes the model learned by the current classifier, $L$ denotes the labeled training set, $x_i \in L$ denotes the unlabeled samples, and $h'$ denotes the classifier obtained by adding the $h$ labeled samples to the training set and retraining them. This $\Delta$ function is the classifier prediction error before adding the labeled samples minus the classifier prediction error trained after adding the labeled samples. If $\Delta > 0$, it means that the performance of the classifier has improved, and the labeled sample with the largest $\Delta$ value is the one with the highest confidence.

3.2. Cotraining in RUL Prediction

CNN is one of the most representative algorithms in deep learning. It has two critical structural layers: convolutional and pooling. The convolutional layer computes the convolutional operation of the input data using kernel filters to extract fundamental features. The pooling layer is usually followed to the convolutional layer. In the pooling layer, subsampling is applied to reduce the dimension and avoid overfitting. The typical architecture of CNN is shown in Figure 3.
In RUL prediction, CNN performs error back-propagation based on the BP algorithm. By combining optimization methods such as gradient descent algorithm to train the weight parameters of each layer, the local feature extraction of the input data can be achieved, and the abstract high-dimensional spatial features can be produced. It is then fitted by a fully convolutional neural network (FCN) to achieve the prediction of RUL.

The traditional recurrent neural network (RNN) has certain memory ability. However, when dealing with long-time series, RNN is prone to gradient explosion or disappearance and cannot learn the relevant information of input data. LSTM is a variant of RNN used in deep learning and has long-term memory ability. The architecture of an LSTM consists of units called memory cells, and the memory capacity of the cells can be improved by introducing “gates” into the cells. The LSTM cell has been transformed and generalized by many researchers in recent years, which mainly include LSTM with forget gates, LSTM without forget gates, and LSTM with peephole connections. Considering that LSTM with forget gates is the most widely used LSTM unit, this study takes it as the basic unit structure of the LSTM unit. The internal structure of this unit is shown in Figure 4. According to the figure, the internal operation process of the LSTM model is as follows:

\[
\begin{align*}
    f_t &= \sigma(W_{fh}h_{t-1} + U_{fx}x_t + b_f) \quad (2) \\
    i_t &= \sigma(W_{ih}h_{t-1} + U_{ix}x_t + b_i) \quad (3) \\
    o_t &= \sigma(W_{oh}h_{t-1} + U_{ox}x_t + b_o) \quad (4) \\
    c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (5) \\
    h_t &= o_t \cdot \tanh(c_t) \quad (6)
\end{align*}
\]

where \(f_t, i_t, o_t\) denote the forget gate, input gate and output gate at moment \(t\). \(c_t, h_t, x_t\) denote the cell state, hidden state and cell input at the moment \(t\). \(W\) and \(U\) denote the weights of the hidden state and the cell input.
The forget gate decides which information from the previous cell state should be forgotten. When $f_t = 1$, the information will be completely retained; when $f_t = 0$, the information will be completely forgotten.

A general flowchart of cotraining in RUL prediction is given in Figure 5. First, two prediction networks are selected and trained separately on the failure data (labeled training data). Then, the interrupt data (unlabeled training data) are labeled, and the samples with high confidence are added to each other’s training set to expand the training set continuously. In this paper, we set CNN as the prediction Network 1 and LSTM as the prediction Network 2.

### Figure 5. Cotraining in RUL prediction.

Setting CNN as the prediction Network 1 and LSTM as the prediction Network 2, the pseudocode of RUL prediction based on cotraining CNN and LSTM is as follows (Algorithm 2):

---

**Algorithm 2: Pseudocode of RUL prediction based on cotraining**

**Input:**
- $L$—Failure training dataset
- $U$—Suspension training dataset
- $T$—Maximum number of cotraining iterations
- $u$—Suspension pool size

**Training Process:**
1. $L_1 = L; L_2 = L$
2. $h_1 = \text{TrainFun} (L_1, 1); h_2 = \text{TrainFun} (L_2, 1)$
3. **Repeat** $T$ times
   4. Create a pool $U'$ of $u$ suspension units by random sampling from $U$
   5. **for** $j = 1$ to 2
   6. **for** each $X_u \subset U'$
   7. $L_u^j = h_j(X_u)$
   8. $h'_j = \text{TrainFun} \left( L_j \cup \{X_u, L_u^j\}, j \right)$
   9. $\Delta_jX_u = \sum \left( L_j^T - h_j(X_u) \right)^2 - \sum \left( L_j^T - h'_j(X_u) \right)^2$
   10. **end**
   11. **if** there exists an $\Delta_jX_u > 0$
      12. $X^*_j = \arg \max_{X_u \subset U'} \left( \Delta_jX_u \right); L^*_j = h_j(X^*_j)$
      13. $\pi_j = \left\{ (X^*_j, L^*_j) \right\}; U' = U' \setminus \pi_j$
   14. **else**
      15. $\pi_j = \emptyset$
   16. **end**
   17. **end**
   18. **if** $\pi_1 = \emptyset$ & & $\pi_2 = \emptyset$ **exit**
   19. **else** $L_1 = L_1 \cup \pi_2; L_2 = L_2 \cup \pi_1$
   20. $h_1 = \text{TrainFun}(L_1, 1); h_2 = \text{TrainFun}(L_2, 2)$
   21. **end**

**Testing Process**
22. $L^T = \omega_1 h_1(X) + \omega_2 h_2(X)$ for any test data $X$
After obtaining the final prediction result $L^P = \omega_1 h_1(x) + \omega_2 h_2(x)$, the sequential quadratic programming (SQP) method can be used to optimize the ensemble weights of different prediction networks. The formula is as follows:

$$\min E = \sum_{x_i \in \psi} (L_i - (\omega_1 h_1(x_i)) + \omega_2 h_2(x_i))^2$$

subject to:

$$\begin{align*}
\omega_1 + \omega_2 &= 1 \\
0 &\leq \omega_1 \leq 1 \\
0 &\leq \omega_2 \leq 1
\end{align*}$$

(7)

3.3. The Bearing RUL Prediction Process Based on Cotraining

The process of cotraining-based approach for RUL prediction of bearings is shown in Figure 6. The specific steps are as follows:

- Extracting time, frequency and time–frequency domain features from vibration signals, conducting feature selection to reduce the input data volume of the network, and splitting the data into training set and test set,
- Determining the degradation starting point and the interrupt operation point of the bearing according to the degradation stages.
- Cotraining CNN and LSTM with labeled training data to obtain the HI.
- Inputting the test set data into cotrained model to obtain the HI of test bearings and predicting their RUL.

Figure 6. The process of cotraining-based RUL prediction of bearings.
4. Experimental Results

4.1. Experimental Results on Benchmark Dataset

4.1.1. Dataset

Rolling bearings are an essential part of mechanical equipment. Although the bearing fault can be visually displayed in disassembly and replacement process, it is difficult to evaluate quantitatively the degree of failure and monitor the running status information in real time. Features extracted from vibration signal can determine the health status of rolling bearings.

The IEEE Reliability Institute and the FEMTO-ST Institute organized the IEEE PHM 2012 Data Challenge in 2012 [40]. The challenge provided a dataset for predicting the RUL of the bearings. In this paper, we used this dataset to train and evaluate the proposed approach. The experimental platform is shown in Figure 7.

![Figure 7. The experimental platform of PHM 2012.](image1)

In the rotating part, the motor power is 250 W, and the maximum speed is 2830 rpm, which can ensure the speed of the second shaft is 2000 rpm. In the load part, this part is a pneumatic jack, which provides 4000 N dynamic load for the bearing. The load diagram is shown in Figure 8.

![Figure 8. The load diagram of rotating part.](image2)
In the testing part, the degradation data of bearing mainly consists of two parts, namely, vibration data and temperature data. The vibration sensor consists of two micro-accelerometers positioned at 90° to each other. The first one is placed on the vertical axis and the second one is placed on the horizontal axis. Two accelerometers are placed on the outer ring of the bearing along the radial direction, and the sampling frequency is 25.6 kHz. The temperature sensor is a resistance temperature detector placed in a hole close to the outer bearing ring with a sampling frequency of 0.1 Hz (Figure 9).

![The vibration sensor.](image)

Therefore, the dataset includes three different working conditions:

- Load 4000 N, speed 1800 rpm
- Load 4200 N, speed 1650 rpm
- Load 5000 N, speed 1500 rpm

The specific training set and test set are shown in Table 1.

| Datasets     | Operation Conditions                  |
|--------------|---------------------------------------|
| Training set |                                       |
| Bearing 1-1  | Bearing 2-1                            |
| Bearing 1-2  | Bearing 2-2                            |
| Bearing 1-3  | Bearing 2-3                            |
| Bearing 1-4  | Bearing 2-4                            |
| Bearing 1-5  | Bearing 2-5                            |
| Bearing 1-6  | Bearing 2-6                            |
| Bearing 1-7  | Bearing 2-7                            |
| Bearing 3-1  | Bearing 3-2                            |
| Test set     |                                       |
| Bearing 1-3  | Bearing 2-3                            |
| Bearing 1-4  | Bearing 2-4                            |
| Bearing 1-5  | Bearing 2-5                            |
| Bearing 1-6  | Bearing 2-6                            |
| Bearing 1-7  | Bearing 2-7                            |
| Bearing 3-3  |                                       |

4.1.2. Health Indicator (HI) Construction and Health Stage Division

Features extracted from vibration signals can determine the health status of rolling bearings. However, the original signal obtained from the sensor is the high-dimensional time-series data mixed with external noise, which makes it unsuitable for use directly in health-status monitoring. Feature extraction techniques and methods, in this case, can map the high-dimensional data into low-dimensional features to reduce the redundant information. Deep learning can extract deep features of degradation in vibration signal and avoid the interference of subjective factors from manual feature extraction.

In real-world tasks, there are numerous factors that affect the health status of the rolling bearing in different ways. If all the parameters are considered as the input of the network, problems such as training difficulties and overfitting may occur, resulting in inaccurate forecasting. In order to reduce the complexity of network training, the time, frequency and time–frequency domain-based features of the signal were selected first to reduce the input data volume of the network.
Time domain-based features include dimensional and dimensionless features. The dimensional feature can well represent the degradation trend of the rolling bearing but are not sensitive to rolling failure, which can be largely influenced by different working conditions. Dimensionless features are not significantly influenced by different working conditions and are more sensitive to rolling failure. If the signal sequence collected by sensor in each sampling time period is $x_i$, $x_i = [x_1, x_2, x_3, \cdots, x_N]$, where $N$ is the number of sampling points, Tables 2 and 3 give the dimensional and dimensionless features used in this paper.

### Table 2. Dimensional time domain-based features.

| No. | Feature                      | Expression                  |
|-----|------------------------------|-----------------------------|
| 1   | Root Mean Square Value       | $X_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$ |
| 2   | Mean Value                   | $\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i$ |
| 3   | Standard Deviation           | $X_e = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ |
| 4   | Square Root Amplitude        | $X_r = \left[\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right]^2$ |
| 5   | Absolute Mean Amplitude      | $X_p = \frac{1}{N}\sum_{i=1}^{N}|x_i|$ |
| 6   | Peak Value (Maximum Value)   | $X_{max} = \max\{|x_i|\}$ |
| 7   | Peak-to-Peak Value           | $X_{p-p} = \max(x_i) - \min(x_i)$ |

### Table 3. Dimensionless time domain-based features.

| No. | Feature                      | Expression                  |
|-----|------------------------------|-----------------------------|
| 1   | Skewness                     | $X_{ske} = \sum_{i=1}^{N}(x_i - \bar{x})^3/(N-1)\bar{x}$ |
| 2   | Kurtosis                     | $X_{kur} = \sum_{i=1}^{N}(x_i - \bar{x})^4/(N-1)\bar{x}^2$ |
| 3   | Skewness Factor              | $I_{ske} = X_{ske}/X_{RMS}$ |
| 4   | Kurtosis Factor              | $I_{kur} = X_{kur}/X_{RMS}$ |
| 5   | Peak Value Factor            | $I_p = X_{max}/X_{rms}$ |
| 6   | Impulse Factor               | $I_i = X_{p}/X_{r}$ |
| 7   | Waveform Factor              | $I_w = X_{rms}/X_{p}$ |
| 8   | Margin Factor                | $I_m = X_{max}/X_{r}$ |

Frequency domain-based features can reflect the failure type and the corresponding failure degree of the rolling bearing. These features of the signal are extracted from the spectral signal. Spectral signal is obtained by Fourier transformation of the time-domain signal, which describes the frequency components of original signal and the amplitude of each frequency component. Fast Fourier transformation (FFT) can reduce the amount of calculation and improve the calculation speed. In real-world tasks, FFT is often used to calculate the frequency spectrum of signal. The calculation formula is as follows:

$$X_N(k) = \begin{cases} X_1(k) + W_N^kX_2(k) & k = 0, \cdots, \frac{N}{2} - 1 \\ X_1\left(k - \frac{N}{2}\right) - W_N^kX_2\left(k - \frac{N}{2}\right) & k = \frac{N}{2}, \cdots, N - 1 \end{cases}$$

where $X_1(k)$ is the discrete Fourier transform of the even items in the time-series $x(i)$, and $X_2(k)$ is the discrete Fourier transform of the odd items.
If the sampling frequency of the original signal is $F_o$, after FFT, the frequency of the $k$th frequency point in the spectrum is $f_k = (k - 1) * F_o / N$. Since the signal spectrum is obtained, frequency domain-based features can be obtained from the spectrum. Table 4 gives the frequency domain-based features used in this paper.

Table 4. Frequency domain-based features.

| No. | Feature              | Expression                        |
|-----|----------------------|-----------------------------------|
| 1   | Gravity Frequency    | $S_{FC} = \frac{\sum_{k=0}^{N-1} fX(k)}{N}$ |
| 2   | Frequency-Domain Amplitude Average | $S = \frac{1}{N} \sum_{k=0}^{N-1} X(k)$ |
| 3   | Frequency-Domain Standard Deviation | $S_{RVF} = \sqrt{\frac{\sum_{k=0}^{N-1} (f - S)^2 X(k)}{\sum_{k=0}^{N-1} fX(k)}}$ |
| 4   | Root-Mean-Square Frequency | $S_{RMSF} = \sqrt{\frac{\sum_{k=0}^{N-1} f^2 X(k)}{\sum_{k=0}^{N-1} fX(k)}}$ |

Time domain- and frequency domain-based features can intuitively show parts of the inherent information of the original signal. However, since original vibration signal of the rolling bearing is a nonstationary signal, it is difficult to accurately describe the change law of original signal only using time domain- and frequency domain-based features. The time–frequency analysis method is introduced in this case to analyze the original signal. The wavelet packet decomposition is a commonly used time–frequency analysis method, which can well analyze nonstationary signals. Compared with wavelet decomposition, wavelet packet decomposition can decompose both the low-frequency and the high-frequency part of the signal. Original signal after wavelet packet decomposition will be decomposed into each sub-band, so as to realize further time–frequency localized analysis. Figure 10 shows the structure of three-layer wavelet packet decomposition, where $x(i)$ denotes the original vibration signal, $H$ denotes the low-frequency component, and $G$ denotes the high-frequency component.

![Figure 10. The structure of three-layer wavelet packet decomposition.](image)

The original signal is decomposed into two parts, high frequency and low frequency, through wavelet packet decomposition. The feature information of the original signal is retained while obtaining the deep information, which is beneficial to the analysis of nonstationary signals. The base wavelet has great influence on feature extraction of the bearing. In this paper, we select the base wavelet according to the variation rate of energy fluctuation. First, the energy of each band is calculated as a percentage of the overall signal energy $E_n$, where $n = 1, 2, \cdots, 2^j$, and then the energy fluctuation parameter of $E_{flu}$ is defined as:

$$E_{flu} = \frac{\max (E_n^i) - \text{mean} (E_n^i)}{\max (E_n^i) - \min (E_n^i)} \quad (9)$$

As can be seen from the equation, calculate the $E_{flu}$ is a normalized gauge, and the value of $E_{flu}$ is between [0, 1] as the signal changes during transmission. The energy
distribution is more uniform in the actual health state of the bearings, while the energy distribution is unbalanced in the fault state, and the two values may differ significantly. Therefore, in order to make the data in different states comparable, based on Equation (9), we can calculate the energy function parameters corresponding to the vibration signal under normal and fault state of the bearing: $E_{nor}, E_{fau}$, then calculate the rate of change of energy fluctuations:

$$E' = \frac{(E_{fau} - E_{nor})}{E_{nor}} \times 100\%$$  \hspace{1cm} (10)$$

The larger $E'$ is, the more the features of the fault signal deviate from the normal signal and the better the fault feature extraction is. Wavelet basis function corresponding to $E'_{\text{max}}$ is the most optimal wavelet basis function for decomposing the bearing signal using wavelet packets.

In this paper, we choose db3, db8, haar and db4 for comparison. The respective energy fluctuation parameters and the rates of change are shown in Table 5. It can be seen from the table that the maximum energy fluctuation rate of change $E'_{\text{max}}$ is the haar wavelet basis function, so it is the optimal wavelet basis function for the decomposition of the bearing signal.

Table 5. Energy fluctuation parameters and the rates of change.

| Wavelet Function | Energy Fluctuation Parameters | Rate of Change $E'/\%$ |
|------------------|-------------------------------|------------------------|
| db3              | Normal                        | 0.48                   | 77.47                  |
|                  | Fault                         | 0.86                   |                        |
| db8              | Normal                        | 0.51                   | 67.48                  |
|                  | Fault                         | 0.86                   |                        |
| haar             | Normal                        | 0.47                   | 87.41                  |
|                  | Fault                         | 0.86                   |                        |
| db4              | Normal                        | 0.49                   | 75.70                  |

We use haar wavelet basis function to decompose the signal of the bearing with three layers of wavelet packets. Eight sub-bands are obtained and the energy ratio of the sub-bands is used as the time–frequency based feature. The energy feature of each sub-band is defined as follows:

$$E^l_j = \sum_{i=1}^{n} (x^l_j(i))^2$$  \hspace{1cm} (11)$$

where $j$ denotes the number of decomposition layers, $l$ denotes the number of nodes obtained by decomposing each layer, and $n$ denotes the length of the node signal $x^l_j(i)$. The energy ratio after wavelet packet decomposition is:

$$p^l_j = \frac{E^l_j}{\sum_{l'=0}^{y-1} E^l_j}$$  \hspace{1cm} (12)$$

According to the relevant research, when using vibration signals to track bearing degradation, the horizontal vibration signals often contain more degradation information than the vertical vibration signals. In this paper, we use only the horizontal vibration signals of bearings as experimental data for subsequent research. Taking the bearing 1-1 as an example, Figure 11 shows a schematic diagram of the time-domain, frequency-domain, and time–frequency domain features of the horizontal vibration signal. After all features are extracted, the features need to be standardized by the following formula:

$$x_i = \frac{2x_i - \max(x) - \min(x)}{\max(x) - \min(x)}$$  \hspace{1cm} (13)$$
However, not all of the above feature parameters can better reflect the degradation state of the bearing. In order to prevent redundant features from having a negative impact on the accurate evaluation of the bearing degradation state, the above feature parameters need to be further screened to remove the feature parameters that are not sensitive to the bearing state. In general, the feature parameters that can better describe the bearing degradation state should have good monotonicity, robustness and high correlation with the bearing degradation process. At the same time, they should have a certain ability to identify different stages of bearing degradation. Therefore, we choose monotonicity, correlation, robustness and identifiability as indicators to further screen the selected feature parameters. At the same time, in order to more comprehensively evaluate the sensitivity of the degradation features, the four evaluation indicators are linearly combined to obtain a comprehensive indicator, and the formula is as follows:

$$F(X) = w_1 \text{Mon}(X) + w_2 \text{Corr}(X, T) + w_3 \text{Rob}(X) + w_4 \text{Ide}(X, C)$$  \hspace{1cm} (14)

where $T$ is the time, $C$ is the life stage of the bearing ($C = [C_1, C_2, \ldots, C_n]$), and $w_1, w_2, w_3, w_4$ is the corresponding weights of monotonicity index, correlation index, robustness index.
and identifiability index, respectively. The larger the value of comprehensive index $F$, the more sensitive the selected feature is to the degradation process of bearings.

Since the degradation process of bearings is a monotonous and irreversible process, the selected degradation features need to reflect the overall degradation trend of bearings, so the monotonicity of degradation features should occupy a relatively large weight in the comprehensive index. At the same time, in the subsequent experiments, we found that most of the robustness indicators of the extracted features are at a relatively high level, which reduces the discrimination of the robustness indicators on such features and is not conducive to the screening of degradation features. Therefore, the weight of the robustness indicators in the comprehensive indicators should be reduced. To sum up, we set the weights of the monotonicity, correlation, robustness and identifiability indicators in the comprehensive indicators to 0.4, 0.2, 0.2 and 0.2 respectively. We screen the degradation features under each working condition and select the first 10 features with the largest comprehensive index to build the degradation feature set. The degradation features we finally obtained include: (1) frequency-domain amplitude average, (2) root mean square, (3) square root amplitude, (4) peak-to-peak value, (5) impulse factor, (6) peak value factor, (7) kurtosis factor, (8) peak value, (9) waveform factor, (10) first frequency sub-band energy ratio of the three-layer wavelet packet decomposition. Taking the bearing 1-1 as the example, the obtained features are shown in Figure 12.

Theoretically, the nondegraded stage of the rolling bearing can hardly provide any degradation information, and the monitoring signal of the nondegraded stage and the degraded stage are two independent distributions. The training set of the neural network is not suitable for two or more distributions, so the neural network is not adoptable to learn the information from the nondegraded stage. According to the degradation trend of the bearing during the whole operating time, this paper divides bearing degradation process into three stages: stable degradation, rapid degradation, and rapid failure period.

- Stable degradation period: the rolling bearing is in normal operating conditions, but the degradation has begun and will continue, and the signal features of degradation are not obvious;
- Rapid degradation period: the operation of the rolling bearing becomes more and more unstable, and the signal features of degradation are extremely obvious;
- Rapid failure period: it is the period from rapid degradation to bearing failure, the features of failure are extremely obvious.

Three stages of bearing degradation process are defined as follows:

$$
T = \begin{cases} 
\text{Stable degradation period} & 0 < t < t_0 \\
\text{Rapid degradation period} & t_0 < t < t_1 \\
\text{Rapid failure period} & t_1 < t 
\end{cases}
$$

(15)

In the testing phase of the model, the HI of the test rolling bearing can be obtained by inputting test set rolling bearing features into the trained network model, as shown in Figure 13a. In the figure, $\hat{t}_0$ denotes the predicted degradation starting value, and $\hat{H}_I(t)$ denotes the rolling bearing HI constructed by the predictive model. From the obtained rolling bearing HI, RUL prediction can be achieved (Figure 13b). Generally, the rolling bearing RUL prediction can be divided into three stages. The first period is before the degradation starting point, the rolling bearing is in a healthy status in this period, only general attention is required, and the RUL index remains relatively stable (we generally define this period as when the RUL $\geq 55\%$). The second period is after the bearing enters the rapid degradation period, the RUL decreases with the reduction of HI, and it is a period that requires careful attention. The third period is when the rolling bearing enters the rapid failure period (we generally define this period as when the RUL $\leq 5\%$). Some unstable changes may occur at any time, and it is prone to catastrophic damage if the rolling bearing continues to operate.
Figure 12. The obtained features of bearing 1-1 (10 s): (a) frequency-domain amplitude average, (b) root mean square, (c) square root amplitude, (d) peak-to-peak value, (e) impulse factor, (f) peak value factor, (g) kurtosis factor, (h) peak value, (i) waveform factor, (j) first frequency sub-band energy ratio of the three-layer wavelet packet decomposition.

Figure 13. RUL prediction of rolling bearing: (a) the HI of the test bearing; (b) RUL prediction.
The bearing degradation starting point needs to be determined first to construct a more accurate HI model. The degradation starting point of the rolling bearings in PHM 2012 dataset under three operating conditions is shown in Table 6.

Table 6. The degradation starting point of the rolling bearing in PHM 2012 dataset.

| Bearing | Degradation Starting Point (10 s) | Bearing | Degradation Starting Point (10 s) | Bearing | Degradation Starting Point (10 s) |
|---------|----------------------------------|---------|----------------------------------|---------|----------------------------------|
| 1-1     | 1325                             | 2-1     | 875                              | 3-1     | 491                              |
| 1-2     | 827                              | 2-2     | 195                              | 3-2     | 1585                             |
| 1-3     | 1352                             | 2-3     | 1946                             | 3-3     | 312                              |
| 1-4     | 1083                             | 2-4     | 742                              |         |                                  |
| 1-5     | 2410                             | 2-5     | 2263                             |         |                                  |
| 1-6     | 2415                             | 2-6     | 686                              |         |                                  |
| 1-7     | 2198                             | 2-7     | 222                              |         |                                  |

It can be seen from the table that the degradation starting point of the bearing can be varied due to the influence of various operating conditions. The earliest and latest degradation starting points are at 7% and 94% of the whole bearing life cycle. In order to improve RUL prediction efficiency, we set the interrupt operation of the bearing at 95% of the whole life cycle (Figure 14). The bearing that exceeds 95% of its whole life cycle is seen as entering the rapid failure period. Taking the degradation signal features from 5% to 95% of its whole life cycle as the dataset and the degradation percentage as RUL output label.

Figure 14. Interrupt data (unlabeled).

4.1.3. Comparison and Analysis of RUL Prediction Results

In this paper, a CNN and an LSTM are trained separately using a small quantity of labeled data. The two networks are then co-trained on large quantities of unlabeled data, adding unlabeled samples with high confidence to each other’s training sets to obtain rolling bearing HI. Finally, the monitoring data are input into HI model to obtain the HI of the monitoring bearing, and the RUL prediction of the monitoring rolling bearing is realized. The network parameters of CNN and LSTM are shown in Tables 7 and 8.
Various indicators can evaluate the performance of RUL prediction. In this paper, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) are selected to evaluate the prediction results, which are calculated as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_{pre})^2}
\]

(16)

\[
MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - y_{pre}}{y_i} \right|
\]

(17)

where \(y_i\) and \(y_{pre}\) denote the prediction result and the true value at moment \(i\), and \(n\) denotes the number of samples in the test set.

We first compare the prediction results of proposed approach with individual CNN and LSTM under same operation condition. Taking the rolling bearing under operation condition 1 as an example, bearing 1-1 and bearing 1-2 are selected as the failure data (with labels), and the remaining bearings as the interrupt data (unlabeled). Interrupt the data at 95% of the whole life cycle, and add one rolling bearing at a time as interrupt training data. CNN and LSTM are trained separately using the small quantity of labeled data. The two networks are then cotrained on large amounts of unlabeled data, adding unlabeled samples with high confidence to each other’s training sets to obtain rolling bearing HI. Finally, the monitoring data are input into HI model to obtain the HI of the monitoring bearing, and the RUL prediction of the monitoring rolling bearing is realized.

The RUL prediction curves obtained by continuously increasing interrupt training data are compared with the actual RUL in Figure 15. It can be seen from the figure that by increasing interrupt training data, the RUL prediction curve is much closer to the true RUL. It demonstrated that the learning performance of cotraining can be improved by continuously increasing the training data. It also verifies that when using semisupervised learning for prediction tasks, within a certain range, the higher percentage of the data is used for training, the better the results will obtain.
Various indicators can evaluate the performance of RUL prediction. In this paper, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) are selected to evaluate the prediction results, which are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_{ij} - \hat{y}_{ij})^2}$$  

(16)

$$MAPE = \frac{100}{n} \sum \frac{|y_{ij} - \hat{y}_{ij}|}{y_{ij}}$$  

(17)

where $y_{ij}$ and $\hat{y}_{ij}$ denote the prediction result and the true value at moment $i$, and $n$ denotes the number of samples in the test set.

We first compare the prediction results of proposed approach with individual CNN and LSTM under same operation condition. Taking the rolling bearing under operation condition 1 as an example, bearing 1-1 and bearing 1-2 are selected as the failure data (with labels), and the remaining bearings as the interrupt data (unlabeled). Interrupt the data at 95% of the whole life cycle, and add one rolling bearing at a time as interrupt training data. CNN and LSTM are trained separately using the small quantity of labeled data. The two networks are then cotrained on large amounts of unlabeled data, adding unlabeled samples with high confidence to each other's training sets to obtain rolling bearing HI. Finally, the monitoring data are input into HI model to obtain the HI of the monitoring bearing, and the RUL prediction of the monitoring rolling bearing is realized.

The RUL prediction curves obtained by continuously increasing interrupt training data are compared with the actual RUL in Figure 15. It can be seen from the figure that by increasing interrupt training data, the RUL prediction curve is much closer to the true RUL. It demonstrated that the learning performance of cotraining can be improved by continuously increasing the training data. It also verifies that when using semisupervised learning for prediction tasks, within a certain range, the higher percentage of the data is used for training, the better the results will obtain.

![Figure 15. RUL prediction vs. true RUL (under operation condition 1).](image)

The RMSE and MAPE values of the test set on CNN, LSTM, and the proposed cotraining are provided in Table 9. It can be seen from the table that CNN performs better than LSTM in RUL prediction with a small quantity of labeled data, and the cotraining can obtain better prediction results than individual CNN and LSTM. Meanwhile, by adding interrupt training data, the RUL prediction results can be further improved. However, it is worth noting that the RMSE and MAPE value of cotraining become larger after adding four interrupt training data compared to adding three interrupt training data. A reasonable explanation is that the accuracy of RUL prediction has entered a “bottleneck.” Generally, the effective method to improve prediction accuracy is mainly to increase or replace variables rather than expanding the training sample size.

| Interrupt Training Data | Evaluation Indicator | CNN | LSTM | Cotraining |
|-------------------------|----------------------|-----|------|------------|
| None                    | RMSE                 | 83.44 | 79.48 | 78.05 |
|                         | MAPE                 | 19.80 | 17.65 | 17.01 |
| 1                       | RMSE                 | -    | -    | 63.41 |
|                         | MAPE                 | -    | -    | 13.56 |
| 2                       | RMSE                 | -    | -    | 58.19 |
|                         | MAPE                 | -    | -    | 14.33 |
| 3                       | RMSE                 | -    | -    | 55.06 |
|                         | MAPE                 | -    | -    | 13.68 |
| 4                       | RMSE                 | -    | -    | 55.73 |
|                         | MAPE                 | -    | -    | 13.89 |

For comparison and analysis, the RUL prediction result under different operation conditions, bearing 1-3, 2-3 and 3-3 are selected as the test bearings, and the remaining 14 bearings are the training bearings. The RMSE and MAPE of training process and testing process are shown in Tables 10 and 11, namely, the test bearing HI construction error and the test bearing RUL prediction error.

| Bearings  | RMSE | MAPE |
|-----------|------|------|
| Bearing 1-3 | 0.0765 | 0.0489 |
| Bearing 2-2 | 0.0724 | 0.0563 |
| Bearing 3-3 | 0.0420 | 0.0267 |
Table 11. Test bearing RUL prediction error.

| Bearings | RMSE (10 s) | MAPE (10 s) |
|----------|-------------|-------------|
| Bearing 1-3 | 59.33       | 19.31       |
| Bearing 2-2 | 54.72       | 11.06       |
| Bearing 3-3 | 4.11        | 5.70        |

Since there are few publications available about cotraining-based approaches for RUL prediction. In order to verify the improvements of the proposed approach, this paper compares the RUL prediction results of the proposed cotraining with the SAE+LSTM stacking network [41] and the CNN+LSTM stacking network [38] in the existing literature using RMSE on PHM 2012 dataset. By taking the bearing 2-2 as the example, the RMSE value of the SAE+LSTM stacking network is 137.12, the RMSE value of the CNN+LSTM stacking network is 49.36, and the RMSE value of the proposed cotraining is 54.72. It is obviously that the RUL prediction result of the proposed method is significantly better than that of the SAE+LSTM stacking network. Compared with the CNN+LSTM stacking network, the RMSE value is slightly higher. The reason is that there are enough training samples in the dataset, and the deep degradation features obtained from supervised learning can better reflect degradation process of the rolling bearing. However, the proposed method can obtain similar RUL prediction results with supervised learning by artificially setting a small number of labeled training samples. Therefore, the proposed approach in this paper is effective.

4.2. Experimental Results on Real-World Cases

4.2.1. Case Description

The pickling line five-stand unit is the crucial production equipment in the Second Silicon Steel Plant of Wuhan Iron and Steel (Group) Company, China. Due to congenital equipment defects, the failure of burning bearings often occur, and the unit was forced to shut down for unplanned maintenance, which may seriously affect safe production. In response to this urgent problem, the regular on-site maintenance was conducted for precision testing of the gearbox, collecting vibration data and applying the proposed approach to predict the RUL of bearings. As a consequence, the bearings were replaced before they entered the rapid failure period, and the predictive maintenance was achieved, which effectively avoided the occurrence of accidents. The structure of the gearbox is shown in Figure 16 and the layout of the vibration measuring points is shown in Figure 17. In the past, burning bearing accidents had occurred at the position of measuring point 1, which was the focus of attention.

Figure 16. The structure of the gearbox (1, 2, 3, 4 represents 4 bearings respectively).
The bearing is in the stable degradation period.

4.2.2. Vibration Signal Test Process and RUL Prediction

The vibration signal testing process and RUL prediction results are presented in Table 12. According the research object, we set the bearing that exceeds 97% of its whole life cycle as entering the rapid failure period, take the degradation features from 50% to 97% of its whole life cycle as the dataset and degradation percentage as the RUL output label. In order to verify the effectiveness of the proposed approach, the disassembled offline bearings’ health-state judgment was adopted for comparison. It can be seen from the offline bearing that the eccentric sleeve was severely worn, the inner ring gap was large, the outer ring had impact marks, and the inner ring was severely worn. The bearing could be burnt or broken at any time, which was in good agreement with RUL prediction results. The example diagrams of the offline bearing are shown in Figure 18.

Table 12. The vibration test process and RUL prediction results.

| Test Time       | Time–Domain Waveform of Measuring Point 1 | Spectrogram of Measuring Point 1 | RUL Prediction                                      |
|-----------------|------------------------------------------|---------------------------------|-----------------------------------------------------|
| 1 June 2021     | ![Time–Domain Waveform of Measuring Point 1](image1.png) | ![Spectrogram of Measuring Point 1](image2.png) | The bearing is in the stable degradation period       |
| 18 February 2022| ![Time–Domain Waveform of Measuring Point 1](image3.png) | ![Spectrogram of Measuring Point 1](image4.png) | The bearing is in the rapid degradation period (85% of the whole life cycle, RUL = 12%) |
| 23 May 2022     | ![Time–Domain Waveform of Measuring Point 1](image5.png) | ![Spectrogram of Measuring Point 1](image6.png) | The bearing is in the rapid failure period. (RUL < 5%) |

Figure 17. The layout of the vibration measuring points.
5. Conclusions

This paper innovatively proposed a cotraining based semisupervised approach for RUL prediction of bearings. In this approach, a CNN and an LSTM are first cotrained on large quantities of unlabeled data by adding unlabeled samples with high confidence to each other’s training set to obtain the health indicator (HI), then the monitoring data are input into HI and RUL prediction is realized. The RMSE and MAPE value are used and the IEEE PHM 2012 dataset is adopted to realize the improvement of the proposed approach. The results show that the proposed approach has confirmed the validity of the cotraining-based approach in RUL prediction of bearings. However, due to the limitations of network selection and parameter settings, the approach still has some room for improvement. Further research can be carried on the following:

- In this paper, a CNN and an LSTM were selected as the initial network for cotraining. Further work can be carried out based on combing different networks for cotraining, comparing the RUL prediction results of different combinations, and exploring the superiority and general rules of different combinations in various application scenarios.
- In this paper, only two networks were used for cotraining. Further work can be carried out based on increasing the number of cotraining networks to cotraining multiple networks, and integrating multiple prediction results.

Author Contributions: Conceptualization, X.Y. and L.W.; methodology, X.Y.; software, Z.Z.; validation, X.Y., X.X. and L.W.; formal analysis, X.X.; investigation, L.W.; resources, X.Y.; data curation, Z.Z.; writing—original draft preparation, X.Y.; writing—review and editing, Z.Z.; visualization, X.X.; supervision, L.W.; project administration, Z.Z.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant 52205537).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mech. Syst. Signal Process. 2014, 42, 314–334. [CrossRef]
2. You, M.Y.; Meng, G. Updated proportional hazards model for equipment residual life prediction. Int. J. Qual. Reliab. Manag. 2011, 28, 781–795. [CrossRef]
3. Si, X.S.; Wang, W.; Chen, M.Y.; Hu, C.H.; Zhou, D.H. A degradation path-dependent approach for remaining useful life estimation with an exact and closed-form solution. Eur. J. Oper. Res. 2013, 226, 53–66. [CrossRef]
4. Hu, C.; Youn, B.D.; Wang, P.; Yoon, J.T. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliab. Eng. Syst. Saf. 2012, 103, 120–135. [CrossRef]
5. Wang, H.F. Prognostics and Health Management for Complex system Based on Fusion of Model-based approach and Data-driven approach. Phys. Procedia 2012, 24, 828–831. [CrossRef]
6. Diana, B.B.; Victor, G.T.G.; Mario, G.B.; Jorge, L.R. An adaptive ARX model to estimate the RUL of aluminum plates based on its crack growth. Mech. Syst. Signal Process. 2017, 82, 519–536. [CrossRef]
7. Wen, L.; Dong, Y.; Gao, L. A new ensemble residual convolutional neural network for remaining useful life estimation. *Math. Biosci. Eng.* 2019, 16, 862–880. [CrossRef] [PubMed]

8. Bai, G.; Wang, P. Prognostics Using an Adaptive Self-Cognizant Dynamic System Approach. *IEEE Trans. Reliab.* 2016, 65, 1427–1437. [CrossRef]

9. Barbieri, M.; Nguyen, K.T.P.; Diversi, R.; Medjahe, K.; Tilli, A. RUL prediction for automatic machines: A mixed edge-cloud solution based on model-of-signals and particle filtering techniques. *J. Intell. Manuf.* 2021, 32, 1421–1440. [CrossRef]

10. Yin, S.; Xie, X.; Sun, W. A nonlinear process monitoring approach with locally weighted learning of available data. *IEEE Trans. Ind. Electron.* 2017, 64, 1507–1516. [CrossRef]

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

12. Hu, C.H.; Pei, H.; Si, X.S.; Du, D.B.; Pang, Z.N.; Wang, X. A prognostic model based on DBN and diffusion process for degrading bearing. *IEEE Trans. Ind. Electron.* 2020, 67, 8767–8777. [CrossRef]

13. 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]

14. He, R.; Dai, Y.; Lu, J.; Mou, C. Developing ladder network for intelligent evaluation system: Case of remaining useful life prediction for centrifugal pumps. *Reliab. Eng. Syst. Saf.* 2018, 180, 385–393. [CrossRef]

15. Larsen, S.; Hooper, P.A. Deep semi-supervised learning of dynamics for anomaly detection in laser powder bed fusion. *J. Intell. Manuf.* 2022, 33, 457–471. [CrossRef]

16. Zhang, Y.; Wen, J.; Wang, X.; Jiang, Z. Semi-supervised learning combining co-training with active learning. *Expert Syst. Appl.* 2014, 41, 2372–2378. [CrossRef]

17. Hady, M.; Schwenger, F.; Palm, G. Semi-supervised learning for tree-structured ensembles of RBF networks with Co-Training. *Neural Netw.* 2010, 23, 497–509. [CrossRef]

18. Kuo, H.K.J.; Lee, C.H. Discriminative training of natural language call routers. *IEEE Trans. Speech Audio Process.* 2003, 11, 24–35. [CrossRef]

19. Lu, J.; Song, L. Multi-Agent Multi-View Collaborative Perception Based on Semi-Supervised Online Evolutive Learning. *Sensors* 2022, 22, 6893. [CrossRef] [PubMed]

20. Frigieri, E.P.; Ynoguti, C.A.; Paiva, A.P. Correlation analysis among audible sound emissions and machining parameters in hardened steel turning. *J. Intell. Manuf.* 2017, 30, 1753–1764. [CrossRef]

21. Xu, Q.; Lu, S.; Jia, W.; Jiang, C. Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning. *J. Intell. Manuf.* 2020, 31, 1467–1481. [CrossRef]

22. Kang, D.; Ming, X.; Xiaofei, Z. Phase difference correction method for phase and frequency in spectral analysis. *Mech. Syst. Signal Process.* 2000, 14, 835–843. [CrossRef]

23. Siegel, D.; Zhao, W.; Lapira, E.; AbuAli, M.; Lee, J. A comparative study on vibration-based condition monitoring algorithms for wind turbine drive trains. *Wind. Energy* 2014, 17, 695–714. [CrossRef]

24. Peng, E.B.; Zhang, H.G.; Wang, H.Y. Mechanics and Pretreatment Research of Vibration Signals for the Numerical Control Machine. *Adv. Mater. Res.* 2013, 676, 162–165. [CrossRef]

25. Wu, W.; Huang, Y.; Chen, W. High order spectrum’s coupling performance based on mechanical fault diagnosis. *Meas. Diagn.* 2012, 32, 130–134. [CrossRef]

26. Liu, H.; Zhou, J.; Xu, Y.; Zheng, Y.; Peng, X.; Jiang, W. Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing* 2018, 315, 412–424. [CrossRef]

27. Fumee, E.; Oneto, L.; Anguita, D. Condition based maintenance in railway transportation systems based on big data streaming analysis. *Procedia Comput. Sci.* 2015, 33, 437–446. [CrossRef]

28. Yuan, G.; Hao, J.; Zheng, X.; Zhang, J. Tool Life Prediction in Titanium High Speed Milling Processes Based on CNC Real Time Monitoring Data Driven. *China Mech. Eng.* 2018, 29, 457–462, 470. [CrossRef]

29. Yao, F.; He, W.; Wu, Y.; Ding, F.; Meng, D. Remaining useful life prediction of lithium-ion batteries using a hybrid model. *Energy* 2022, 248, 123622. [CrossRef]
36. Babu, G.S.; Zhao, P.; Li, X.L. Deep convolutional neural network based regression approach for estimation of remaining useful life. *Database Syst. Adv. Appl.* **2016**, *2016*, 214–228. [CrossRef]

37. Wu, J.Y.; Wu, M.; Chen, Z.; Li, X.L.; Yan, R. Degradation-Aware Remaining Useful Life Prediction with LSTM Autoencoder. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]

38. Mao, W.; He, J.; Tang, J.; Li, Y. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Adv. Mech. Eng.* **2018**, *10*, 1–18. [CrossRef]

39. Gómez, J.L.; Villalonga, G.; López, A.M. Co-Training for Deep Object Detection: Comparing Single-Modal and Multi-Modal Approaches. *Sensors* **2021**, *21*, 3185. [CrossRef]

40. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An Experimental Platform for Bearings Accelerated Life Test. In Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012.

41. Cao, N.; Jiang, Z.; Gao, J. Automatic Diagnosis Method of Rolling Bearing Based on LSTM-SAE Network. *Adv. Asset Manag. Cond. Monit.* **2020**, *166*, 647–659. [CrossRef]