Meta-Learning Based Early Fault Detection for Rolling Bearings via Few-Shot Anomaly Detection

Wenbin Song, Di Wu, Member, IEEE, Weiming Shen, Fellow, IEEE, Benoit Boulet, Senior Member, IEEE

Abstract—Early fault detection (EFD) of rolling bearings can recognize slight deviation of the health states and contribute to the stability of mechanical systems. In practice, very limited target bearing data are available to conduct EFD, which makes it hard to adapt to the EFD task of new bearings. To address this problem, many transfer learning based EFD methods utilize historical data to learn transferable domain knowledge and conduct early fault detection on new target bearings. However, most existing methods only consider the distribution drift across different working conditions but ignore the difference between bearings under the same working condition, which is called Unit-to-Unit Variability (UtUV). The setting of EFD with limited target data considering UtUV can be formulated as a Few-shot Anomaly Detection task. Therefore, this paper proposes a novel EFD method based on meta-learning considering UtUV. The proposed method can learn a generic metric based on Relation Network (RN) to measure the similarity between normal data and the new arrival target bearing data. Besides, the proposed method utilizes a health state embedding strategy to decrease false alarms. The performance of proposed method is tested on two bearing datasets. The results show that the proposed method can detect incipient faults earlier than the baselines with lower false alarms.

Index Terms—Relation network, Few-shot anomaly detection, Meta-learning, Early fault detection

I. INTRODUCTION

Rolling bearings are one of the most commonly used components in mechanical systems. The fault of rolling bearings is the main cause of mechanical equipment failure [1]. Early fault detection (EFD) can identify the incipient fault symptoms of rolling bearings, which enables predictive maintenance to be taken before severe failure occurring. Therefore, conducting EFD for rolling bearings can prevent mechanical systems break down and economic loss caused by the failure of bearings [2].

EFD aims at assessing the health states of sequential data and monitoring small changes of states in time series data. The health states of bearings remain almost unchanged during the normal operating stage and begin to degrade after an early fault appears [3]. EFD can be formulated as an anomaly detection problem, which is different from fault diagnosis tasks. This is mainly because the purpose of EFD is identifying the abnormal states of bearings. Different faults are all equally treated as anomalies in EFD tasks. A straightforward solution of EFD is constructing a metric to measure the deviation of new arrival data and comparing with normal data to detect early faults. Lu et al. [4] detected early faults by measuring the deviation values between generative sequences and current signal states with a deep architecture based on Deep Neural Network (DNN) and Long Short-Term Memory Network (LSTM). Liu et al. [5] calculated the deviations between the predicted values by AutoRegressive Integrated Moving Average (ARIMA) and the deep features by stacked denoising autoencoder (SDAE) to alarm early faults. However, most of these methods require a certain amount of normal data of the target bearing, such as 300 or 500, to extract sensitive features and develop a robust EFD model. The excessive amount of the required target bearing data to start the model can make it difficult to adapt quickly to EFD tasks on new bearings.

In practice, historical whole lifecycle data can be collected on the same equipment with the target bearing, which can be used to learn the knowledge of pattern recognition of faults. However, the degradation processes of bearings caused by different faults are distinct. This difference along with the different operating conditions leads to the domain drift problem. The states of bearings are gradually changing, which is hard to learn a generic boundary between normal states and faults for EFD under different conditions [6]. To address the problem of limited target bearing data, many research efforts have been concentrated on transfer learning based EFD methods. Transfer learning can learn the pattern recognition knowledge in a source domain and transfer it to tasks on a target domain to improve the performance [7] [8] [9]. Mao et al. [10] utilized a transfer learning method to obtain a feature extractor. The features of normal data extracted by the feature extractor are used to train a support vector data description (SVDD) model for EFD. Mao et al. [11] proposed a deep domain adaptation method with joint adversarial training to detect early faults of bearings. However, most of these methods assume that data collected from different bearings under the same working condition follow the same data distribution. These methods only consider the distribution drift problem across different working conditions, but ignore the unit-to-unit variability (UtUV) across different bearings.

The UtUV refers to degradation trends of bearings differing from each other due to the varying working conditions and health states [12]. In consideration of UtUV and limited avail-
able target bearing anomaly data, the EFD problem can be formulated as a few-shot anomaly detection (FSAD) task. Meta-learning is a typical few-shot learning method which can adapt to new tasks faster than otherwise by the meta-knowledge extracted across tasks generated from the provided data through a meta-learner [13][14]. There have been a lot of research efforts on meta-learning based fault diagnosis. Hu et al. [15] proposed a Task-Sequencing Meta-Learning based fault diagnosis method to detect faults with few target data available. Wu et al. [16] utilized Relation Network (RN) for few-shot fault diagnosis under two situations, including transferring across different working conditions and from artificial fault to natural failure. Wang et al. [17] proposed a fault comparison deep neural network (FCDNN) based on Siamese Net to conduct few-shot fault diagnosis. However, most existing methods require few normal data and fault data of target domain to fine tune the model or serve as the support set. In the setting of FSAD considering UtUV, the fault data of target bearing is unavailable because the final failure type of the target bearing is unknown and cannot be predicted. Therefore, this paper proposes an meta-learning based EFD method considering UtUV, which can detect early faults with limited normal data of target bearing. A Health State Assessment model based on Unsupervised Sequence Segmentation Convolutional Neural Network (USSCNN) [6] is utilized to identify the states of historical whole lifecycle data collected from bearings other than the target bearing. The labeled historical data are further used to conduct meta-learning based on RN. The final prediction of the health states of the new arrival data are obtained by a Health State Embedding Strategy (HSES) based on the RN. The main contribution of our work can be summarized as follows:

1. Different from existing anomaly detection based EFD methods, the proposed method can detect early faults with limited target bearing normal data available. The proposed RN based method can learn a transferable metric to measure the similarity of the current sample to normal samples.
2. Unlike most existing methods, the proposed method considers the distribution drift across working conditions and different units under the same working condition. The proposed method utilized multiple historical data to train an RN, which can adapt to the EFD task of new bearings quickly.
3. In order to decrease false alarms, a Health State Embedding Strategy is proposed in this paper. The support sets consisted of normal data are dynamically enlarged by the new arrival data identified as normal. The final states of data are determined by integrating the results obtained by multiple support sets.

The rest of this paper is organized as follows. Section II introduces the theoretical background of related approaches. Section III illustrates the details of the proposed EFD model. Section IV presents the experiment details and results. Section V concludes the paper and discusses some future work.
feature module is used to extract feature maps of support set and query set with $M$ channels. After that, each feature map of the query set is channel-wise concatenated with the feature maps of individual classes in the support set and generates $C$ concatenated feature maps with $2 \times M$ channels. The relation scores of concatenated feature maps can be calculated by the relation module. The relation scores between the samples in the query set and the support set can be converted to a one-hot label, which indicates the class of the sample.

III. PROPOSED METHOD

A. Problem Definition

The early fault detection problem with limited normal data available can be treated as a task of few-shot anomaly detection (FSAD). In this situation, a small number of $K$ normal samples of the target bearing are available and serve as the support set. The fault data of the target bearing is unavailable, which is different from the setting of typical few-shot tasks. In practice, the fault type and degradation trend of the target bearing are unpredictable and unknown at the beginning of normal operating. Therefore, we have to learn a metric to measure the states deviation of the target bearing based on the support set.

However, it is tough to learn a generic and effective metric only relying on the support set with limited samples. Therefore, a training set with several historical whole lifecycle data will be utilized to perform meta-learning. The training data will contribute to extract the transferable latent relationship between the normal samples themselves and between normal samples and faulty samples. Let $S^1, \ldots, S^m$ denote the multiple historical whole lifecycle data, where $m$ is the number of historical data. Generally, a snapshot of signals will be collected at the same interval, and the historical bearing data consist of multiple snapshots from intact to complete failure. Let $S^i = \{s^i_1, \ldots, s^i_j, \ldots, s^i_n\}$ represent the $i$th training data, where $s^i_j$ represents the snapshot collected at time $t_j$, $n$ is the number of snapshots of the $i$th training data. The number of snapshots differs from different training data.

There are only two states of snapshots, i.e. normal and faulty, but the labels of historical data are unknown. In order to conduct meta-learning on training set, it is necessary to assess the states of the historical data.

B. Framework of the Proposed Method

The framework of the proposed method is shown in Fig. 2. There are two main steps in the proposed method, i.e. training process and predicting process. The historical whole lifecycle data are only used in the training process. In the training process, the states of snapshots of individual historical whole lifecycle data are identified by the USSCNN-based Health State Assessment (HSA) model. Although the label space of the training set and target bearing is the same, the data distributions of training set and target bearing vary from each other due to different working conditions and fault types. Similarly, the normal data and faulty samples in different historical data cannot be treated as samples in the same classes. This property fits to the setting of RN, and the labelled historical data can be used to train the RN. The details of training process is presented in Section III-D.

In the predicting process, the well-trained RN is utilized to calculate the relation scores between the support set and the new arrival data batch of the target bearing. In order to improve the robustness of the proposed method, a health state embedding strategy (HSES) is utilized in this stage. The health states of the new arrival data will be obtained by combining the results of relation scores between multiple support sets and the new data.

C. Health State Assessment Model

The first step of the training process is assessing the health states of the historical data. As mentioned in Section II-A, the historical data can be divided to two classes by the time when an early fault appears. Let $S = \{s_1, \ldots, s_j, \ldots, s_n\}$ represent a set of whole lifecycle data, where $n$ is the number of snapshots of the historical data. The objective is to find the time $t_e$ when the early fault appears. If $j < e$, the states of $s_j$ are normal, otherwise, the states are faulty. In this paper, the
Algorithm 1 HSA: Health State Assessment Model

Input: Snapshots of the historical data $s_1, \cdots, s_n$, Number of iteration $N$
1: Conduct FFT to extract frequency domain features and reshape 1D features to 2D images $x_1, \cdots, x_n$.
2: Randomly select the early fault start time $e^* = e$, initialize best fitness value $f^* = 0$.
3: for $i = 1, \cdots, N$ do
4: Define pseudo labels by Eq. (1).
5: Train a CNN model $C$ with $\{x_j, \hat{y}_j\}_{j=1}^n$.
6: Calculate the fitness value $f(e)$ by Eq. (2).
7: if $f(e) < f^*$ then
8: Update $e^* = e$, $C^* = C$, $f^* = f(e)$
9: end if
10: Update $e$ based on SA
11: end for
Output: Early fault start time $e^*$, CNN model $C^*$.

Algorithm 2 RN-FSAD: Relation Network Based Few-shot Anomaly Detection

Input: Multiple historical data $S^1, \cdots, S^m$, Support size $K$, Query size $T$
1: for $i = 1, \cdots, m$ do
2: for $e = 1, \cdots, Epoch$ do
3: Randomly select $K$ successive normal samples from $S^i$ as support set $S$.
4: Randomly select $T$ normal samples other than $S$ and faulty samples from $S^i$ as query set $Q$.
5: Calculate relation scores by Eq. (4).
6: Calculate loss by Eq. (5).
7: Update $f_0 \leftarrow \theta, g_\varphi \leftarrow \varphi$
8: end for
9: end for
Output: Feature extraction module $f_0$, relation module $g_\varphi$.

HSA model based on USSCNN is conducted to search for the time $t_e$ of individual historical data. The pseudo code of the HSA model is shown in Algorithm 1.

The raw historical signals are processed by Fast Fourier Transform (FFT) to extract frequency domain features. Subsequently, the 1D frequency features of $s_j$ are reshaped to a 2D matrix and further converted to a grey scale image $x_j$. After that, the early fault start time $e^*$ is randomly initialized as $e$ and the pseudo label $\hat{y}_j$ can be aligned to each sample $x_j$ as follows.

$$\hat{y}_j = \begin{cases} 0, & j < e \\ 1, & j \geq e \end{cases} \quad (1)$$

The accuracy of the CNN model trained by samples with pseudo labels $\{x_j, \hat{y}_j\}_{j=1}^n$ and the proportion of the normal samples defined by $e$ are combined as the fitness function as follow.

$$f(e) = \alpha \times \frac{e-1}{n} - ACC(e) \quad (2)$$

where $ACC(e)$ is the training accuracy of CNN. $e$ is the defined early fault start time. $n$ is the total number of snapshots. $\alpha$ is the weighting parameter. In this paper, $\alpha$ is set to 0.002 to avoid the algorithm seeking for extremely small $e$ with a relatively low accuracy. The SA is utilized to optimize the fitness function and find the early fault appearance moment $t_{e^*}$. The optimal $e^*$ is the earliest time that can classify the states of samples defined by $e^*$ accurately.

The states of historical data defined by $t_e$ should be identified accurately and easily because of the differences of the intrinsic features. Therefore, the iteration number of the CNN training process in Line 11 of Algorithm 1 should not be large. In this paper, the iteration number is set to be in the range of 10 to 30.

D. EFD Based on Relation Network

1) RN Based Few-Shot Anomaly Detection: The early fault detection with limited normal data can be formulated as a task of few-shot anomaly detection. There are two main differences in the training process between the RN Based FSAD (RN-FSAD) method and traditional RN. One is that there is only one type of data in the support set since the faulty samples of target bearings are not available in the beginning of operation. The other is that the support set in the RN-based EFD method should be $K$ successive samples other than randomly selected samples in the normal data. Because new data arrive successively in online EFD and new identified normal samples will be added to the support sets in the health state embedding strategy. The training process should simulate the sampling strategy of the support set. The details will be illustrated in Section II-D2.

There are two modules in the RN, including a feature embedding module $f_0$ and a relation module $g_\varphi$. Algorithm 2 presents the pseudo code of RN-FSAD. The training process conducts episode based training [22] to simulate the FSAD setting. In each training epoch, a subset of $K$ successive samples $s_p^i, \cdots, s_{p+K-1}^i$ are selected from the historical data to serve as the support set, where $1 \leq p \leq e - K$. Meanwhile, $T$ normal samples other than the support set and $T$ faulty samples are randomly selected as the query set. The selecting of the support set and query set is to mimic the situation of online EFD.

The deep features of the support set and query set are extracted by the feature embedding module $f_0$.

$$f_j^i = f_0(s_j^i) \quad (3)$$

where $j = s, \cdots, s + K - 1$. The feature maps of the query set are denoted as $f_0(q_k)$, where $q_k$ represents the sample in the query set, $k = 1, \cdots, 2M$. The feature map of the support set $f_S$ is integrated by element-wise sum of all feature maps in the support set. The relation scores of the support set to the samples of query set can be calculated as follow.

$$r_k = g_\varphi(C(f_S, f_0(q_k))) \quad (4)$$

where $C(\cdot, \cdot)$ is the channel-wise concatenation operation. The loss function can be defined as follow.

$$L = \frac{1}{2M} \sum_{k=1}^{2M} (r_k - y_k)^2 \quad (5)$$
The relation scores between the kth sample $x^*_k$ and all support sets can be represented as a vector $[r^*_1, \cdots, r^*_n]$. The state of $x^*_k$ identified by each support set is determined through comparing the relation score with a threshold $r_t$. If $r^*_j \geq r_t$, the state identified by the jth support set is normal, otherwise, the state is abnormal. The final decision of the state will be given by voting strategy.

3) Robust Early Fault Alarm Criterion: The early fault can be detected based on the health states obtained by the proposed method. However, alarming early fault once an abnormal sample appears will lead to a high false alarm rate. Abnormal states may appear in the normal data due to the influence of noise and imperfection of the prediction model. Therefore, a commonly used robust strategy \cite{23} \cite{24} for early fault alarm is conducted in this paper.

Let $X = \{x_1, \cdots, x_K\}$ represent the new arrival batch samples and $y_1, \cdots, y_K$ represent the corresponding labels. If $x_i$ is normal, $y_i = 0$, otherwise, $y_i = 1$. If the number of detected abnormal states in the batch samples exceeds a threshold $p$, i.e. $\sum_{k=1}^{K} y_i \geq p$, an early fault is considered to have occurred.

IV. Experiments

Following the practice of existing work \cite{3} \cite{25} \cite{26}, two frequently used run-to-failure bearing datasets are utilized to validate the performance of the proposed method, including IEEE PHM Challenge 2012 dataset \cite{27} and XJTU-SY dataset \cite{28}.

A. Datasets

1) PHM2012: The IEEE PHM Challenge 2012 dataset was collected by PRONOSTIA test platform with a rotating part, a loading part, and a data collection part. The bearing is driven by a motor through the shaft in the rotating part. Accelerated bearing degradation experiments of bearings can be conducted by applying an extra pressure on the bearing through the loading part. The snapshots of signals are collected
by accelerometer sensors with a sampling frequency of 25.6 KHz in horizontal and vertical directions.

There are several run-to-failure bearing data under three working conditions in this dataset. In this paper, the target bearings are the 1st bearing and the 3rd bearing with the speed of 1800 rpm and a load of 4000 N, and they are abbreviated as PHM1_1 and PHM1_3 respectively. The rest of the whole lifecycle data are used as training set.

2) XJD: The XJD dataset conducted run-to-failure experiments on 15 rolling bearings under three different working conditions. The signals are collected with a sampling frequency of 25.6 kHz. In this paper, the 5th bearing under the first working condition is chosen for testing, which is denoted as XJD1_5 for short. The training data contain XJD1_1-1_3, 1_5 and XJD_2-1_2_5.

### TABLE I

**The Structure of Feature Embedding Module and Relation Module**

| Module               | Layer                  | Parameters | Output Size |
|----------------------|------------------------|------------|-------------|
| Feature Embedding    | Input                  | /          | 1×32×32     |
| Module               | Convolution Layer (ReLU) | 5×5        | 6×28×28     |
| Module               | Max Pool Layer 1       | 2×2        | 6×14×14     |
| Module               | Convolution Layer 2 (ReLU) | 5×5        | 16×10×10    |
| Module               | Max Pool Layer 2       | 2×2        | 16×5×5      |
| Relation Module      | Convolution Layer (ReLU) | 3×3        | 32×3×3      |
| Module               | Max Pool Layer 1       | 2×2        | 32×1×1      |
| Module               | Fully Connected Layer 1| 16          | 16          |
| Module               | Fully Connected Layer 2| 1           | 1           |

### B. Experiment Setting

1) Structure of RN: The structures of the feature embedding module and the relation module in the experiments are shown in Table I. The feature embedding module contains two convolution layers and outputs feature maps with size of 16×5×5. The feature maps of support set and query set are channel-wise concatenated to generate images with size of 32×5×5. The relation module consists of one convolution layer and two fully connected layers. The relation scores of query set to support set calculated by the relation module is in the range of zero to one.

2) Evaluation Metrics: In order to evaluate the performance of early fault detection, two commonly used evaluation metrics [4][10][23] and the fitness score are adopted in the experiments:

(1) the location (LO) of the sample where an early fault appears;
(2) the number of false alarms (FA). False alarms refer to normal states identified as faults by the EFD method before the early fault occurs.
(3) the fitness score is defined as follow.

\[ Score = w_1 \times LO + w_2 \times FA \]  \hspace{1cm} (7)

where \( w_1 \) and \( w_2 \) are weight parameters, and \( w_1 + w_2 = 1 \). In real application, decreasing false alarm is more significant than detecting early fault early. Therefore, we set \( w_1 = 0.4 \) and \( w_2 = 0.6 \) in this paper.

3) Baseline: To verify the effectiveness of the proposed method, seven baseline methods are introduced as follows.

1. Robust early fault detection based on deep transfer learning (REFD+DTL) [24]
2. Self-Adaptive Deep Feature Matching (SDFM) [29]
3. Online Early Fault Detection Based on Deep Transfer Learning (OEFD+DTL) [10]
4. Safe Semi-Supervised SVM (S4VM) + second-order difference of radius margin bound (SODRMB) [23]
5. ARIMA+Domain Adaptation (DA) with the same working condition [3]
6. ARIMA+Domain Adaptation (DA) under different working conditions [3]
7. RN-FSAD without HSES
8. RN-FSAD

Methods 1 [24] utilizes a deep transfer learning method to conduct robust early fault detection. Methods 2 [29] assumes that the distributions of bearings under the same working condition are the same. The historical data are used train an SDAE to extract deep features. The EFD is conducted by a Self-Adaptive Deep Feature Matching (SDFM) in the online stage. Method 3 [10] utilizes deep transfer learning through fine tuning the feature extraction module of VGG16 with auxiliary bearing data. The normal data of auxiliary bearing data are used to train a SVDD model to detect early faults of the target bearing. Method 4 [23] is an early fault detection method based on S4VM and utilizes the SODRMB of SVM as the health index. Method 5 and Method 6 [5] extract common features of the target bearing and auxiliary bearings based on domain adaptation and detect early faults based on ARIMA. Method 5 and 6 conduct domain adaptation across data from the same working condition and different working conditions respectively.

Since the FSAD based early fault detection problem has not been well-studied, we conducted Method 4, 5 and 6 with the same experiment setting as the proposed method, i.e. only 10 target bearing data are required. It should be noted that Method 1, 2, 3 assume that the data from bearings under the same working condition follow the same distribution. These methods utilize historical data to develop EFD models and need no target bearing data. Apart from the methods mentioned above, we also test the proposed method without HSES to validate the effectiveness of the proposed embedding strategy.

4) Health State Assessment Results of the Training Set: The health state assessment results of PHM2012 and XJD datasets are listed in Table II. The results consist of the early fault location of each bearing and the percentage of normal states, which indicate that the degradation trends vary from bearings due to the different working conditions and fault types, especially in XJD dataset. The percentages of normal states bearings under the first and the second working condition differ from 20% to 65% and from 28% to 90% respectively. The failure of outer race is occurred on XJD1_1-1_3, 2_2, 2_4 and 2_5, but the percentage of normal states is different as well.

5) EFD Results of PHM2012: In the experiments of PHM1_1 and 1_3, the parameters \( K \) and \( p \) for early fault alarm are set to be 10 and 0.8 respectively. The thresholds \( r_t \)
TABLE II
THE HEALTH STATE ASSESSMENT RESULTS OF PHM2012

| Dataset | Bearing | Early Fault Location | Total Num of Samples | Percentage of Normal States |
|---------|---------|----------------------|----------------------|----------------------------|
| PHM     | PHM1_1  | 1261                 | 2801                 | 45.02%                     |
|         | PHM1_2  | 650                  | 871                  | 74.63%                     |
|         | PHM1_3  | 1274                 | 2375                 | 53.64%                     |
|         | PHM2_1  | 120                  | 911                  | 13.17%                     |
|         | PHM2_2  | 160                  | 797                  | 20.08%                     |
|         | PHM2_3  | 490                  | 313                  | 91.13%                     |
|         | XJD1_1  | 931                  | 1476                 | 63.08%                     |
|         | XJD1_2  | 378                  | 1932                 | 19.57%                     |
|         | XJD1_3  | 697                  | 1896                 | 36.76%                     |
|         | XJD1_5  | 409                  | 624                  | 65.54%                     |
| XJD     | XJD2_1  | 532                  | 5892                 | 90.33%                     |
|         | XJD2_2  | 160                  | 797                  | 20.08%                     |
|         | XJD2_3  | 362                  | 6396                 | 56.63%                     |
|         | XJD2_4  | 361                  | 304                  | 71.63%                     |
|         | XJD2_5  | 1441                 | 4068                 | 35.42%                     |

TABLE III
COMPARING RESULTS OF THREE BEARINGS

| Method               | PHM1_1 | PHM1_3 | XJD1_5 |
|----------------------|--------|--------|--------|
| 1. REFD+DTL          | 1400   | 0      | 2246   |
| 2. SDFM              | 1374   | 42     | 1193   |
| 3. OEF+DTL           | 1236   | 81     | 1667   |
| 4. S4VM+SOBMB        | 1531   | 11     | 2243   |
| 5. ARIMA+DA (same)   | 2454   | 876    | 2094   |
| 6. ARIMA+DA (different) | 741   | 694    | 2249   |
| 7. RN-FSAD without HSES | 1374 | 10    | 1275   |
| 8. RN-FSAD           | 1397   | 0      | 409    |

respectively with no false alarm. The early fault detection results of PHM1_1 is shown in Fig. 4(a). Although there are many samples before the early fault appearing identified as abnormal, the robust early fault alarm criterion ensures that those misleading labels will not alarm early fault. After the early fault occurring, the states of samples are all identified as abnormal, which means that the fault samples are easier to be classified and there is a high risk for false alarms. The distributions of original data and features extracted by feature embedding module of PHM1_1 are shown in Fig. 4(b) and Fig. 4(c). Comparing Fig. 4(b) with Fig. 4(c), the features extracted by the RN are easier to distinguish between normal and abnormal data in the feature space.

6) EFD Results of XJD: The parameters $K$ and $p$ for early fault alarm are set to be 10 and 0.6 in the experiments of XJD1_5. The threshold $r_t$ to determine the states of the new data is set to be 0.8 and the volume of the support set is set to be 10. The EFD result of XJD1_5 is the 1085th sample with no false alarm. The same as the PHM1_1 and 1_3, the robust early fault criterion contributes to decrease the false alarms of the proposed method.

7) Comparative Results: The comparative results of LO and FA are listed in Table III. Among three experiments, the proposed method generates the least false alarms comparing with other methods. The detected early fault locations of PHM1_1 and PHM1_3 by Method 2 are earlier than the proposed method, but it has much more false alarms than the proposed method. Similarly, Method 3 and Method 6 detect the incipient fault earlier on PHM1_1, but generate more false alarms than the proposed method. The results of the proposed method are better than Method 1-6 on the rest bearings both in the location of early faults and the false alarm number. Comparing with Method 7, the proposed method detects the incipient fault earlier with less false alarms on XJD1_5. Despite Method 7 detects the location of early faults earlier than the proposed method on PHM1_1 and 1_3, it has much more false alarms than the proposed method. The results of LO and FA prove that the embedding strategy can decrease false alarms effectively and improve the robustness of the proposed EFD method.
The results of fitness scores are shown in Fig. 5. As for the fitness scores, the proposed method achieved the lowest scores on PHM1_3 and XJD1_5, and the third lowest score on PHM1_1. More importantly, the average fitness scores of the proposed method with and without HSES are the lowest and second lowest among the baselines, which proves the effectiveness of the proposed few-shot anomaly detection architecture for EFD.

V. CONCLUSION AND FUTURE WORK
In this paper, an EFD method of rolling bearings based on RN considering UtUV is proposed. The proposed method can learn a generic metric based on RN with historical data and detect early faults with limited normal data of the target bearing. In order to decrease the false alarms of EFD, the proposed method utilizes a health state embedding strategy. The proposed method is tested on three bearings in the PHM2012 and XJD datasets. The results show that the proposed method can detect incipient faults earlier than other methods with the lowest number of false alarms. More importantly, limited target bearing data are required to conduct effective early fault detection by the proposed method. The embedding strategy is also proven to be effective in improving the reliability of the proposed method.

Nevertheless, the proposed method has some limitations. On one hand, it has to set a threshold to determine the states of samples manually. On the other hand, the proposed method cannot transfer the knowledge across different platforms. Therefore, our future research work will be conducted in the following aspects. First, this method will be further developed to learn the threshold automatically. Moreover, an adaptive EFD method will be studied to adjust to different platforms for real industry applications.

REFERENCES
[1] Y. Xu, D. Zhen, J. X. Gu, K. Rabeyee, F. Chu, F. Gu, and A. D. Ball, “Autocorrelated envelopes for early fault detection of rolling bearings,” Mechanical Systems and Signal Processing, vol. 146, p. 106990, 2021.
[2] X. Li, S. Cao, L. Gao, and L. Wen, “A threshold-control generative adversarial network method for intelligent fault diagnosis,” Complex System Modeling and Simulation, vol. 1, no. 1, pp. 55–64, 2021.
[3] S. Yang, Z. Liu, and G. Lu, “Early change detection in dynamical bearing degradation process based on hierarchical graph model and adaptive inputs weighting fusion,” IEEE Transactions on Industrial Informatics, vol. 17, no. 5, pp. 3186–3196, 2020.
[4] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, “Early fault detection approach with deep architectures,” IEEE Transactions on instrumentation and measurement, vol. 67, no. 7, pp. 1679–1689, 2018.
[5] K. Liu, W. Mao, H. Shi, and X. Liang, “A new unsupervised online early fault detection framework of rolling bearings based on granular feature forecasting,” IEEE ACCESS, vol. 9, pp. 159684–159698, 2021.
[6] W. Song, W. Shen, L. Gao, and X. Li, “An early fault detection method of rotating machines based on unsupervised sequence segmentation convolutional neural network,” IEEE Transactions on Instrumentation and Measurement.
[7] C. Zhao, G. Liu, and W. Shen, “A dual-view alignment-based domain adaptation network for fault diagnosis,” Measurement Science and Technology, vol. 32, no. 11, p. 115102, 2021.
[8] C. Zhao, G. Liu, W. Shen, and L. Gao, “A multi-representation-based domain adaptation network for fault diagnosis,” Measurement, p. 109650, 2021.
[9] G. Liu, W. Shen, L. Gao, and A. Kusiak, “Predictive modeling with an adaptive unsupervised broad transfer algorithm,” IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 2, pp. 443–456, 2019.
[10] W. Mao, L. Ding, S. Tian, and X. Liang, “Online detection for bearing incipient fault based on deep transfer learning,” Measurement, vol. 152, p. 107278, 2020.
[11] W. Mao, L. Ding, Y. Liu, S. S. Afshari, and X. Liang, “A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault,” ISA transactions, 2021.
[12] N. Li, Y. Lei, T. Yan, N. Li, and T. Han, “A wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability,” IEEE Transactions on Industrial Electronics, vol. 66, no. 3, pp. 2092–2101, 2018.
[13] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” ACM computing surveys (csur), vol. 53, no. 3, pp. 1–34, 2020.
[14] J. Vanschoren, “Meta-learning: A survey,” arXiv preprint arXiv:1810.03548, 2018.
[15] Y. Hu, R. Liu, X. Li, D. Chen, and Q. Hu, “Task-sequencing meta learning for intelligent few-shot fault diagnosis with limited data,” IEEE Transactions on Industrial Informatics, 2021.
[16] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, “Few-shot transfer learning for intelligent fault diagnosis of machine,” Measurement, vol. 166, p. 108202, 2020.
[17] C. Wang and Z. Xu, “An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis,” Neurocomputing, 2021.
[18] Y. Cheng, H. Zhu, K. Hu, J. Wu, X. Shao, and Y. Wang. “Health degradation monitoring of rolling element bearing by growing self-organizing mapping and clustered support vector machine,” IEEE Access, vol. 7, pp. 135 322–135 331, 2019.
[19] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, “Learning to compare: Relation network for few-shot learning,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1199–1208.
[20] G. Koch, R. Zemel, R. Salakhutdinov et al., “Siamese neural networks for one-shot image recognition,” in ICML deep learning workshop, vol. 2. Lille, 2015.
[21] J. Snell, K. Swersky, and R. S. Zemel, “Prototypical networks for few-shot learning,” arXiv preprint arXiv:1703.05175, 2017.
[22] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra et al., “Matching networks for one shot learning,” Advances in neural information processing systems, vol. 29, pp. 3630–3638, 2016.
[23] W. Mao, S. Tian, J. Fan, X. Liang, and A. Safian, “Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation,” Journal of Manufacturing Systems, vol. 55, pp. 179–198, 2020.
[24] W. Mao, D. Zhang, S. Tian, and J. Tang, “Robust detection of bearing early fault based on deep transfer learning,” Electronics, vol. 9, no. 2, p. 323, 2020.
[25] A. V. Malawade, N. D. Costa, D. Muthirayan, P. P. Khargonekar, and M. A. Al Faruque, “Neuroscience-inspired algorithms for the predictive maintenance of manufacturing systems,” IEEE Transactions on Industrial Informatics, vol. 17, no. 12, pp. 7981–7990, 2021.
[26] M. Xia, T. Li, T. Shu, J. Wan, C. W. De Silva, and Z. Wang, “A two-stage approach for the remaining useful life prediction of bearings using deep neural networks,” IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3703–3711, 2018.
[27] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, “PRONOSTIA: An experimental platform for bearings accelerated degradation tests,” in IEEE International Conference on Prognostics and Health Management, PHM’12. IEEE Catalog Number CFP12PHM-CDR, 2012, pp. 1–8.
[28] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” IEEE Transactions on Reliability, vol. 69, no. 1, pp. 401–412, 2018.
[29] W. Mao, J. Chen, X. Liang, and X. Zhang, “A new online detection approach for rolling bearing incipient fault via self-adaptive deep feature matching,” IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 2, pp. 443–456, 2019.