An Early Fault Detection Method of Rotating Machines Based on Multiple Feature Fusion with Stacking Architecture

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

Abstract—Early fault detection (EFD) of rotating machines is important to decrease the maintenance cost and improve the mechanical system stability. One of the key points of EFD is developing a generic model to extract robust and discriminative features from different equipment for early fault detection. Most existing EFD methods focus on learning fault representation by one type of feature. However, a combination of multiple features can capture a more comprehensive representation of system state. In this paper, we propose an EFD method based on multiple feature fusion with stacking architecture (M2FSA). The proposed method can extract generic and discriminative features to detect early faults by combining time domain (TD), frequency domain (FD), and time–frequency domain (TFD) features. In order to unify the dimensions of the different domain features, Stacked Denoising Autoencoder (SDAE) is utilized to learn deep features in three domains. The architecture of the proposed M2FSA consists of two layers. The first layer contains three base models, whose corresponding inputs are different deep features. The outputs of the first layer are concatenated to generate the input to the second layer, which consists of a meta model. The proposed method is tested on three bearing datasets. The results demonstrate that the proposed method is better than existing methods both in sensibility and reliability.

Index Terms—Ensemble learning, Stacking, One-Class SVM, SDAE, Early fault detection.

I. INTRODUCTION

Rotating machines are commonly used in many fields such as transportation, electrical power generation and manufacturing. Faults of rotating machines can reduce the stability of the entire system and lead to catastrophic accidents [1]. Early fault detection (EFD) aims to identify incipient faults at an early stage, which can be seen as a fundamental step for condition based maintenance (CBM) or predictive maintenance. CBM means that once early fault occurs, maintenance measures will be taken in advance to revert the machine back to normal. Therefore, it is very significant to conduct proper and reliable early fault detection for rotating machines, so as to guarantee the stable operation of the whole system and reduce maintenance costs [2].

An early fault is the state when there are symptoms of appearance of incipient faults [3]. The association between performance degradation and early faults of rotating machines can be discovered using feature information in the whole lifecycle data [4]. One of the key points of EFD is developing a generic model to extract robust and discriminative features from different equipment for early fault detection. Vibration based techniques are widely used for fault detection of rotating machines. It has been proved that vibration analysis is effective to monitor the health state of rotating machines [5]. The representation which can reflect the dynamic health condition precisely can be obtained by vibration feature extraction effectively. Existing vibration feature extraction methods include time-domain (TD) analysis, frequency-domain (FD) analysis, and time–frequency-domain (TFD) analysis [6]. TD features include Root Mean Square (RMS) [7], kurtosis [8], and skewness [9]. TD features are sensitive to fault occurrence, but it is hard to obtain an indication of the early fault location based on TD features [10]. FD features are widely used in vibration based fault diagnostics. A typical extraction method is Fast Fourier Transform (FFT) [11]. TFD feature extraction methods include Short-time Fourier Transform (STFT) [12] and Discrete Wavelet Transform (DWT) [13].

The fault vibration signal at an early stage is usually very weak and hard to be identified [14]. Machine learning methods have been widely applied in the field of EFD detection for its powerful ability of feature representation. Wen et al. [14] proposed an early fault detection method named graph modeled singular values (GMSVs), which constructs the graph by utilizing singular values as inputs. The proposed method enhanced the ability for detecting faults. Mao et al. [15] proposed a semi-supervised architecture to detect early faults. At first, Stacked Denoising Autoencoder (SDAE) is used to extract deep features from Hilbert-Huang Transform (HHT) marginal spectrum of raw signals. Then the semi-supervised support vector machine (S4VM) is utilized to identify the health state of the new arrived data collected from the target bearing. Shao et al. [5] proposed an EFD detection model based on enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy. The proposed method constructs a novel monitoring index to indicate the bearing performance degradation. Mao et al. [16] proposed an online early fault detection approach for
rolling bearings based on Self-adaptive Deep Feature Matching (SDFM). The proposed method assesses health states of training data based on Singular Value Decomposition (SVD) and Kurtosis criterion. Then the detection model based on Support Vector Data Description (SVDD) is applied to identify early faults using deep features obtained by Denoising Autoencoder (DAE). However, most of these methods have to set a threshold manually to judge the state of new arrival data or develop robust model with auxiliary historical data, which make it hard to implement in practical. Besides, most of existing research efforts are mainly focused on using one type of features for EFD. Every single feature can reflect the health condition of rotating machines to some degree, but a combination of multiple features will contribute to construct a more comprehensive representation.

Ensemble learning is a technique to combine multiple learners to obtain better predictive performance than one single learner \[17\] \[18\]. It has achieved great performance in many real applications. Zhang et al. \[19\] proposed a multidimensional feature fusion and stacking ensemble mechanism (MFFSEM) to detect abnormal conditions of network traffic information. Li et al. \[20\] utilized an improved stacking ensemble learning method to detect faults of sensors in heating, ventilation, and air conditioning systems. Liu et al. \[21\] utilized multidimensional feature fusion and ensemble learning to develop a fault diagnosis model for the braking system of heavy-haul trains. These applications show the great ability of ensemble learning in integrating multiple features, which is beneficial to generate a more discriminative representation for EFD. To our best knowledge, ensemble learning based methods have not been well studied in the field of EFD. Therefore, this paper proposes an EFD method for rotating machines based on multiple feature fusion with stacking architecture (M2FSA). The multiple features extracted from raw signals include TD, FD and TFD features. To unify the dimensions of different domain features, three SDAEs with different structures are used to learn three types of deep domain features, three SDAEs with inputs of different deep domain features. The outputs of different structures are concatenated as the input to the meta model of the second layer to obtain a final prediction of the equipment state. In this paper, the One-Class Support Vector Machine (SVM) are selected as the base models and the meta model of the proposed method. The contribution of this paper can be summarized as follows.

1. Unlike existing EFD methods, this paper utilized multiple feature fusion with stacking architecture (M2FSA) to develop an anomaly detection model for early fault detection. The proposed method integrates multiple deep features with a hierarchical ensemble architecture to extract generic features for EFD of different rotating machines.
2. Different from other AD-based EFD methods, the decision boundary can be learned only with normal data of the target rotating machines and no extra historical data is required. Besides, the state of the equipment can be obtained without setting a threshold. These characteristics make the proposed method easy to implement in real applications.

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 ensemble learning-based EFD model. Section IV presents the experiment details and comparative results. Section V concludes the paper and discusses some future work.

II. BACKGROUND

A. One-Class SVM

Support Vector Machine (SVM) can handle classification and regression tasks in a supervised pattern. The training process of SVM requires samples from multiple classes. Different from general SVM, One-Class SVM only needs one type of data, i.e. normal data or positive data, to build a decision boundary, and detect outliers or negative samples of new data. The One-Class SVM aims to find a hyperplane \( \langle w, x \rangle - \rho = 0 \) that separates the data from different classes with maximal margin. This problem can be expressed as follows:

\[
\min_{w, \xi_1, \cdots, \xi_n, \rho} = \frac{1}{2} \|w\|^2 + \frac{1}{en} \sum_{i=1}^{n} \xi_i - \rho
\]

subject to:

\[
\langle w, x \rangle \geq \rho - \xi_i, \quad i = 1, \cdots, n
\]

\[
\xi_i \geq 0, \quad i = 1, \cdots, n
\]

where \( \xi_i \) are slack variables and \( v \) adjusts the penalization of the slack variable in the loss function. Sequential Minimal Optimization (SMO) of the dual form is an efficient way to deal with this problem. The corresponding Wolfe dual is:

\[
\min_{\alpha_1, \cdots, \alpha_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \langle x_i, x_j \rangle
\]

subject to:

\[
0 \leq \alpha_i \leq 1/(vn)
\]

\[
\sum_{i=1}^{n} \alpha_i = 1
\]

where \( \{ \alpha_1, \cdots, \alpha_n \} \) are Lagrange multipliers. The decision function below can be used to judge whether a new sample \( x \) is the same class as the original set or an anomaly.

\[
g(x) = \text{sign}(\langle w, \phi(x) \rangle - \rho)
\]

B. Stacked Denoising Autoencoder

Autoencoder (AE) is an unsupervised method that can extract low-dimensional features of raw data by minimizing the reconstruction loss of the raw data. As shown in Fig. 1(a), an AE contains an encoder and a decoder. A low-dimensional representation of input can be learned by the encoder. The representation learned by encoder is further used to reconstruct the raw input by the decoder. The forward calculation of a given sample \( x \) can be expressed as:

\[
h = \sigma(Wx + b)
\]

\[
z = \sigma'(W'h + b')
\]
where $\sigma$ and $\sigma'$ are activation functions, which provide the nonlinear fitting capability to the network. $W$ and $b$ represent the weight matrix and bias of the encoder, $W'$ and $b'$ are the weight matrix and bias of the decoder. $h$ is the output of the encoder. $z$ is the reconstruction of the raw sample $x$.

The objective of AE network is minimizing the reconstruction error. A Denoising Autoencoder (DAE) has similar structure with AE, except for adding a little noise to the input to learn a more robust representation and enhance the generalization ability. The new samples with noise are represented as $x' = [x'_1, \cdots, x'_n]$, and mean-square error (MSE) is adopted as loss function as follows:

$$L(x', z) = \frac{1}{n} \sum_{i=1}^{n} \|x'_i - z_i\|^2$$

By stacking multiple DAEs, a Stacked Denoising Autoencoder (SDAE) can be constructed as shown in Fig. 1(b).

**III. PROPOSED METHOD**

The main idea of stacking is improving the heterogeneity of base models and combining the results of base models with a meta model to enhance the generalization ability. In this section, the proposed multiple feature fusion with stacking architecture for EFD is presented. The base models are One-Class SVM models with different input features, including deep TD, FD and TFD features. The outputs of the base models are integrated by a One-Class SVM meta model.

**A. Framework of the Proposed Method**

The framework of the proposed M2FSA based EFD method is shown in Fig. 3. In the training process, the proposed model is trained by a few normal data only. At first, Time Domain (TD) features, Frequency Domain (FD) features and Time-Frequency Domain (TFD) features are extracted from raw signals. TD features are statistical features of raw signals. FD features and TFD features are extracted by FFT and HHT respectively. However, the dimensions of FD features and TFD features are much higher than TD features. More importantly, these features contain noise information. In order to eliminate the interference of noise and unify the dimension of different features, SDAE is conducted to extract deep features in three domains.
Subsequently, three One-Class SVM base models are trained by deep TD, FD and TFD features of training set respectively. The outputs of the three base models are distances from each sample to the hyperplanes learned by One-Class SVM. As described in Section II-C, the outputs of base models in training set can be used as inputs to train the One-Class SVM meta model.

In the definition of One-Class SVM, the distance $d$ from the sample to the hyperplane can be seen as the confidence level of normal or abnormal state. If $d \geq 0$, the larger the $d$ is, the more likely the sample is normal. Similarly, if $d < 0$, a smaller $d$ means that the sample deviates more from normal state. In EFD tasks, the normal samples and severe fault samples are easily to be distinguished. However, the samples around the appearance of early faults are located around the hyperplane. The decision of the health states of these samples have a low confidence level. Simply detecting the states with a single means that the sample deviates more from normal state. In EFD tasks, the normal samples and severe fault samples are easily to be distinguished. However, the samples around the appearance of early faults are located around the hyperplane. The decision of the health states of these samples have a low confidence level. Simply detecting the states with a single.

In the prediction phase, deep features are extracted by SDAE models. Subsequently, the distances from samples to hyperplanes are calculated by the base models. At last, the state of equipment are obtained by the meta model. The details of the training and prediction processes are summarized as follows:

![Fig. 4. Training process of the ith Base Model B(i)](image)

1) Training Process of the proposed method: The steps of the base model training process are shown in Fig. 4. The training set is consisted of normal state data only. Let $X_{\text{train}} = [x_1, \ldots, x_n]$ represent the training data, where $n$ is the number of training samples. At first, the domain features corresponding to the ith base model $B(i)$ are extracted, which are denoted as $\{x(i)_j\}_{j=1}^n$. Secondly, the SDAE model $f(i)$ is utilized to extract deep domain features of the training data. Let $Z_{\text{train}}^{(i)} = [z^{(i)}_1, \ldots, z^{(i)}_n]$ represent the training set with deep features. In order to deal with the problem of overfitting, the $K$-fold cross validation strategy is utilized in the training process of base models. As shown in Fig. 4, the training data with deep features are firstly divided into $K$ sub-training sets, namely $S^{(i)}_1, \ldots, S^{(i)}_K$. Then, the ith Base Model $B(i)$ is trained $K$ times independently with different data. In each time, a subset $S^{(i)}_k$ is selected as validation data and the rest of the training set are served as training data. In the end of training, $K$ sub-models of the Base Model $B(i)$ are obtained, which are denoted as $B^{(i)}_1, \ldots, B^{(i)}_K$. The corresponding hyperplanes can be represented by $P^{(i)}_1, \ldots, P^{(i)}_K$, where $P^{(i)}_k : w^{(i)}_k \cdot x - \rho^{(i)}_k = 0$. The distances from samples of validation subset $S^{(i)}_k$ to the hyperplane $P^{(i)}_k$ obtained by the corresponding sub-model can be calculated as follows.

$$D^{(i)}_k = \langle w^{(i)}_k, S^{(i)}_k \rangle - \rho^{(i)}_k$$  \hspace{1cm} (8)

where $i = 1, \ldots, N$, $N$ is the number of base models. $k = 1, \ldots, K$, $D^{(i)}_k, S^{(i)}_k \in \mathbb{R}^M$, $M$ is the number of samples in the subset $S^{(i)}_k$. In the training process, the distances of each validation subsets are concatenated to generate the outputs $O^{(i)}$ of training data by the base model $B(i)$ as shown in Eq. (9).

$$O^{(i)} = C(D^{(i)}_1, \ldots, D^{(i)}_K)$$  \hspace{1cm} (9)

Algorithm 1 Multiple Feature Fusion with Stacking Architecture (M2FSA)

Input: Training data $\{x_j\}_{j=1}^n$, Number of cross validation $K$, New arrival data $x_{\text{new}}$

1: Extract TD $\{x^{(1)}_j\}_{j=1}^n$, FD $\{x^{(2)}_j\}_{j=1}^n$ and TFD $\{x^{(3)}_j\}_{j=1}^n$ features of $\{x_j\}_{j=1}^n$.
2: for $i = 1, 2, 3$ do
3: Train deep model SDAE $f_i$ by $\{x^{(i)}_j\}_{j=1}^n$.
4: end for
5: for $i = 1, 2, 3$ do
6: Extract deep features of training data by $x^{(i)}_j = f(i)(x_j)$
7: Randomly divide $\{z^{(i)}_j\}_{j=1}^n$ into $K$ subsets $S^{(i)}_k = \{S^{(i)}_{k,j}\}_{j=1}^n$ for $k = 1, \ldots, K$ do
8: Train the base One-Class SVM model $B^{(i)}_k$ by $S^{(i)}_k = \{S^{(i)}_{k,j}\}_{j=1}^n$.
9: Calculate the outputs of $S^{(i)}_k$ by $B^{(i)}_k$ with Eq. $8$.
10: end for
11: Obtain the training outputs $O^{(i)}$ by $B^{(i)}_k$ with Eq. $9$.
12: end for
13: Train the meta One-Class SVM $g(\cdot)$ by $[O^{(1)}, O^{(2)}, O^{(3)}]$
14: end for
15: for $m = 1, 2, 3$ do
16: Extract deep features of $x_{\text{new}}$ by $z_{\text{new}} = f(1)(x_{\text{new}})$
17: Calculate the new arrival data output $\hat{o}_{\text{new}}^{(m)}$ of $B^{(i)}_m$ by Eq. $10$.
18: end for
19: Predict the state of $x_{\text{new}}$ by $y = g(\hat{o}_{\text{new}}^{(1)}, \hat{o}_{\text{new}}^{(2)}, \hat{o}_{\text{new}}^{(3)})$

Output: Health state $y$ of new arrival data.
where $C()$ is the concatenation operation. Each base model is trained by one type of features independently. In the meta model training process, the outputs of all base models are combined as the input of $N$-dimension to the meta model. All of the base model outputs of training data are used to train the meta model $g(\cdot)$.

2) Prediction Process of the proposed method: In the prediction process, let $x_{new}$ represent the new arrival data. The raw TD, FD and TFD domain features are extracted from $x_{new}$. Subsequently, the corresponding SDAE model in each domain is utilized to obtain the deep features. Let $z_{new}^{(1)}$, $z_{new}^{(2)}$, $z_{new}^{(3)}$ represent the deep TD, FD and TFD domain features respectively. Different from the training process of base models, the prediction output of the sample $x_{new}$ in base model $B^{(i)}$ is the average of the distance from deep features of $x_{new}$ to each hyperplanes of sub-models of $B^{(i)}$. The calculation of the output of base model $B^{(i)}$ is given by Eq. (10).

$$o_{new}^{(i)} = \frac{1}{K} \sum_{k=1}^{K} [\langle w_k^{(i)} , z_{new}^{(i)} \rangle - \rho_k^{(i)}]$$  \hspace{1cm} (10)

After obtaining the base features of different base models, the features are merged and fed into the meta model to make a final decision of the equipment state. The pseudo code of M2FSA is shown in Algorithm 1.

IV. EXPERIMENTS

The performance of the proposed method is evaluated on three frequently used run-to-failure bearing datasets, i.e., IEEE PHM Challenge 2012 dataset [24], IMS dataset [25] and XJTU-SY dataset [26].

A. Datasets

1) IEEE PHM 2012: The IEEE PHM Challenge 2012 dataset was collected by PRONOSTIA test platform. A rotating part, a loading part, and a data collection part are installed on the PRONOSTIA test platform. The bearing is driven by a motor through the shaft in the rotating part. An extra pressure is applied on the bearing to accelerate the degradation process by the loading part. Accelerometer sensors with a sampling frequency of 25.6 KHz are installed in horizontal and vertical directions to collect vibration signals. Besides, temperature data are collected by a temperature sensor with a frequency of 10 Hz. It takes a few hours to conduct an accelerated degradation experiment and collect the run-to-failure data on this test platform.

There are several run-to-failure bearing data under three working conditions in this dataset. The conditions are listed in Table I. 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.

2) IMS: Three degradation experiments are conducted on four bearings in the IMS dataset. The vibration signals are collected with a sampling frequency of 20 kHz. At the end of the first test, an inner race fault and a ball fault occurred on bearing 3 and bearing 4 respectively. Similarly, outer race faults occurred on bearing 1 at the end of the second test and bearing 3 at the end of the third test. In this paper, we choose the bearing 1 with an outer race fault in the second experiment for testing and denote it as IMS2_1 for short.

3) 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, vibration signals in the horizontal direction are used to detect the early fault, and the 5th bearing under the first working condition is chosen for testing. The chosen bearing is denoted as XJD1_5 for short.

B. Experiments

1) Data processing: In the time domain, the extracted statistics features include mean value, maximum value, minimum value, median, mean absolute value, peak-to-peak value, standard deviation, RMS, skewness, kurtosis, crest factor, form factor, margin factor and pulse factor. These features contain most of the commonly used time domain features, which reflect the amplitude, energy and distribution of a signal over time domain. Their calculating equations are shown in Table II, where $n$ is the number of points in a signal. In the frequency and time-frequency domains, FFT and HHT are conducted to extract FD and TFD features respectively.

2) Deep feature extraction by SDAE: In this section, deep features of TD, FD, TFD are extracted by three different SDAEs. The first 500 samples of four target bearings are used to train the feature extraction model and the early fault

---

**Table I**

| Working Condition | Motor Speed/rpm | Load/N |
|-------------------|-----------------|--------|
| 1                 | 1800            | 4000   |
| 2                 | 1650            | 4200   |
| 3                 | 1500            | 5000   |

**Table II**

| No. | Time domain features | Equation |
|-----|----------------------|----------|
| 1   | Mean value           | $s_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$ |
| 2   | Maximum value        | $s_2 = \max(x)$ |
| 3   | Minimum value        | $s_3 = \min(x)$ |
| 4   | Median               | $s_4 = \text{median}(x_i)$ |
| 5   | Mean absolute value  | $s_5 = \frac{1}{n} \sum_{i=1}^{n} |x_i|$ |
| 6   | Peak-to-peak value   | $s_6 = s_2 - s_3$ |
| 7   | Standard deviation   | $s_7 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - s_1)^2}$ |
| 8   | Root mean square     | $s_8 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ |
| 9   | Skewness             | $s_9 = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - s_1)^3}{(x_i - s_3)^3}$ |
| 10  | Kurtosis             | $s_{10} = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - s_1)^4}{(x_i - s_3)^4} - 3$ |
| 11  | Crest factor         | $s_{11} = \max(|x_i|)/s_5$ |
| 12  | Form factor          | $s_{12} = \frac{\max(|x_i|)}{\sqrt{n}}$ |
| 13  | Margin factor        | $s_{13} = \frac{\max(|x_i|)}{\sum_{i=1}^{n} |x_i|/n}$ |
| 14  | Pulse factor         | $s_{14} = \frac{\max(|x_i|)}{s_5}$ |
TABLE III
Comparing results of existing EFD methods (Bold indicates the best result)

| NO. | Method               | PHM1_1 LO | PHM1_1 FA | Score  | PHM1_3 LO | PHM1_3 FA | Score  | IMS2_1 LO | IMS2_1 FA | Score  | XJD1_5 LO | XJD1_5 FA | Score  | Average Score |
|-----|----------------------|-----------|-----------|--------|-----------|-----------|--------|-----------|-----------|--------|-----------|-----------|--------|----------------|
| 1   | BEMD+AMMA            | 1900 /    | 760.00    | 1600 / | 640.00    | 535      | 214.00 | 1090 /    | 436.00    | 512.50 |
| 2   | RMS+CC               | 1690 /    | 676.00    | 1330 / | 532.00    | 546      | 218.40 | 1392      | 516.00    | 497.00 |
| 3   | FDDA                 | 1392 45   | 583.80    | 1837 38| 757.60    | 546      | 0      | 1392 15   | 565.80    | 531.40 |
| 4   | S4VM+SODRM           | 1480 3    | 593.80    | 1290 1| 516.00    | 535      | 2      | 215.20    | 1392 1    | 516.00 |
| 5   | Ours                 | 1136 0    | 454.40    | 1327 0| 530.80    | 535      | 0      | 214.00    | 1085 0    | 434.00 |

Method 1: BEMD+AMMA  
Method 2: RMS+Correlation Coefficient (CC)  
Method 3: FDDA  
Method 4: S4VM+SODRM  
Method 5: Ours

According to the proposed method, the early fault occurs at the 1136th sample of PHM1_1 with

C. Comparative Results with Existing EFD Methods

1) Methods for comparison: To verify the effectiveness of the proposed method, four comparative methods are introduced as follows.

- BEMD+AMMA: Bandwidth Empirical Mode Decomposition (BEMD) and Adaptive multiscale morphology analysis (AMMA).
- RMS+CC: Root Mean Square (RMS) and Correlation Coefficient (CC).
- FDDA: Fault Detection with Deep Architecture (FDDA).
- S4VM+SODRM: Semi-Supervised Learning (S4VM) and Online Dimension Reduction (SODRM).
no false alarm. The proposed method detects the early fault occurred at the 1327\textsubscript{th} sample of PHM1\_3, and there is no false alarm as well. The early fault is thought to be appeared at the 535\textsubscript{th} sample in IMS2\_1 and the 1085\textsubscript{th} sample in XJD1\_5, respectively. There are no false alarm in these bearings as well. The results of the false alarms indicate the high reliability of the proposed method on detecting early faults precisely. The comparative experiment results are presented in Table IV. It should be noted that there are no false alarms in Methods 1 and 2 because they calculate the fault characteristic frequency and correlation coefficient of online data to alarm early fault directly. Compared with Method 1 and 2, the proposed method detects faults earlier on all experiments with no false alarms. Compared with Method 3, the proposed method detects early fault earlier in all four bearings with less false alarms. Although the proposed method does not provide a better early fault detection of PHM1\_3 than Method 4, it generates less false alarms than Method 4. Moreover, the proposed method achieves the lowest average score of 408.30, which is the best among all methods. The results show that the proposed method can detect incipient fault earlier and precisely with higher reliability.

D. Comparative Results of One-Class SVM with Different Features

In order to validate the effectiveness of the feature fusion strategy of the proposed method, we also conduct One-Class SVM with different features to detect early faults.

1) Methods for comparison: The comparative methods are introduced as follows.

Method 1: Deep TD+One-Class SVM
Method 2: Deep FD+One-Class SVM
Method 3: Deep TFD+One-Class SVM
Method 4: Ours-Raw
Method 5: Ours-PCA
Method 6: Ours

Method 1, 2 and 3 are One-Class SVM models trained by deep TD features, FD features and TFD features respectively. Method 4 removes the SDAE models of the proposed method which utilizes raw TD, FD, TFD features to develop M2FSA model. Method 5 replaces SDAE of the proposed method with Principal component analysis (PCA) to extract domain features.

2) Experiment results: The comparative results is shown in Table IV. Compared with Methods 1, 2 and 3, the proposed method can give a more robust and earlier alarm for incipient faults than these methods. Although the detected locations of early faults in PHM1\_1 of Method 2 and 3 are a little better than the proposed method, the false alarms are larger than the proposed method. The results of the proposed method are better than Method 1, 2 and 3 on other bearings both on the location of early faults and false alarms. The comparative results show that the results of early fault detection with multiple features are more generic than using any single feature.

Compared the proposed method with Method 4 and 5, Method 4 detects the early fault of PHM1\_3 earlier than the proposed method, but it generates more false alarms. Similarly, the detected early fault location of XJD1\_5 by Method 5 is a little better than the proposed method, but the false alarms is much larger. The detection results of Method 4 and 5 on other bearings are not only later than the proposed method, but also have more false alarms. Besides, the average score of the proposed method is the best among these methods. Although Method 4 achieves the same average score as the proposed method, it generates a large number of false alarms on three bearings. The results prove that SDAE contributes to extract discriminative features for early fault detection and improve the reliability of the EFD model.

To further evaluate the performance of the proposed method, the features of PHM1\_3 extracted by different methods are visualized in Fig. 5. The deep TD, FD and TFD domain features shown in Fig 5(a), 5(b) and 5(c) indicate that severe faults and normal data do not have a clear boundary in the feature space. The overlap of normal and abnormal data leads to a lot of false alarms in the results of early fault detection based on these features. Especially in Fig. 5(c), the incipient fault data are submerged in normal and severe fault data, which makes the early fault detection of PHM1\_3 by TFD features generate the largest false alarms in all methods. Although the fault data of Ours-PCA are separated from the majority of normal data, the overlap of abnormal data and a few normal data causes a small number of false alarms. As shown in Fig. 5(d) and 5(f), the severe fault data and normal data of Ours-Raw and the proposed method are both separated clearly. But the incipient fault data of the proposed method are farther from normal data than Ours-Raw. The results show that the proposed method can extract discriminative features so that improve the reliability of early fault detection.

V. CONCLUSION AND FUTURE WORK

In this paper, a new early fault detection method for rotating machines based on multiple feature fusion with stacking
architecture is proposed. The main idea involves integrating multiple features extracted from raw signals to generate a more comprehensive representation by stacking ensemble learning. The proposed method is evaluated on three frequently used run-to-failure bearing datasets. The experimental results show that the performance of the proposed method is better than conducting early fault detection based on One-Class SVM using any single type of features. More importantly, the proposed method has stronger generalization ability than using any single feature, which is more practical in application. Furthermore, the SDAE is proved to be more effective than other feature extraction methods. Compared with existing methods, the proposed method generates the lowest false alarms and detects incipient faults earlier than other methods. The analysis of the extracted features indicates that the proposed method can extract more discriminative features for early fault detection with higher reliability.

The limitations of the proposed method include the following aspects for real applications. Firstly, a few normal samples of the target equipment are required to train the proposed early fault detection model. Secondly, the proposed method can only detect the occurrence of an early fault, but cannot predict the failure type of the equipment. In the future, a failure type prediction model can be developed based on the degradation trend extracted by the proposed method. Besides, transfer learning or meta learning based early fault detection method can be further explored to address the cold start problem.

REFERENCES

[1] B. Peng, S. Wan, Y. Bi, B. Xue, and M. Zhang, “Automatic feature extraction and construction using genetic programming for rotating machinery fault diagnosis,” IEEE Transactions on Cybernetics, 2020.

[2] A. Silva, A. Zarzo, J. M. M. González, and J. M. Muñoz-Guijosa, “Early fault detection of single-point rub in gas turbines with accelerometers on the casing based on continuous wavelet transform,” Journal of Sound and Vibration, vol. 487, pp. 107583, 2020.

[3] A. Glowacz, W. Glowacz, Z. Glowacz, and J. Kozik, “Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals,” Measurement, vol. 113, pp. 1–9, 2018.

[4] F. Chen, M. Cheng, B. Tang, W. Xiao, B. Chen, and X. Shi, “A novel optimized multi-kernel relevance vector machine with selected sensitive features and its application in early fault diagnosis for rolling bearings,” Measurement, vol. 156, p. 107583, 2020.

[5] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, and W. Zhantao, “Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing,” Knowledge-Based Systems, vol. 188, p. 105022, 2020.

[6] N. Lu, H. Hu, T. Yin, Y. Lei, and S. Wang, “Transfer relation network for fault diagnosis of rotating machinery with small data,” IEEE Transactions on Cybernetics, 2021.

[7] N. Li, Y. Lei, J. Lin, and S. X. Ding, “An improved exponential model for predicting remaining useful life of rolling element bearings,” IEEE Transactions on Industrial Electronics, vol. 62, no. 12, pp. 7762–7773, 2015.

[8] Y. Hong, M. Kim, H. Lee, J. J. Park, and D. Lee, “Early fault diagnosis and classification of ball bearing using enhanced kurtogram and Gaussian mixture model,” IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 12, pp. 4746–4755, 2019.

[9] M. Sohail, C.-H. Kim, and J.-M. Kim, “A hybrid feature model and deep-learning-based bearing fault diagnosis,” Sensors, vol. 17, no. 12, p. 2876, 2017.

[10] Y. Xu, D. Zhen, J. X. Gu, K. Rabeye, 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.

[11] M. Sun, H. Wang, P. Liu, S. Huang, P. Wang, and J. Meng, “Stack autoencoder transfer learning algorithm for bearing fault diagnosis based on class separation and domain fusion,” IEEE Transactions on Industrial Electronics, 2021.

[12] J. Burriel-Valencia, R. Puche-Panadero, J. Martínezm-Roman, A. Sapena-Bano, and M. Pineda-Sánchez, “Short-frequency fourier transform for fault diagnosis of induction machines working in transient regime,” IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 3, pp. 432–440, 2017.

[13] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, and Y. Zhang, “Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform,” Computers in industry, vol. 106, pp. 48–59, 2019.

[14] X. Wen, G. Lu, J. Liu, and P. Yan, “Graph modeling of singular values for early fault detection and diagnosis of rolling element bearings,” Mechanical Systems and Signal Processing, vol. 145, p. 106956, 2020.

[15] W. Mao, S. Tian, J. Fan, X. Liang, and A. Sañan, “Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation,” Journal of Manufacturing Systems, vol. 55, pp. 179–198, 2020.

[16] 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.

[17] X. Huang, D. Wu, and B. Boulet, “Ensemble learning for charging load forecasting of electric vehicle charging stations,” in 2020 IEEE Electric Power and Energy Conference (EPEC). IEEE, 2020, pp. 1–5.

[18] Z. Mao, M. Xia, B. Jiang, D. Xu, and P. Shi, “Incipient fault diagnosis for high-speed train traction systems via stacked generalization,” IEEE Transactions on Cybernetics, 2020.

[19] H. Zhang, J.-L. Li, X.-M. Liu, and C. Dong, “Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection,” Future Generation Computer Systems, vol. 122, pp. 130–143, 2021.

[20] G. Li, Y. Zheng, J. Liu, Z. Zhou, C. Xu, X. Fang, and Q. Yao, “An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information,” Journal of Building Engineering, p. 102812, 2021.

[21] Z. Liu, M. Zhang, F. Liu, and B. Zhang, “Multidimensional feature fusion and ensemble learning-based fault diagnosis for the braking system of heavy-haul train,” IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 41–51, 2020.

[22] Y. Cheng, J. Wu, H. Zhu, S. W. Or, and X. Shao, “Remaining useful life prognosis based on ensemble long short-term memory neural network,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–12, 2020.

[23] S. Cui, Y. Yin, D. Wang, Z. Li, and Y. Wang, “A stacking-based ensemble learning method for earthquake casualty prediction,” Applied Soft Computing, vol. 101, p. 107038, 2021.

[24] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Vernier, “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.

[25] H. Qiu, J. Lee, J. Lin, and G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics,” Journal of sound and vibration, vol. 289, no. 4-5, pp. 1066–1090, 2006.

[26] 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.

[27] Y. Li, M. Xu, X. Liang, and W. Huang, “Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings,” IEEE Transactions on Industrial Electronics, vol. 64, no. 8, pp. 6506–6517, 2017.

[28] Z. Guo, G. Jiang, H. Chen, and K. Yoshihira, “Tracking probabilistic correlation of monitoring data for fault detection in complex systems,” in International Conference on Dependable Systems and Networks (DSN’06). IEEE, 2006, pp. 259–268.

[29] 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.