IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 2, 2021, accepted February 8, 2021, date of publication February 12, 2021, date of current version February 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058907

Online Incipient Fault Detection Method Based on Improved ℓ_1 Trend Filtering and Support Vector Data Description

QINGFENG WANG[1,2], XIAOJIN LIU[1,2], BINGKUN WEI[1,2], AND WENWU CHEN[3]

[1]Beijing Key Laboratory of Health Monitoring and Self-Recovery of High-End Machinery Equipment, Beijing University of Chemical Technology, Beijing 100029, China
[2]Diagnosis and Self-Recovery Engineering Research Center, Beijing University of Chemical Technology, Beijing 100029, China
[3]State Key Laboratory of Safety and Control for Chemicals, China Petroleum and Chemical Corporation Research Institute of Safety Engineering, Qingdao 266071, China

Corresponding author: Qingfeng Wang (wangqf2422@buct.edu.cn)

This work was supported by the China Petroleum and Chemical Corporation Ministry of Science and Technology under Grant 320059 and Grant 319022-1.

ABSTRACT Poor model generalization, missing or false alarms, and heavy dependence on expert's experience are some of the major problems which exist in traditional incipient fault detection (IFD) methods. An IFD rolling bearing application method based on combination of improved λ_1 trend filtering (L1TF) and support vector data description (SVDD) is proposed. First, spectral distance index and multi-scale dispersion entropy based on normal vibration data, which is sensitive to incipient faults, are extracted. The improved λ_1 trend filter (IL1TF) method is employed for processing the feature values and obtaining a trend factor with less fluctuation and better incipient fault indication ability. Then, after determining the kernel function bandwidth of the SVDD by analyzing the characteristics of the training data, a suitable offline SVDD model is trained. Finally, incipient faults are identified by estimating the distance between the trend factor of the real-time data and the center of the hypersphere in the SVDD model. This method employs full performance of SVDD to detect abnormal data files, while reducing the influence of abnormal data files on the model via IL1TF. Furthermore, the method increases the discrimination between the incipient fault data and the normal data. By utilizing Intelligent Maintenance Systems of University of Cincinnati bearing laboratory data and Chinese petrochemical company's centrifugal pump bearing engineering data, the effectiveness of the constructed model is demonstrated. In addition, the proposed method is compared against existing representative IFD methods. The results indicate that the method proposed in this paper can solve false alarms and detect incipient failure data files more accurately without depending on the external expert's experience. This is of great significance for providing guidelines to enterprises which employ predictive maintenance techniques.

INDEX TERMS Improved ℓ_1 trend filtering, support vector data description, kernel function bandwidth determination, incipient fault detection.

I. INTRODUCTION

As one of the most commonly employed parts in rotating machinery, rolling bearings operate under harsh working environment and are often prone to failures. Their performance degradation is an important issue that threatens the safe operation of rotating machinery [1]. Rolling bearings degradation process during the entire life cycle can be roughly divided into four stages: normal operation stage, incipient failure stage, severe failure stage, and failure stage [2]. Therefore, quick and accurate identification of the time required for the normal operation state to achieve incipient failure state is particularly important. This, in turn, would provide sufficient time for applying the predictive maintenance strategy and reducing the possibility of production accidents.

Currently, online incipient fault detection (IFD) needs to meet certain requirements in the actual engineering application environment. Completing the online IFD should be done

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

within the time limit and database limit of online applications (generally only historical normal operating status data). The problem of false alarms caused by abnormal data files should be solved within the time limit while detecting incipient faults of rolling bearings. Incipient fault detection methods should not rely on prior knowledge to set key hyperparameters for IFD models. Furthermore, they should not depend on expert knowledge to determine if an incipient failure has occurred. Lastly, IFD models must have good generalization.

Bearing vibration signal analysis is one of the most commonly employed methods for detecting incipient faults. In recent years, researchers have proposed many IFD methods which can be divided into physical model-based methods and data-driven methods [3]. In [4]–[6], an IFD method based on prior physical knowledge, which requires the establishment of an accurate physical model, has been investigated. However, the internal stress and other physical quantities of the running rolling bearings are relatively difficult to characterize, which often results in poor model generalization [7]. An IFD rolling bearings method based on the analysis of fault characteristic frequencies is established in [8]–[10]. These methods do not require pre-training models and have beneficial online application capabilities. However, they have certain disadvantages. The weak incipient fault characteristic signals may be submerged within the noise signal. Equipment vibration signal noise is different, and the preprocessing methods such as noise reduction are also different. Hence, this results in poor model generalization. Furthermore, a significant amount of time is consumed to perform online mode decomposition and noise reduction, which is not conducive for engineering applications.

Data-driven IFD methods mainly use the intrinsic characteristics of data for fault detection research. These methods have received wide attention from many researchers since they do not require precise physical knowledge. Zan *et al.* [11] proposed a method for evaluating performance deterioration degree based on the joint approximative diagonalization of eigen-matrices and particle swarm optimization support vector machine. Rai *et al.* [12] used a method based on the empirical mode decomposition and k-medoids clustering. Zhou *et al.* [13] utilized wavelet packet energy entropy and radial basis function neural network to detect the time of the incipient bearing failure. Mao *et al.* [14] proposed a semi-supervised architecture and depth feature representation method for bearing online IFD, which was based on stacked denoising auto-encoder and semi-supervised support vector machine. In addition, Lu *et al.* [15] proposed a deep-structured framework to detect the incipient fault, which was based on deep neural network (DNN) and long short-term memory (LSTM). These methods can achieve IFD by constructing bearing performance degradation models. Zan's method can accurately predict the performance degradation trend and the remaining service life of rolling bearings for small samples; However, the conventional particle swarm algorithm that easily falls into local optimum [16]. Rai and Zhou's method is able to timely diagnose the incipient bearings failure, and Mao's method performs well on bearing online IFD. However, the models training require fault samples of bearings in these IFD methods. Lu's method can detect incipient failures timely, but he did not explain how to choose appropriate parameters of DNN and LSTM.

As a typical one-class classification algorithm, the support vector data description (SVDD) requires only normal samples to train a sensitive outlier detection model [17]. Thus, the SVDD is widely used in data-driven IFD methods. Mao *et al.* [18] proposed a rolling bearing IFD method based on self-adaptive deep feature matching and the SVDD. Yang *et al.* [19] proposed an IFD method based on incremental weighted SVDD. The method can solve the problem that the SVDD is susceptible to slight fluctuations which cause false alarms. With the purpose of investigating the small number of fault samples and the susceptibility of the model to abnormal data, Zhu *et al.* [20] constructed three one-class classifiers: projection support vector data description (PSVDD), projection K-means (PK-means) and projection K-center (PK-centers). In order to reduce the input features impact on the performance of a degradation model, Wang *et al.* [21] proposed an IFD method that combines canonical variate analysis with the SVDD. Zhang *et al.* [22] proposed an improved incremental SVDD algorithm to achieve online update of the normal sample set, and to improve the accuracy and computational efficiency of the offline part. Aiming at the weak incipient faults that are easily covered by system interference and noise, Zhang *et al.* [23] employed normal and faulty samples to construct a robust SVDD model based on the traditional SVDD, this model improves the calculation of the spherical radius. K-nearest neighbor algorithm has been widely used for the classification, where distance from the target sample to the classification center is taken as the weight of the slack variable in the SVDD model [24]. In this way, a defect that considers only the distance between the center of the hypersphere and the training sample, not the data in the high-density area is avoided. The aforementioned researchers have promoted the engineering application of the SVDD. However, there are still some shortcomings that should be further addressed. The bandwidth value of the kernel function [19], [20], [22], which is a key parameter of the SVDD, has not been illustrated in detail. Furthermore, the model constructed in [21] has low sensitivity and delayed time for incipient fault detection. The model proposed in [23] requires training by a quantity fault case data. Lastly, false alarms caused by the abnormal data fluctuations have not been considered for the model in [24].

SVDD is usually employed together with a kernel function which maps the input feature values to a high-dimensional space. Thus, low-dimensional linear inseparability problem is converted into a linearly separable problem. The Gaussian kernel function is widely used as kernel function. The SVDD model based on Gaussian kernel function has two key parameters: penalty factor $\boldsymbol{C}$, and kernel function bandwidth $\boldsymbol{\sigma}$. These factors need to be determined in advance. Previous investigation results indicate that the SVDD is not susceptive

to the parameter $\boldsymbol{C}$. On the contrary, performance of the SVDD largely relies on the $\boldsymbol{\sigma}$ selection [25], [26]. In recent years, many researchers have carried out a significant amount of research on $\boldsymbol{\sigma}$ selection. Tax and Duin [27] proposed a method of optimal $\boldsymbol{\sigma}$ selection by minimizing the support vector to sample ratio. Khazai *et al.* [28] suggested a process of optimizing $\boldsymbol{\sigma}$ parameter based on the maximum distance between the samples. Wang *et al.* [29] proposed a method for optimizing $\boldsymbol{\sigma}$ using the boundary tightness. Liu *et al.* [30] suggested a method for locating the maximum value of the variance-square root mean-ratio (VSRMR) for selecting $\boldsymbol{\sigma}$. These methods have made significant contributions to further application of the SVDD. However, the methods also have some shortcomings. The models proposed in [17] and [28] are computationally intensive and, consequently, time consuming. Furthermore, underfitting is prone to appear for the model [24]. Lastly, according to the verification conducted by the authors, the method [30] is not suitable when magnitude of the input feature value is relatively small.

In summary, the SVDD has wide applications in the bearings IFD field. The current research hotspot of IFD methods based on the SVDD is to improve the usability, robustness, generalization and accuracy of the IFD model in order to satisfy online engineering applications. When employing the SVDD model to detect the occurrence of incipient faults, same event (caused by real-time data fluctuation) alarm repetition can be reduced or avoided. In order to adequately utilize the SVDD, its key parameters have to be quantified and determined. In this paper, with the purpose of solving the problems such as incipient faults detection difficulties in engineering practice, and overcoming the deficiencies of the above provided investigations, an online bearing IFD method based on improved ℓ_1 trend filter (IL1TF) and SVDD is proposed.

The ℓ_1 trend filter (L1TF) is improved on the H-P filter by Kim *et al.* [31]. Compared with the H-P filter, the L1TF has excellent piecewise linearity and can extract linear time series trend with great accuracy. Gowri *et al.* [32] used a combination of variational mode decomposition (VMD) and L1TF to enhance speech signals in strong Gaussian white noise signals. Selvin *et al.* [33] applied L1TF to image the noise reduction. Ottersten *et al.* [34] improved the L1TF and proposed the ℓ_1 mean filter to accurately calculate the turning point. L1TF can adequately extract the piecewise linear trend of the time series. However, the performance of L1TF depends on whether the selected regularization parameter λ is appropriate. Currently, there is no suitable method to solve this problem. Moreover, in the practical engineering application process, requirements of L1TF processing for the full life cycle data cannot be met [35]. In this paper, a combined sliding window for sequential data reading is proposed. The consistent slope of the filtered data onto the sliding combined window is employed to adaptively select the regularization parameter λ, which provides a new direction of L1TF application in engineering.

IL1TF can reduce the impact of occasional feature value abnormal samples, expand the deterioration trend near the incipient fault occurrence moment, and increase the sensitivity of the IFD model while solving the problem of false alarms. In this paper, an approximation is made to determine the bandwidth of the kernel function according to the divergence of the filtered training data in the high-dimensional kernel function matrix. Furthermore, the value problem of key parameters which affect the SVDD performance is solved. The contribution of this paper can be summarized as follows:

1) The L1TF is improved, providing a new method for the selection of L1TF key parameters and the real-time filtering of online data. The feature values trend factor obtained after the IL1TF process can reduce both impact of occasional abnormal samples and data fluctuations, increase the degradation trend near the incipient failure samples, and provide theoretical basis for distinguishing the incipient failure samples and false alarm samples in the online detection phase.

2) The kernel function bandwidth is approximately calculated by the divergence of training data onto the high-dimensional kernel function matrix. The results indicate that the hypersphere trained by the kernel function bandwidth and calculated by the proposed method is of moderate size, i.e., there are no over-fitting and under-fitting problems.

3) The online IFD method based on IL1TF and the SVDD proposed in this paper can quickly and accurately identify the incipient fault samples. Furthermore, it can solve the problems of false alarm interference and significant data fluctuations following the alarm.

The remainder of this paper is arranged as follows: In Section 2, some basic theoretical knowledge of L1TF and SVDD is introduced. In Section 3, details about the proposed method are described. The method is verified against laboratory data and engineering case data in Section 4. Comparison of the results is conducted in Section 5. Finally, conclusions and future work are summarized in Section 6.

II. REVIEW OF ℓ_1 TREND FILTERING AND SUPPORT VECTOR DATA DESCRIPTION

A. ℓ_1 TREND FILTERINGS

Kim *et al.* [31] proposed the L1TF based on the H-P filter. Both H-P trend filter and L1TF utilize the least square cost function. The difference is that H-P filter uses the ℓ_2 norm when calculating second-order difference matrix, while the L1TF uses ℓ_1 norm. The ℓ_1 norm nullifies many second-order difference terms. Thus, the resulting trend term is piecewise linear. Therefore, the ℓ_1 trend filter is suitable for estimating the trend, the inflection point or slope of a time series.

Suppose that time series $y_i (i = 1, \cdots, n)$ consists of a basic trend sequence x_i and a random variable sequence z_i. Trend filtering is employed to estimate the trend sequence from y_i. This trend estimation is achieved by minimizing the weighted objective function, which is shown in Eq. (1):

$$(1/2)\sum_{i=1}^{n}(y_i - x_i)^2 + \lambda \sum_{i=2}^{n-1} |x_{i-1} - 2x_i + x_{i+1}| \qquad (1)$$

Eq. (1) can be converted into the matrix norm form of Eq. (2):

$$(1/2)\left\|y-x\right\|_2^2+\lambda\left\|Dx\right\|_1 \tag{2}$$

where $y=(y_1,y_2,\cdots,y_n)\in R^n$, $\|u\|_1=\sum_i|u_i|$, represents the ℓ_1 norm of $\boldsymbol{u}$, and $D\in R^{(n-2)\times n}$ is a second-order difference matrix:

$$D=\begin{bmatrix}1&-2&1&&\\&1&-2&1&\\&&\ddots&\ddots&\ddots\\&&&1&-2\ 1\end{bmatrix} \tag{3}$$

In Eq. (2), the first part represents the gap between the actual series y_i and the trend series x_i, which is measured by the least squares cost function $(1/2)\|y-x\|_2^2$. The second part represents the penalty for the smoothness of x_i, expressed by the ℓ_1 norm of the second-order difference matrix $\|Dx\|_1$. The two parts are connected by the regularized parameter λ to control the weight. The regularization parameter λ is a non-negative parameter for controlling the balance between the estimated trend and the signal redundancy. The weighted objective function is a strictly convex function of unique solution. Therefore, x_i has only one solution. L1TF is used to filter the bearing monitoring signal. Thus, true degradation trend of the bearing well is obtained.

B. SUPPORT VECTOR DATA DESCRIPTION

SVDD is a one-class classifier. Its core idea is to map the given training data onto a high-dimensional space and find a minimum hypersphere. In that way, as many of the training samples as possible are contained in the hypersphere. Samples that do not belong in this category are located outside the hypersphere. Its mathematical model is constructed as follows:

Presuming a training sample set $x_i=\{x_1,x_2,\cdots,x_l\}$, where $\boldsymbol{l}$ is the number of samples in this training set. Define a hypersphere, with center $\boldsymbol{a}$ and radius $\boldsymbol{r}$, which containing almost all training samples. Considering that there may be a small number of errors in the training samples, in order to improve the algorithm's robustness to the outliers in the training samples, a slack variable $\xi_i>0$ is introduced to allow some samples to be distributed outside the hypersphere. The SVDD optimization problem is described as:

$$\begin{cases}\min\varepsilon(R,a,\xi_i)=r^2+C\sum_i\xi_i\\ s.t.\ \|x_i-a\|^2\le r^2+\xi_i\\ \xi_i\ge 0\end{cases} \tag{4}$$

where $C>0$ is the penalty parameter used to control the trade-off between the hypersphere volume and the errors. In this paper, default value of $\boldsymbol{C}$ is set to 0.3. In order to solve the unconstrained optimization problem, a Lagrangian equation is constructed:

$$L(r,a,\alpha_i,\xi_i)=r^2+C\sum_i\xi_i-\sum_i\alpha_i\{r^2+\xi^2-(x_i^2-2ax_i+a^2)\}-\sum_i\gamma_i\xi_i \tag{5}$$

where $\alpha_i\ge 0$ and $\gamma\ge 0$ are Lagrange multipliers. Through partial derivation of Eq. (5), Eq. (6) is obtained:

$$\begin{cases}\dfrac{\partial L}{\partial R}=0:\sum_i\alpha_i=1\\ \dfrac{\partial L}{\partial a}=0:a=\sum_i\alpha_ix_i\\ \dfrac{\partial L}{\partial\xi_i}=0:C-\alpha_i-\gamma_i=0\end{cases} \tag{6}$$

For the sample located in the hypersphere, $\alpha_i=0$. For the sample located on the boundary of the hypersphere, $0<\alpha_i<\boldsymbol{C}$. Lastly, for the sample located outside the hypersphere, $\alpha_i=\boldsymbol{C}$. Samples located on and outside the boundary of the hypersphere are called the support vectors (*SV'S*).

By combining Eq. (6) and Eq. (5), an optimized function is obtained:

$$\begin{cases}\max L=\sum_i\alpha_iK(x_i,x_j)-\sum_{i,j}\alpha_i\alpha_jK(x_i,x_j)\\ s.t.\ 0\le\alpha_i\le C,\quad\sum_i\alpha_i=1\end{cases} \tag{7}$$

where $K(\cdot)$ is the kernel function, which maps the training samples from the low-dimensional spaces onto the high-dimensional space, where the training samples can be more adequately described. Gaussian kernel function is employed:

$$K(x_i,x_j)=\exp(-\frac{\left\|x_i-x_j\right\|^2}{\sigma^2}),\quad\sigma>0 \tag{8}$$

In the high-dimensional space, the generalized radius $\boldsymbol{r}$ of the hypersphere can be obtained as the distance from the *SV'S* on the hypersphere interface to the center of the hypersphere:

$$\begin{aligned}r^2&=\|\phi(x_{sv})-\phi(a)\|^2\\&=K(x_{sv},x_{sv})-2\sum_i\alpha_iK(\mathrm{x}_i,x_{sv})+\sum_{i,j}\alpha_i\alpha_jK(x_i,x_j)\end{aligned} \tag{9}$$

For the real-time input test sample $\boldsymbol{z}$, its distance from the center of the sphere can be calculated by:

$$\begin{aligned}d&=\|\phi(z)-\phi(a)\|\\&=K(z,z)-2\sum_i\alpha_iK(z,x_i)+\sum_{i,j}\alpha_i\alpha_jK(x_i,x_j)\end{aligned} \tag{10}$$

In the process of bearing online IFD, the extracted normal sample feature value is employed as the training sample input value for SVDD training. Thus, a hypersphere containing normal samples is obtained. The generalized distance $\boldsymbol{d}$ between the real-time online samples and the center of the hypersphere can be calculated via Eq. (10). When $\boldsymbol{d}\le\boldsymbol{r}$, the samples are denoted as normal. On the other hand, when $\boldsymbol{d}>\boldsymbol{r}$, the samples are suspected as failures.

III. PROPOSED METHOD

A. FEATURE EXTRACTION

Extracting a single IFD feature value is usually computationally efficient and the feature value may have a certain physical meaning. However, it is relatively difficult to use a single IFD feature value to fully reflect the fault characteristics. It is very important to extract multi-feature and multi-scale feature values to characterize the comprehensive information of the fault. However, the calculation efficiency of multi-feature and scale feature values is low, which is not suitable for online IFD applications. After weighing the pros and cons, two indicators, the spectral distance index (SDI) and the multi-scale dispersion entropy (MDE), are extracted to characterize the information of the fault. The former can characterize the operation reliability of the equipment [36], and the latter can reflect the characteristics of the fault at multiple scales [37]. The experimental validation results in this paper show that SDI and MDE are stable and reliable, sensitive to incipient faults, and suitable for their online detection.

1) SPECTRAL DISTANCE INDEX

He *et al.* [36] proposed that SDI and other methods can be used to calculate the operating reliability, establish the operating condition and reliability correlation mapping model of rotating equipment, and achieve real-time operational reliability assessment. SDI represents the operating reliability of the equipment. Its value gradually decreases, which characterizes the equipment performance degradation process from normal to faulty. The calculation method of SDI is shown in Eq. (11):

$$R\left(J_{xy}\right) = \frac{1}{1+\alpha \cdot J_{xy}} \tag{11}$$

where $J_{xy}(J_{xy} \geq 0)$ represents the $\boldsymbol{J}$ divergence between the normal condition signal and the signal designated for evaluation. Parameter α is the sensitivity coefficient, which controls the degree of J_{xy} influence on the SDI. In this paper, default value of 0.5 is employed. Mathematical expression of $\boldsymbol{J}$ divergence is as follows:

$$J_{x,y} = \frac{1}{2P}\sum_{k=1}^{P}\left[\frac{S_y(k)}{S_x(k)} + \frac{S_x(k)}{S_y(k)}\right] - 1 \tag{12}$$

where $S_x(k)$ and $S_y(k)$ are the auto-power spectrum of signals $x(t)$ and $y(t)$, respectively. $\boldsymbol{P}$ is the number of auto-power spectrum lines.

2) MULTISCALE DISPERSION ENTROPY

Under complex operating conditions, the fault vibration bearing signal often shows nonlinearity. Sample entropy (SE), permutation entropy (PE), multi-scale sample entropy (MSE) and multi-scale permutation entropy (MPE) have been widely used in the field of rolling bearing fault diagnosis since they can extract the nonlinear characteristics of signals. However, SE calculation is usually lengthy and susceptible to sudden signal changes. Although PE computes faster than SE, however the calculation process does not account for the relationship between the signal amplitudes. Due to the shortcomings of the sample entropy and the permutation entropy, Rostaghi and Azami [38] proposed Dispersion Entropy (DE), which overcomes some of the PE and SE defects. Furthermore, DE considers the relationship between amplitudes, and has the advantages of relatively fast calculation speed and less susceptibility to sudden changes.

Costa *et al.* [39], [40] proposed MSE on the basis of SE, and applied it to analyze the complexity of biological signals, which achieved good results. By coarse-graining the sequence, the sequence expression in different scales can be obtained. Azami *et al.* [41] proposed multiscale dispersion entropy (MDE) when analyzing biomedical signals. The authors conducted comparative experiments which proved that MDE calculation time is much lower than the one of MSE. MDE can reflect the complexity of time series at different scales. DE is a nonlinear dynamic method that characterizes the complexity and irregularity of time series. The DE calculation is shown as follows:

1) As shown in Eq. (13), the normal distribution function is used to map the time series $x = \left\{x_i,\ \text{i=1}, 2, \cdots, l'\right\}$ to $y = \left\{y_j, j = 1, 2, \cdots, l'\right\}, y \in (0, 1)$:

$$y_i = \frac{1}{\sqrt{2\pi}\beta}\int_{-\infty}^{x_i} e^{-\frac{(t-\mu)^2}{2\beta^2}}dt \tag{13}$$

where $\boldsymbol{\mu}$ is mean, and $\boldsymbol{\beta}$ is standard deviation.

2) As shown in Eq. (14), linear transformation is used to map $\boldsymbol{y}$ to the range of $1 \sim \boldsymbol{c}$. Thus, the sequence $z_j^{(c)}$ is obtained:

$$z_j^{(c)} = int(cy_j + 0.5) \tag{14}$$

where $\boldsymbol{c}$ is the number of categories, and ***int*** is the rounding function.

3) Embedding vector $z_i^{(m,c)}$ is calculated by:

$$z_i^{(m,c)} = (z_i^{(c)}, z_{i+d}^{(c)}, \cdots, z_{i+(m-1)d}^{(c)}) \tag{15}$$

where $\boldsymbol{m}$ represents the embedding dimension, and $\boldsymbol{d}$ represents time delay.

4) Each $z_i^{(m,c)}$ is mapped to a dispersion pattern $\pi_{v_0\, v_1 \cdots v_{m-1}}$, where $z_i^{(c)} = v_0, z_{i+d}^{(c)} = v_1, \cdots, z_{i+(m-1)d}^{(c)} = v_{m-1}$.

5) The probability of each dispersion pattern is calculated according to Eq. (16):

$$P(\pi_{v_0\, v_1 \cdots v_{m-1}}) = \frac{\text{num}(\pi_{v_0\, v_1 \cdots v_{m-1}})}{M-(m-1)d} \tag{16}$$

where $\text{num}(\pi_{v_0\, v_1 \cdots v_{m-1}})$ is the mapping number according to $z_i^{(m,c)}$ to $\pi_{v_0\, v_1 \cdots v_{m-1}}$.

6) Based on the Shannon's definition of entropy, DE is calculated as follows:

$$DE = -\sum_{\pi=1}^{c^m} P(\pi_{v_0\, v_1 \cdots v_{m-1}})\ln(P(\pi_{v_0\, v_1 \cdots v_{m-1}}) \tag{17}$$

7) Normalized DE is calculated as follows:

$$NDE = DE(x, m, c, d)/\ln(c^m) \tag{18}$$

According to the provided DE calculation method, it can be seen that when the probability of all dispersion patterns is equal, the degree of irregularity of the data is the highest while DE value has the maximum value. The greater the entropy spread, the higher the complexity of the time series. In [38], parameters are suggested for DE calculation. The embedding dimension $\boldsymbol{m}$ is more suitable when taken as a value 2 or 3. The number of categories $\boldsymbol{c}$ is an integer value from 4 to 8. Lastly, the time delay $\boldsymbol{d}$ is generally taken as 1. In this paper, $\boldsymbol{m} = 3$, $\boldsymbol{c} = 6$, and $\boldsymbol{d} = 1$.

MDE is used to coarse-grain the DE sequence, which can reflect the complexity of the sequence in various scales. The procedure for MDE calculation is as follows:

1) As shown in Eq. (19), coarse-grain processing the time series $x = \{x_i, \mathrm{i} = 1, 2, \cdots, l'\}$ is employed to obtain a sequence of length τ. In order to obtain a coarse-grained signal, the average value of each sequence is calculated. Rostaghi proposed that when $c^m < l'/\tau_{\max}$ is satisfied, the calculation result of DE is reliable [38]. In actual engineering applications, the data length is different, in order to meet the application requirements of most engineering data, $\boldsymbol{\tau}_{\max}$ should not be too large, therefore scale factor $\boldsymbol{\tau}_{\max} = 15$ is used within this paper.

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le \left| \frac{l'}{\tau} \right| \tag{19}$$

2) As shown in Eq. (20), DE of the coarse-grained signal for each scale factor $\boldsymbol{\tau}$ is calculated to obtain $MDE(x^{(\tau)}, m, c, d)$.

$$MDE(x, m, c, d, \tau) = DE(y_j^{(\tau)}, m, c, d) \tag{20}$$

In order to fully utilize the bearing state information contained in various scales and to avoid errors caused by the magnitude difference between the feature values, average NDE value is calculated at different scales to obtain the MDE used in this paper.

B. IMPROVED ℓ_1 TREND FILTERING

The regularization parameter λ is an important parameter for the successful application of L1TF. This parameter is utilized to control the balance between the estimated trend and the signal redundancy. When this parameter approaches zero, the trend of the filtered signal converges on the original signal, $\boldsymbol{Y_t}$. On the contrary, when it approaches infinity, the second-order difference objective function terms become dominant, and the filtered curve converges to the optimal affine fitting straight line, $\boldsymbol{X^{ba}}$. In the field of economics, the value of $\boldsymbol{\lambda}$ is studied by some researchers. For example, Hodrick and Prescott [42] suggested that $\boldsymbol{\lambda} = 1600$ when studying the US quarterly GDP. Bruder [43] employed cross-validation and suggested $\boldsymbol{\lambda}$ selection method of L1TF. Qin *et al.* [44] proposed a modified generalized cross-validation criterion (GCV) to calculate $\boldsymbol{\lambda}$. However, these methods are not very effective when applied to vibration data.

According to the aforementioned analyses, determination of λ depends on the characteristics of the analyzed signal. However, in actual bearing fault diagnosis, these characteristics are related to parameters such as rotation frequency and bearing size. Thus, methods mentioned in previous studies are not suitable for calculating λ. In this paper, some improvements to L1TF are made regarding the problems of difficulty in determining λ and insufficient incipient fault representation of feature values. By utilizing the SDI extracted from the bearing vibration data, IL1TF example is provided:

1) REFERENCE DATA SET FOR NORMAL OPERATION CONDITION IS FILTERED

Step 1: The SDI data set $X_0\,(t_1, t_2, \cdots, t_S)$ for historical normal operating condition is obtained. Parameter $\boldsymbol{S}$ is the length of the normal raw waveform data set.

Step 2: Parameter X_0 of length $\boldsymbol{S}$ is decomposed into several arrays $\left[x_j, x_{j+1}, \cdots, x_{j+M-1}\right]\,(j = 1, 2, \cdots, s - M + 1)$ of length $\boldsymbol{M}$ in chronological order.

Step 3: For $\boldsymbol{j} = 1$, an array $[x_1, x_2, \cdots, x_M]$ of length $\boldsymbol{M}$ is placed into a fixed window w_1. The average value K_M of the curve derivative after using IL1TF of the SDI in the window is calculated. Then, the regularization parameter λ is taken as an arbitrary value between 1 and 100.

Step 4: For $j = 2, 3, \cdots, S - M + 1$, the same method is used to construct $w_2, w_3, \cdots, w_S$. The average value of the curve derivative after using IL1TF of the SDI in the window is calculated. The value of λ is the same as that of Step 3.

Step 5: The smallest absolute value in $K_i(i = M, M + 1, \cdots, S)$ is found and recorded as K_0, where K_i represents the fluctuation of the data in the window. The array corresponding to K_0 represents the subset of length $\boldsymbol{M}$ with the smallest fluctuation selected from the SDI set under normal operation condition with the length of $\boldsymbol{S}$. This array is recorded as the reference data set $X_0(t_M)$.

2) CONSTRUCTION OF THE COMBINED SLIDING WINDOW

A combined sliding window (CSW) with a length of $\boldsymbol{M} + \boldsymbol{N}$ is constructed. Here, $\boldsymbol{M}$ represents the number of data in the fixed array, which is composed of the SDI of the reference data set $X_0(t_M)$. Parameter $\boldsymbol{N}$ represents the number of data in the active array, which is composed of the SDI of the online running data.

3) ADAPTIVE CALCULATION OF λ

Kim [31] proposed that when $\boldsymbol{\lambda}$ is relatively large, the fitting curve will converge to the optimal affine straight line. In other words, when the value of $\boldsymbol{\lambda}$ gradually increases, the fitting curve gradually becomes a line segment. As shown in Fig. 1, the first 45 data files represent normal operating condition SDI, collected from Intelligent Maintenance Systems of University of Cincinnati (IMS) [45]. The specific information about the experimental device is provided in Section 4.1. The last 5 data files are the SDI following the incipient failure.

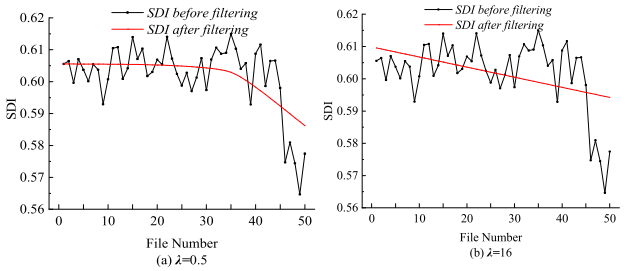

FIGURE 1. Comparison chart of SDI trend before and after LITF at different λ values.

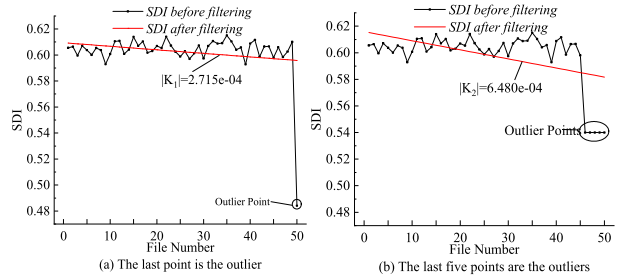

FIGURE 2. L1TF based on adaptive parameter for single and continuous outliers.

In Fig. 1(a), the fitting curve after filtering for $\boldsymbol{\lambda} = 0.5$ is shown. In Fig. 1(b), the fitting line after filtering for $\boldsymbol{\lambda} = 16$ is shown. When λ increases to 16, the slope of the curve no longer changes. It rather becomes a line segment. In this paper, L1TF on the feature values in the CSW of length $\boldsymbol{M} + \boldsymbol{N}$ is performed. Parameter $\boldsymbol{\lambda}$ gradually increases from relatively small to the constant derivative value of each data file in the CSW. This value of $\boldsymbol{\lambda}$ can be considered as a regularization parameter.

4) CALCULATION OF THE TREND FACTOR

The data in the CSW are filtered by L1TF. Parameter $\boldsymbol{\lambda}$ takes the calculation result of the previous step, and defines the absolute derivative value of the filtered line segment as the trend factor (TF). As shown in Fig. 2(a), the last data file in the CSW is an occasional abnormal data file. Since there is only one abnormal data file, it is relatively difficult to affect the overall trend of a total of 50 data files. The absolute value of the filtered line segment $|K_1|$ slope is relatively small. This can reduce the effect of the abnormal data file. As shown in Fig. 2(b), deviation amplitude of a single data file is not as large as that in Fig 2(a). However, the absolute value of the filtered line segment $|K_2|$ slope is larger. When five consecutive data files deviate from the normal data, an increase in incipient failures deterioration trend is observed. As shown in Fig. 3, the SDI changes slightly near the initial failure data files, which causes the CSW active array to gradually deviate from the fixed array. Hence, the filtered line segment has a certain slope, which is used to represent the current state. As indicated in Fig. 3 and Fig. 4, SDI and SDITF extracted from IMS Test 2 laboratory data are considered as an example. The SDI amplitudes at the 291st, 320th, 426th and other abnormal data files prior to filtering are relatively

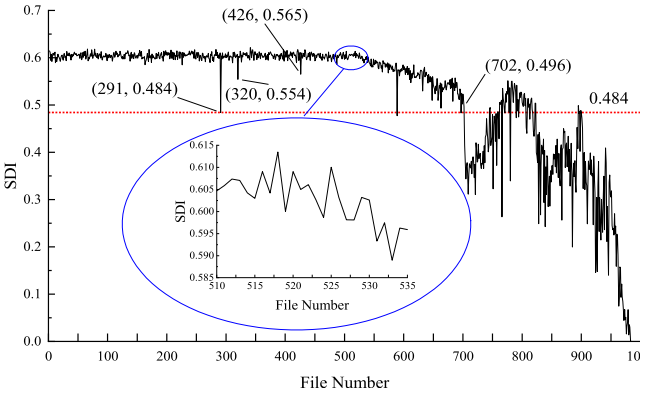

FIGURE 3. The performance degradation trend graph of the feature value SDI extracted from the IMS Test 2 data set.

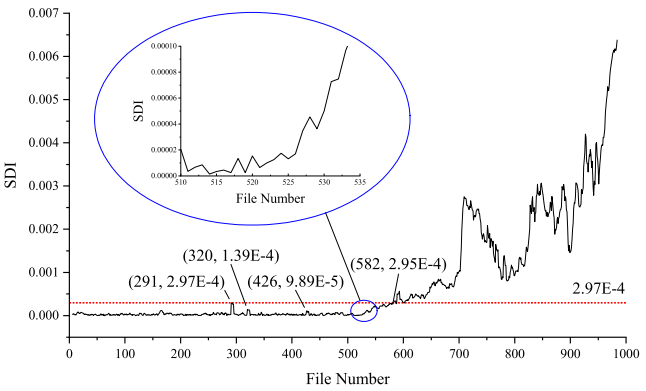

FIGURE 4. The performance degradation trend graph of the feature value SDIIF extracted from the IMS Test 2 data set.

large, which are greatly reduced after filtering. The SDI at the 291st data file prior to filtering is equivalent to the 702nd data file. After filtering, it is equivalent to the SDI at the 582nd data file. As shown in the partial enlarged views of Fig. 3 and Fig. 4, the fluctuation of the filtered data is significantly reduced. Moreover, the monotonicity near the incipient fault data files and the degradation trend are improved. Lastly, the anti-interference ability of the SVDD for abnormal samples and sensitivity to incipient failure samples are improved.

5) THE REAL-TIME TF IS CALCULATED VIA CSW

The SDI data of length N is taken as an example, where CSW of length $LEN = M + N = 50(M = 45, N = 5)$ is considered. The CSW slides from the left (the first real-time SDI) to the right (the last real-time SDI). The real-time spectral distance index trend factor (SDITF) is obtained successively. The process is shown in Fig. 5.

As shown in Fig. 5(a), there are 45 SDI data in the fixed array, while there are 5 SDI data in the active array. L1TF (adaptive calculation $\boldsymbol{\lambda}$) is performed on 50 SDI data in the fixed array and the active array. The SDITF is obtained by calculating the absolute value of derivative of the fitted line segment, which corresponds to the No. 5 real-time SDI data. Then, the SDITF value is recorded as k_5.

As shown in Fig 5(b), the CSW moves one piece of data to the right. The first data on the leftmost side of the active

array in the CSW is removed. The remaining 2^{nd} $\sim 5^{th}$ data and the No. 6 data are utilized to compose a new active array. The new 50 SDI data in the sliding window are filtered via adaptive L1TF. The SDITF, corresponding to the No.6 real-time SDI data, is obtained by calculating the absolute derivative fitted line segment. The SDITF is recorded as k_6.

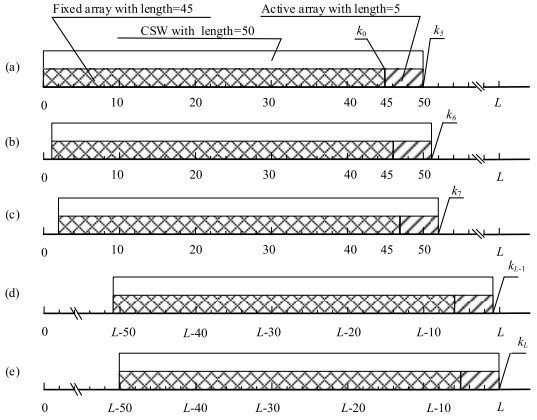

FIGURE 5. Schematic diagram of real-time TF calculation for the forward movement of CSW.

As shown in Fig. 5(c)∼5(e), the CSW slides to the No. $\boldsymbol{L}$ real-time SDI data. 50 new SDI data in the CSW are filtered via adaptive L1TF. The SDITF, corresponding to the No. $\boldsymbol{L}$ real-time SDI data is obtained by calculating the absolute derivative fitted line segment value. The SDITF is recorded as k_L.

In the process of utilizing IL1TF to calculate the TF, when the CSW has an occasional abnormal data file with relatively large amplitude, the calculated result can reduce the influence of the abnormal value. However, the influence of the abnormal value cannot be completely eliminated. When the CSW gradually passes the outlier, it will return to normal. Therefore, theoretically, a single abnormal value can only affect the same number of sample data files as the active window length $\boldsymbol{N}$. Therefore, the influence of incipient false alarm data files can be eliminated by judging whether to continuously alarm $\boldsymbol{N}$ times during the online detection. This provides theoretical support for eliminating the interference of false alarm data files during online monitoring.

The fixed array length in the CSW is relatively long. This results in insensitivity of incipient fault data files detection, and does not increase the degradation trend near the incipient fault data files. The active array in the combined sliding window is significantly long, which causes the CSW to contain abnormal data files for a relatively long time. It will cause the threshold limit to be exceeded for a relatively long time, which is not conducive to online judgment of incipient faults. In addition, too long the window length will increase the calculation time. After weighing the above advantages and disadvantages, the ratio of the fixed array length $\boldsymbol{M}$ to the active array $\boldsymbol{N}$ length in the CSW is 9:1, and the length of CSW is 50.

Most data processing methods, such as L1TF, or H-P trend filtering, can only meet the requirement of either solving the problem of false alarms caused by abnormal values or increasing the deterioration trend near the incipient failure data files. However, they cannot achieve both requirements at the same time. The IL1TF proposed in this paper operates from the perspective of continuous short time series. It overcomes the limitation of a single data file focus, and solves the aforementioned problems via comparison with the normal operating condition data files.

C. DETERMINATION OF SVDD MODEL PARAMETERS

As mentioned in Section 1, the performance of the SVDD depends to a large extent on the choice of $\boldsymbol{\sigma}$. According to the derivation process of the SVDD, the interface of the SVDD is highly dependent on the kernel function. The Gaussian kernel bandwidth $\boldsymbol{\sigma}$ is the only adjustable parameter, which determines the complexity of the SVDD interface. When $\boldsymbol{\sigma}$ decreases, the number of support vectors increases. Thus, the interface is closer to the one of the training samples. This, in turn, improves the description accuracy, but simultaneously causes the model to be oversensitive and overfitting. When $\boldsymbol{\sigma}$ is relatively large, a significant difference is observed between the interface and the training samples. This results in a decrease in accuracy, which is called the problem of underfitting. In this paper, a method for approximate determination of $\boldsymbol{\sigma}$ based on the laboratory IMS data is proposed.

By combining $K(x_i, x_i) = 1$ and Eq. (6), the matrix form of Eq. (7) is expressed as Eq. (21). The kernel matrix $\boldsymbol{K}$ in Eq. (21) is a diagonal matrix with a diagonal of 1. The kernel matrix consists of training samples kernel functions. Since the matrix is a diagonal one, only upper or lower triangular matrix needs to be discussed. The lower triangular matrix is defined as $\boldsymbol{K}_{\text{tu}}$. According to Eq. (21), the information obtained via SVDD in the training process is comprised in the $\boldsymbol{K}$ matrix. The value of the kernel function in the $\boldsymbol{K}$ matrix affects the solution of Eq. (7). Thus, it affects the establishment of the hypersphere. Hence, if kernel matrix with excellent performance can be constructed, the hypersphere performance will also be favorable:

$$L = \alpha^T K \alpha - 1 \tag{21}$$

where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_l]^T$.

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_l) \\ \vdots & \ddots & \vdots \\ k(x_l, x_1) & \cdots & k(x_l, x_l) \end{bmatrix} \tag{22}$$

Evangelista *et al.* [27] suggested that if the $\boldsymbol{K}$ matrix conforms to a uniform distribution, the solution α of Eq. (7) has better sparsity. The variance mean ratio (VMR) is defined as a standard measure of the probability distribution dispersion. Therefore, Eq. (23) can adequately describe the dispersion of values in $\boldsymbol{K}_{\text{tu}}$, while simultaneously eliminating the influence of the data magnitude. However, when $\boldsymbol{\sigma}$ is relatively small, it causes a larger value of $C_V(\sigma)$, while the elements in $\boldsymbol{K}_{\text{tu}}$ are more concentrated. Specifically, $C_V(\sigma)$ cannot reflect the original level of dispersion. Therefore, as shown in Eq. (24),

Evangelista *et al.* proposed a new equation to characterize divergence based on Eq. (23). The square of the variance and the denominator are introduced into $\boldsymbol{\varepsilon}$ to avoid the nullification of the denominator. This contributes to the practicality of the method. However, determination of $\boldsymbol{\varepsilon}$ value has emerged as a new problem. Liu *et al.* [30] proposed the variance-square root mean-ratio (VSRMR) method based on Evangelista *et al.* The formula of this method is shown in Eq. (25).

$$C_V(\sigma) = \frac{s(\sigma)}{m(\sigma)} \tag{23}$$

where $\boldsymbol{s}$ is the standard deviation and $\boldsymbol{m}$ is the mean value.

$$\mathrm{g}(\sigma) = \frac{s^2(\sigma)}{m(\sigma) + \varepsilon} \tag{24}$$

where ε is a positive value as it approaches 0.

$$f(\sigma) = \frac{s^2(\sigma)}{\sqrt{m(\sigma)}} \tag{25}$$

As previously mentioned, relevant investigations focused primarily on selection of the appropriate $\boldsymbol{\sigma}$ values to ensure favorable dispersibility of $\boldsymbol{K}$. At the same time, existing investigations gradually improved the proposed method to avoid relatively small values of $\boldsymbol{\sigma}$, resulting in the false appearance of good dispersion. However, in this paper, the TF calculated via IL1TF is applied to the extracted feature value, which is mostly in the range of $10^{-4} \sim 10^{-5}$. The method of Liu *et al.* still could not handle feature values of this order of magnitude. If elements in the K_{tu} are infinitely close to 1 due to $\boldsymbol{\sigma}$ being too large, poor dispersion of the results is observed (Eq. 8, Eq. 23) Therefore, the optimal value range of $\boldsymbol{\sigma}$ should be close to the value range of the TF in the input SVDD. Due to the value of the TF being relatively small, the optimal value of $\boldsymbol{\sigma}$ behaves in a similar manner, i.e., it exceeds the calculation range of the above square. If input feature value of SDITF and the MDETF are considered as an example, the $f(\boldsymbol{\sigma})$- $\boldsymbol{\sigma}$ curve is monotonously decreasing in the $10^{-3} \sim 10^{-5}$ range. Thus, the optimal value of $\boldsymbol{\sigma}$ is equal to 10^{-3}. As indicated in Fig. 6, the solid line represents the description boundary of the hypersphere on the two-dimensional plane. For $\boldsymbol{\sigma} = 10^{-5}$, all samples become support vectors, and serious overfitting problems occur. For $\boldsymbol{\sigma} = 10^{-3}$, the description boundary is quite smooth and under-fitting occurs. For $\boldsymbol{\sigma} = 0.00019$, good fit between the description boundary and the sample is observed. This means that the value of $\boldsymbol{\sigma}$ is in the range of $10^{-3} \sim 10^{-5}$, and the dispersion of the $\boldsymbol{K}$ matrix is not monotonically decreasing. Therefore, the method of Liu *et al.* is not suitable for the case where the feature values of the input SVDD are particularly small.

Regarding TF values being too small to determine $\boldsymbol{\sigma}$, the feature values of the SVDD input as SDITF and MDETF are taken as examples. First, the SDITF and the MDETF values are randomly extracted out of the 6^{th} to 300^{th} data file under normal operation of the IMS. Test 2 data is set as the input feature values of the SVDD, and the appropriate $\boldsymbol{\sigma}$ is trained via cross-validation method. As shown in Fig. 6(b),

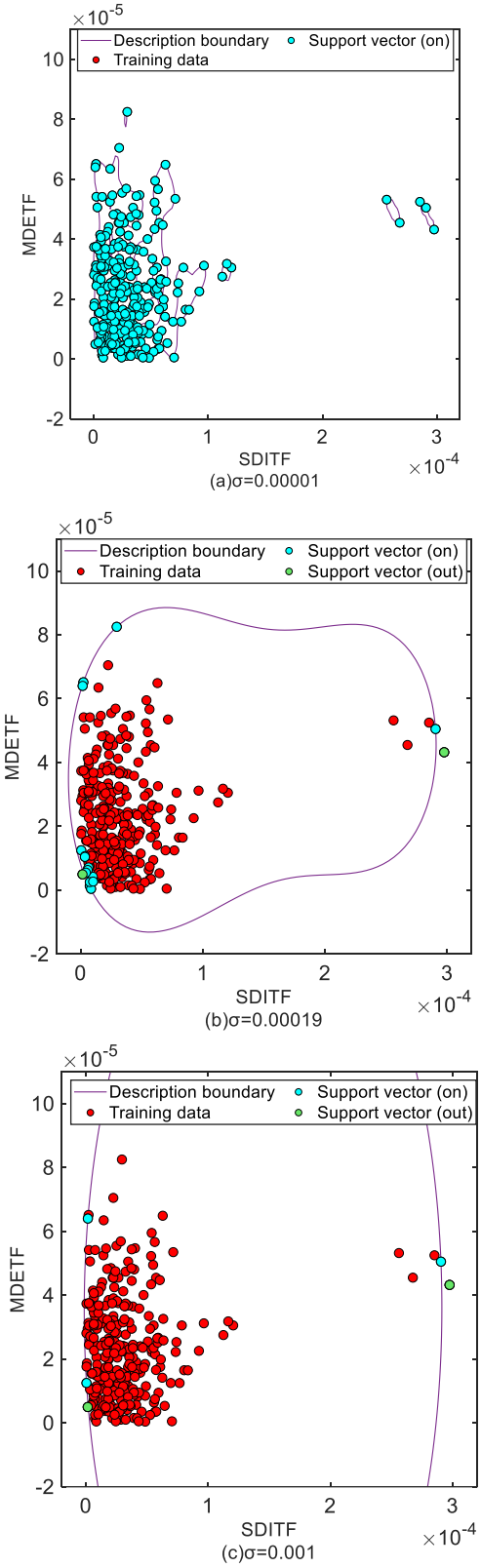

FIGURE 6. Hypersphere description boundary of different σ.

when $\boldsymbol{\sigma} = 0.00019$, the performance of the hypersphere is favorable. Thus, $\boldsymbol{\sigma} = 0.00019$ is taken as the reference value. Then, the value of $\boldsymbol{\sigma}$ is modified according to the

characteristics of the target data. The specific process is as follows:

Step 1: Using the normal operating conditions data of IMS Test 2 as the reference data set, the spatial distance matrix $\boldsymbol{H}$ between SDITF and the MDETF of the reference data set and the target data set is calculated. As shown in Eq. (26), the kernel function matrix $\boldsymbol{K}$ can be calculated by $\boldsymbol{H}$ and $\boldsymbol{\sigma}$. Parameter $\boldsymbol{H}$ is a diagonal matrix with a diagonal consisting of only 0 values. Therefore, only the upper triangular matrix or lower triangular matrix need to be considered. The lower triangular matrix of the reference data set and the target data set spatial distance matrix are denoted as $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, and the kernel function matrix of the reference data set and the target data set are denoted as $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$, respectively.

$$H = \begin{pmatrix} ||X_1 - X_1|| & \dots & ||X_1 - X_l|| \\ \vdots & \ddots & \vdots \\ ||X_l - X_1|| & \cdots & ||X_l - X_l|| \end{pmatrix},$$
$$K = \exp(-\left|\frac{H}{\sigma}\right|) \tag{26}$$

Step 2: Kernel density estimation is a non-parametric method used to estimate the probability density function of variables. The specific calculation process can be found in [46]. In this paper, the kernel density estimation method is used to calculate the distribution of data for $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. The estimated values corresponding to the maximum probability density for $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are denoted as e_1 and e_2, respectively. e_1 and e_2 will be used to represent $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$.

Step 3: The kernel function bandwidth σ_1 of the target data set is calculated according to Eq. (27). For $\sigma_1 = 0.00019$, favorable hypersphere performance of the reference data set is observed. It can be considered that the kernel matrix $\boldsymbol{K}_1$ diverges. Consequently, it is expected that the kernel matrix $\boldsymbol{K}_2$, calculated by the target data set, will also diverge. Therefore, assuming $\boldsymbol{K}_1 = \boldsymbol{K}_2$ and combined with Eq. (26), Eq. (27) is obtained:

$$\sigma_2 = \sigma_1 \frac{e_2}{e_1} \tag{27}$$

This represents an approximation method for determining the kernel function bandwidth of the SVDD based on the perspective of engineering online applications. Compared with the method that Liu and Evangelista determined the kernel function bandwidth by analyzing the dispersion of the kernel matrix K with different kernel function bandwidth, the advantages of the algorithm proposed in this paper are fast calculation speed, simple and no non-convergence phenomenon. Experimental validation results in Sections 4 and 5 indicate that this method can be applied to various equipment and feature values.

D. PROPOSED IFD METHOD

The specific algorithm of the IFD method based on IL1TF and the SVDD proposed in this paper is as follows and the functional schematic diagram is presented in Fig. 7.

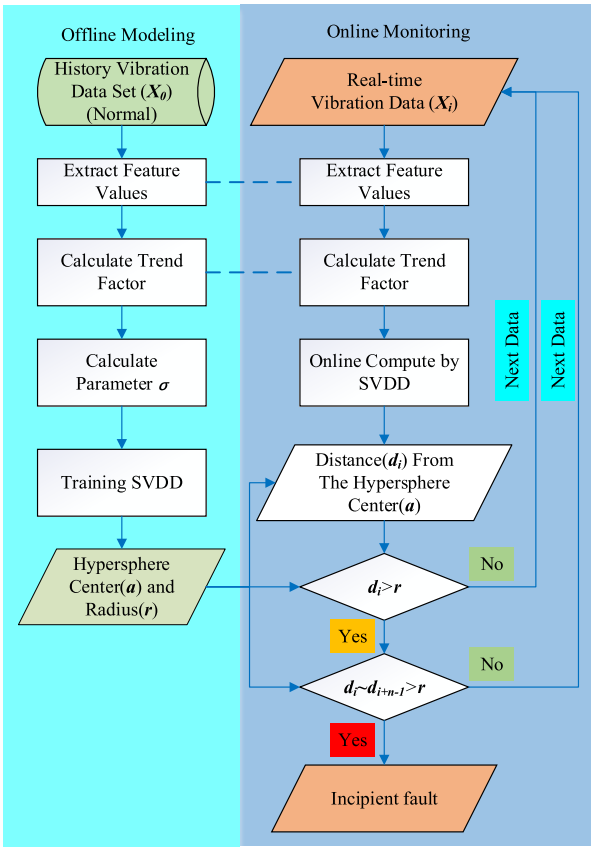

FIGURE 7. IFD functional framework diagram with two working modes of offline training and online monitoring.

IV. EXPERIMENTAL VALIDATION

A. LABORATORY DATA VALIDATION

In order to validate the feasibility of the IFD method proposed in this paper, public IMS Test 2 bearing data set is employed as the model training data. According to Fig. 8, four ZA-2115 double-row bearings are mounted on the main shaft of the experimental platform. The spring mechanism applies a radial load of 26689 N to bearings 2 and 3. Meanwhile, two PCB 353B33 acceleration sensors are mounted on each bearing along the horizontal and vertical axes. NIDAQ card 6062E with a sampling rate of 20 kHz is utilized to collect 20480 points of vibration data every 10 minutes. It is observed that bearing 1 failed on the outer ring after exceeding its design life. During the normal to failure test, 984 data files were collected from February 12, 2004 until February 19, 2004, with each file being composed of 20,480 points for a recording interval of 10 minutes. Following the run-to-failure test, cracks are observed at the outer ring of bearing 1.

By considering the IMS laboratory data as an example, the IFD method data processing for bearings proposed in this paper can be explained:

1) First 300 data files of No. 1 bearing are utilized for envelope spectrum analysis to confirm the fault-free characteristic frequency. Then the SDITF and the MDETF are extracted.

2) Eq. (27) is used to calculate the bandwidth of the kernel function. The SDITF and the MDETF are inputted into the

Algorithm 1 IFD Algorithm

Input: X_0 // Historical raw vibration data set
X_i // Real-time raw vibration data
Output: IFD information (no incipient fault is detected; no information is given)
Step 1: Extracting SDI and MDE from the X_0 data set.
Step 2: Calculating the SDITF and MDETF by IL1TF.
Step 3: Calculating of kernel function bandwidth σ by Eq. (27).
Step 4: SDITF and MDETF, calculated in the Step 2, are inputted for training into the SVDD with the kernel function bandwidth σ calculated in the Step 3 to obtain the center $\boldsymbol{a}$ and radius $\boldsymbol{r}$ of the hypersphere.
Step 5: Extracting SDI and MDE from the X_i, and calculating the SDITF and MDETF by IL1TF.
Step 6: Calculating of the generalized distance $\boldsymbol{d_i}$ by the SVDD.
Step 7: Online comparison of $\boldsymbol{d_i}$ and $\boldsymbol{r}$ is made. If $\boldsymbol{d_i} > \boldsymbol{r}$, a suspected incipient failure sample is detected. If the number of consecutive alarms exceeds the length of CSW active arrays, it is denoted as an incipient fault sample.

TABLE 1. The serial number of the suspicious fault data file calculated by IFD based on the IMS Test 2 data set.

43	210	291	292	318	360	444	445
446	447	457	494	495	496	497	498
521	529	532	533	534	535	536	...

SVDD model to calculate the center $\boldsymbol{a}$ and radius $\boldsymbol{r}$ of the hypersphere.

3) The active window length of the combined sliding window is selected as 5. The SDI calculation result of the No.1 data file is removed. L1TF is conducted for the No.6 data file. The SDITF and the MDETF of No.6 data file are calculated.

4) Above-calculated two trend factors are placed into the trained SVDD model. The distance $\boldsymbol{d}$ from the center $\boldsymbol{a}$ of the hypersphere is calculated and compared against $\boldsymbol{a}$ and $\boldsymbol{d}$ distance calculated from the known normal data. In this way, incipient fault location is determined.

5) Steps 3 and 4 are repeated for all data files starting from No.7 and ending with No.984.

6) Distance curve is drawn from the No.6 to No.984 data files, as shown in Fig. 9. Points where distance $\boldsymbol{d}$ exceeds the recorded radius are shown in Table 1.

According to Table 1 and Fig. 9, events that exceed the dynamic self-learning threshold line have occurred at the following data files: No.43, No.210, No.291, No.318, No.360, No.444, No.457, No.494, No.521, No.529, and No.532. However, starting with data file No. 532, the calculated radius value continuously exceeds the self-learning threshold line. The time at file No.532 is defined as the incipient fault detection moment. By applying the Matlab R2018b software, calculating an IMS Test 2 data file lasts 1.271 seconds. The operation environment of the application computer is configured as I5-8300h CPU, 8G RAM memory, and Windows 10 operating system.

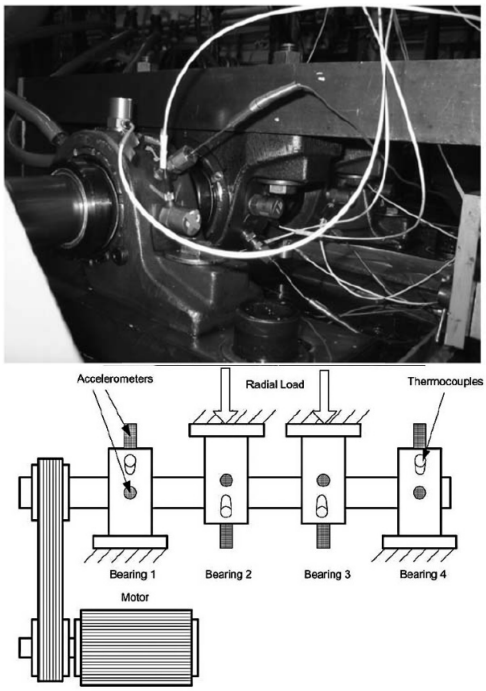

FIGURE 8. IMS bearing test bench and sensor layout [45].

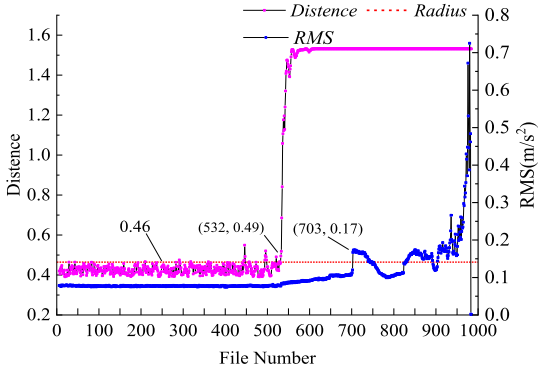

FIGURE 9. Distance and RMS curve calculated based on the IMS Test 2 data set.

Based on fault mechanism analysis, researchers have carried out many investigations on the bearing incipient fault detection via method of fault characteristic frequency extraction signal analysis. Through application of Gabor wavelet coefficient squared envelope spectrum analysis method, Wang *et al.* [47] detected the incipient fault occurrence at the No.533 data file. Using an improved empirical mode decomposition method to extract the incipient fault characteristic frequency, Fan *et al.* [7] detected the incipient fault occurrence at the No.532 data file.

Data-driven incipient fault detection methods have been extensively studied. Zhu *et al.* [48] employed the fruit fly optimization algorithm (FOA) to optimize the SVDD parameters and to establish the FOA-SVDD model. The authors detected the incipient fault at the No.710 data file, which is 1780 minutes later than the method proposed in this article. Using the method of combining the hierarchical Dirichlet

process and the continuous hidden Markov model, Wang *et al.* [49] identified the incipient fault occurrence at the No.576 data file, which is 440 minutes later than the method proposed in this paper. Using wavelet packet energy entropy and RBF neural network, Zhou *et al.* [13] detected the incipient fault in No.533 data file, which is 10 minutes later than the method proposed in this paper. As shown in Fig. 9, RMS did not change significantly until the 703rd data file, which is 1710 minutes later than the method proposed in this paper.

B. ENGINEERING CASE VERIFICATION

In engineering applications, bearings operate in environments where rotational speed and lubricating oil parameters vary, shaft alignment may be poor, and axial clearance may be inappropriate. Data acquisition loops, which include parts such as the signal cable or the data collector, can be susceptible to electromagnetic interference. Laboratory data validated that the constructed model and the method proposed in this paper may not be representative. In order to validate this model in engineering applications, a set of bearing vibration data collected in the actual production process of the factory is used to verify the practicability of the proposed method.

A failure case data set of a P3409A centrifugal pump from a hydrocracking production unit of a Chinese petroleum company is selected as the model training data set and validation data set [50]. As indicated in Fig. 10, two deep groove ball type 6217 bearings are mounted on the shaft of the centrifugal pump, which operates at 2980 rpm Two PCB EXM603C01 High Sensitivity Quartz ICP accelerometers are mounted on the left bearing housing (one horizontal X and one axial Z), and one PCB EXM603C01 High Sensitivity Quartz ICP accelerometer is mounted on the right bearing housing (horizontal X). Raw acceleration data was continuously obtained from 24:00 on December 15, 2013, to 16:00 on January 12, 2014. A total of 332 data files were obtained, with the sampling interval of each data file being two hours. The sampling rate was 25.6 kHz, and each data file was composed of 16,384 points. Cracks were found in the outer ring of the left end bearing during the shutdown phase. As demonstrated in Fig. 11, the acceleration effective value curve is used to characterize the bearing performance degradation curve within a specific period of time. For the No.226 bearing data file, the acceleration effective value exceeds the vibration high alarm line, and an operation warning signal is issued. For the No.329 bearing data file, the acceleration effective value exceeds the vibration high-high alarm line, and an operation danger alarm signal is issued.

Using the IFD model constructed in this paper, the data set starting from file No.6 and ending with the file No.96 are defined as the normal operation data, which are utilized for the model training. Following the steps according to the Section 4.1, the kernel function bandwidth is calculated by Eq. (27) as 0.00061, and the trained SVDD hypersphere boundary is drawn (Fig. 12). According to Fig. 12, the number of support vectors in the SVDD hypersphere is moderate,

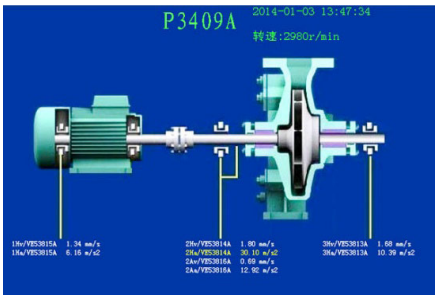

FIGURE 10. Centrifugal Pump measuring point information and sensor layout [50].

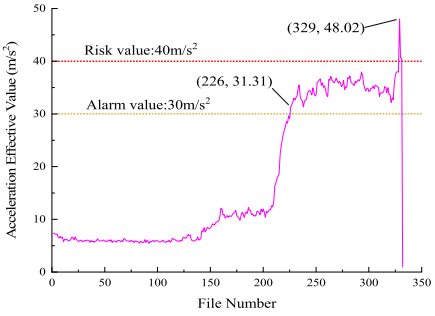

FIGURE 11. The effective value of the acceleration of the left bearing of the centrifugal pump and its fixed threshold alarm line graph.

TABLE 2. The serial number of the suspicious fault data file calculated by IFD based on the engineering data set.

6	70	79	81	107	108	115	116
117	118	119	121	127	128	129	132
133	134	142	143	144	145	146	...

TABLE 3. Fault characteristic frequency of 6217 bearing.

Failure mechanism	characteristic frequency (Hz)
inner ring defect	342.32
outer ring defect	207.67
rolling defect	96
cage defect	18.88

and boundary of the hypersphere can accurately describe the training samples without overfitting.

According to data provided in Table 2 and Fig. 13, the events that exceed the dynamic self-learning threshold line have occurred for the following data files: No.6, No.70, No.79, No.81, No.107, No.115, No.121, No.127, No.132, and No.142. However, starting with data file No. 142, the calculated radius value continuously exceeds the self-learning threshold line. The time at data file No.142 is defined as the incipient fault detection moment. As shown in Fig.14, the envelope spectrum analysis is conducted separately on No.141 and No.142 data files to demonstrate the accuracy of the IFD. Envelope spectrum analysis indicates that characteristic frequencies of the outer race fault are 204.7 Hz and 409.4 Hz. By observing the fault characteristic frequency of 6217 bearing calculated in Table 3, it can be inferred that the outer ring of the bearing has an early fault.

The IFD model proposed in this paper can quickly, accurately, and stably detect the incipient faults of centrifugal pump bearings. Furthermore, it can avoid ''insufficient'' and

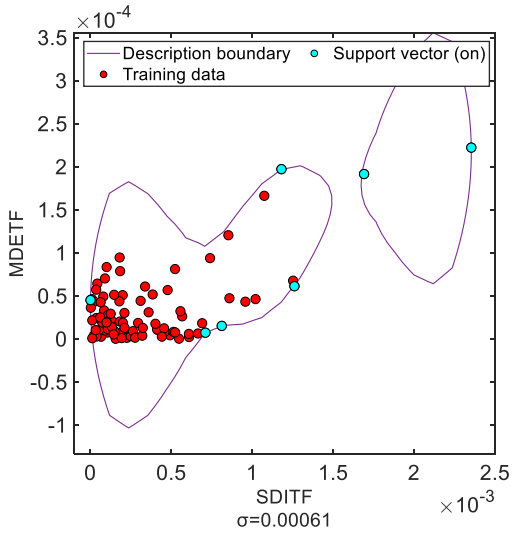

FIGURE 12. Hypersphere description boundary trained by engineering data.

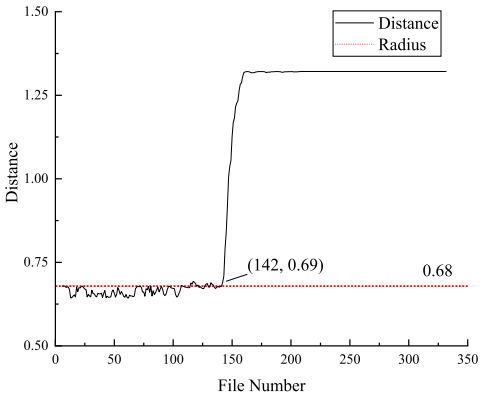

FIGURE 13. The distance curve calculated by IFD based on engineering data.

"excessive" alarms of the same event. Once the equipment component incipient performance degradation is confirmed, the distance curve that characterizes the incipient fault does not show repeated fluctuations. Via utilization of the IFD method proposed in this paper, the incipient fault is detected for No.142 data file, which is 8 days and 6 hours earlier than the traditional method for No.226 fault detection. Similar to the calculation environment in Section 4.1, the average time for calculating a data file is 0.778 seconds, which is significantly lower than the calculation time of 2 hours. This meets the time requirement of online incipient fault detection for engineering application.

V. COMPARISONS

Several data-driven models, such as the Local Outlier Factor (LOF) [51] model, One-class SVM (OCSVM) [52] model and SVDD [17] model are applied to detect the incipient fault occurrence time using the same IMS Test 2 data set. One-class SVM maps the trained data to the high-dimensional feature space via kernel function. Optimal hyperplane in the feature space is solved to achieve the maximum separation between the target data and the coordinate origin. LOF evaluates

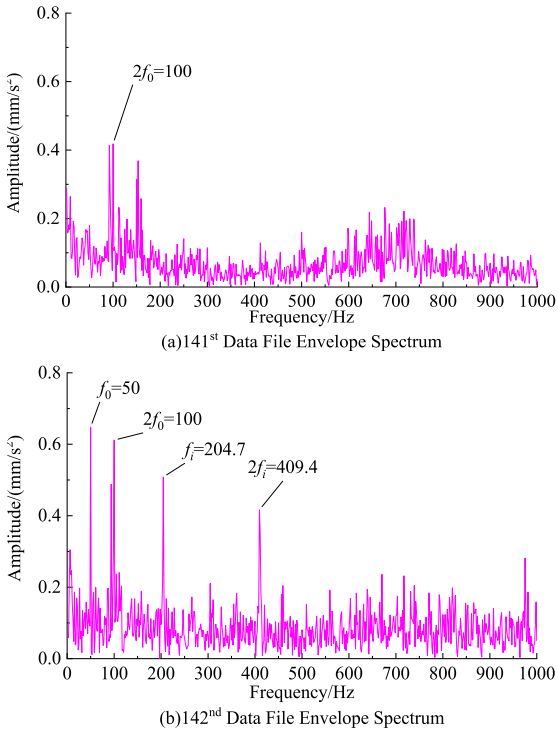

FIGURE 14. Comparison chart of vibration signal envelope spectrum analysis at data file 141 and 142.

whether the target data is an outlier by comparing the density of each point $\boldsymbol{p}$ with the density of the neighboring points. The density ratio of $\boldsymbol{p}$ to the average density of the neighboring points is designated as the LOF. Whether or not $\boldsymbol{p}$ is an abnormal sample can be directly evaluated by comparing the size of the LOF. In addition, two typical time-domain feature values, such as RMS and kurtosis (KU) as the input of IFD model, are compared with SDI and MDE as the input of IFD model.

In order to avoid the numerical difference effect of these characteristic parameters, the maximum and minimum normalization method is used for pretreating. Twelve incipient fault detection models are obtained by cross combination of the feature values. The performance of aforementioned IFD models is compared, such as the starting data file of initial fault detection, whether it relies on prior knowledge, whether there are repeated alarms for the same event and the calculating time required for one data file under same software and hardware operating environment. The comparison results are shown in Table 4 and Fig.16. The IFD method formulated in this paper is denoted as the ILSVDD model.

No. 6 data file to No. 300 data file of the IMS test 2 are used as the normal status samples for the model training. The run-to-failure data files of the IMS test 2, which do not include the file No. 1 data file, are selected as testing data files for the model verification.

The Gaussian kernel bandwidth of the SVDD is calculated via Eq. (27). As shown in Fig. 15(a) and (b), the solid lines are hyperspheres trained by the root mean square trend factor (RMSTF) and the kurtosis trend factor (KUTF), with

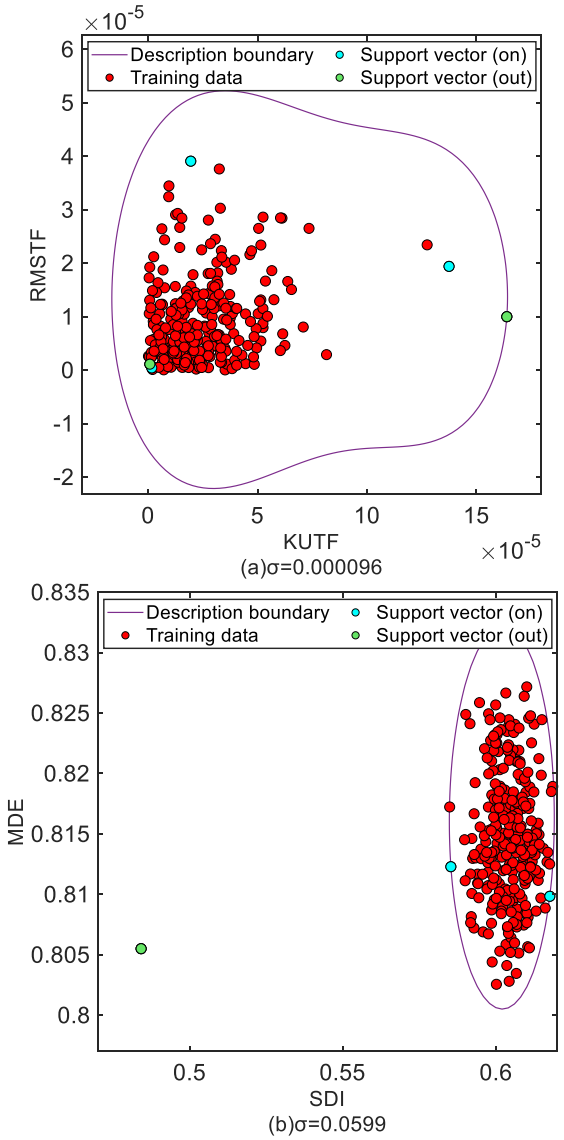

FIGURE 15. Hypersphere description boundary of different feature values.

relatively small numerical magnitude of the feature value. The SDI and MDE have relatively large numerical magnitude of the feature value, while none of them have over-fitting and under-fitting problems. In addition, the proposed hypersphere calculation method performs well in the engineering data (Fig. 12). Therefore, it can be concluded that the gaussian kernel function bandwidth calculation method proposed in this paper is applicable to various devices and eigenvalues, i.e., it can be applied either when magnitude of the eigenvalues is small or large.

The heterogeneous sample ratio (HSR) $\boldsymbol{k}$ is set to 0.01 for the OCSVM model. The value $\boldsymbol{k'}$, which represents the number of neighbors (NN), is set to 10. Lastly, the LOF model threshold is set to 2.

The online alarm criteria are defined as follows: once five consecutive samples are identified as abnormal, the first sample of the five consecutive samples is regarded as the

TABLE 4. A comparison table of the start time of performance degradation detected by several IFD methods and whether the prior knowledge of external experts is required and the calculating time required for one data file.

No.	IFD method	IFD moment	Relying on prior knowledge or not?	Calculating time
1	(RMSTF and KUTF) + OCSVM	534	Yes (HSR)	0.023
2	(RMSTF and KUTF) + LOF	534	Yes (NN and threshold)	0.028
3	(RMSTF and KUTF) + SVDD	534	No	0.024
4	(SDITF and MDETF) + OCSVM	528	Yes (HSR)	1.271
5	(SDITF and MDETF) + LOF	531	Yes (NN and threshold)	1.275
6	ILSVDD	532	No	1.271
7	(RMS and KU) + OCSVM	533	Yes (HSR)	0.001
8	(RMS and KU) + LOF	533	Yes (NN and threshold)	0.006
9	(RMS and KU) + SVDD	533	No	0.002
10	(SDI and MDE) + OCSVM	533	Yes (HSR)	1.239
11	(SDI and MDE) + LOF	533	Yes (NN and threshold)	1.242
12	(SDI and MDE) + SVDD	533	No	1.238

incipient fault data files If there are false and repeated alarm problems, and there is no reasonable online detection strategy, the problems cannot be solved by the selected IFD model.

As shown in Table 4, the ILSVD method proposed in this paper is the only one that can accurately identify the incipient fault at No.532 data file without depending on the prior knowledge. However, it falls behind the method No.4, which can detect the incipient fault at No. 528 data file, and the method No.5, which can detect the incipient fault at No. 531 data file.

By comparison with the method which utilizes IL1IF in Fig. 9 and the method without IL1TF in Fig. 16(a), it can be seen that IL1TF can greatly reduce the amplitude of the abnormal data fluctuation at data files No.291 and No.320. Once the incipient failure of the equipment performance degradation is confirmed, performance degradation trend line obtained by the former method is relatively smooth and without fluctuations. On the contrary, performance degradation trend line obtained by the latter method fluctuates violently. Since the bearing performance degradation trend is generally irreversible, the performance degradation trend line with a large fluctuation trend is somewhat different from the actual situation.

As shown in Fig. 9 and Fig. 16 (a)∼ Fig.16 (c), when the input feature values of the trained SVDD model are different, the incipient fault is detected at data files No.532, 533, 534 and 533. For the same input values, a smoother performance degradation curve can be obtained if IL1TF is applied, which can reflect the equipment performance degradation state more accurately.

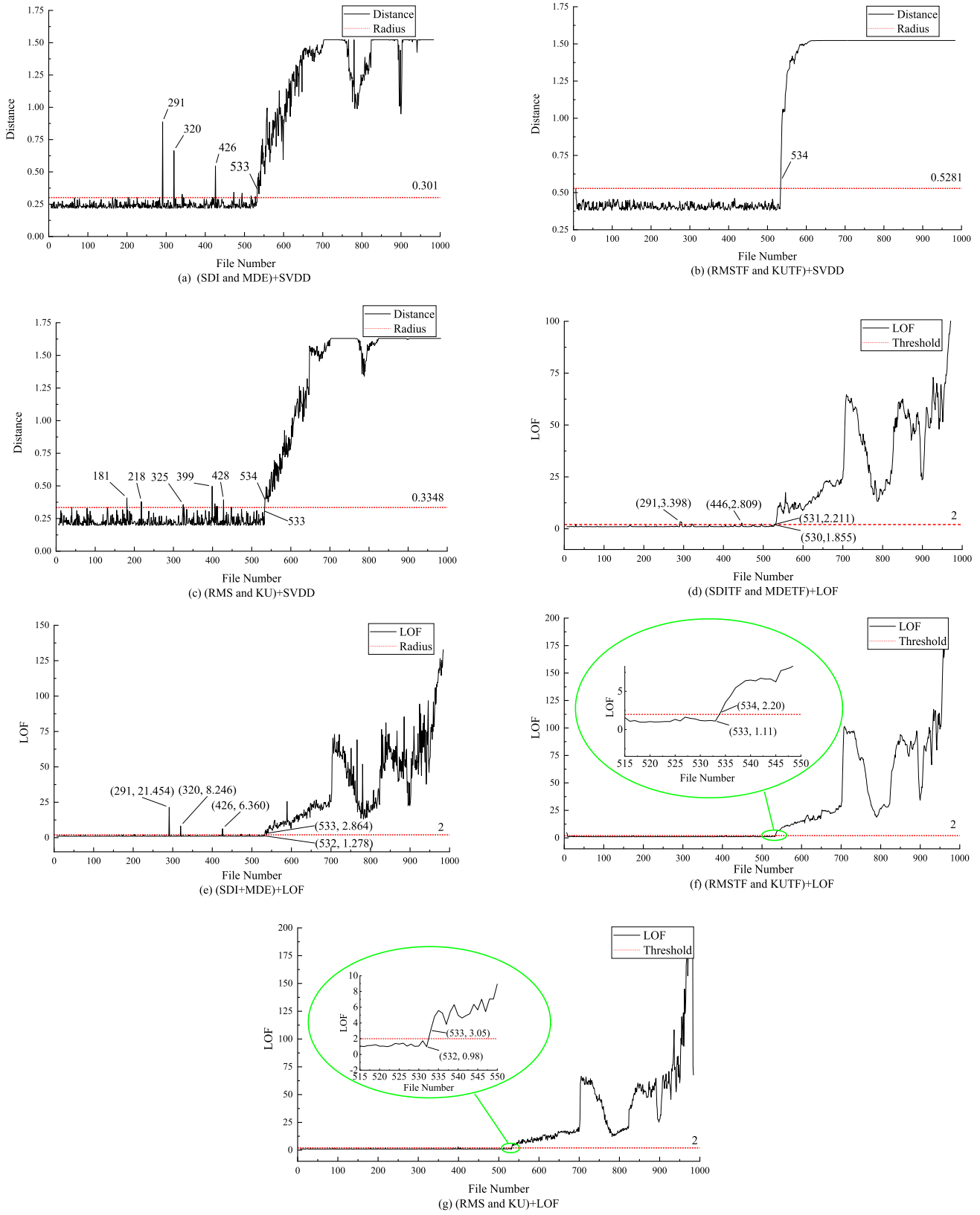

FIGURE 16. Comparison of the starting time of bearing performance degradation detected by different incipient fault detection models based on IMS Test 2 data set.

According to Fig. 9 and Fig. 16, when the model input feature values are the same, the detection sensitivity of the SVDD is higher than that of the LOF. The SVDD incipient fault detection model can be used online without depending on prior knowledge or experience from the outside experts. Regarding model inputs, SDI and MDE are more sensitive than RMS and KU in the incipient fault monitoring model. Based on the observed data, it can be concluded that the SVDD incipient fault detection model which employs IL1TF can effectively reduce the occurrence of repeated alarms.

VI. CONCLUSION

In this paper, a data-driven IFD model was constructed and validated/verified by experimental/engineering case data. The following conclusions are made:

1) The IFD model, which is built based on the IL1TF and the SVDD technology, can be regarded as a "black box". It does not rely on prior knowledge from the outside experts, real-time raw vibration data is required as its input, while the incipient fault warning message is provided as its output.

2) A method for calculating TF is proposed, which replaces the traditional feature values as the input of the SVDD model and increases the distinguishability between the normal status and incipient fault status.

3) Based on the spatial distance matrix of the target data trend factor and the reference data trend factor spatial distance matrix, a method for linear modification of the SVDD kernel function bandwidth is proposed. In this method, there is no need to repeatedly demonstrate the divergence of the kernel matrix and to train a moderately sized hypersphere. Difficulty of determining the bandwidth of the SVDD kernel function when the input feature value is too small has been resolved.

4) A method for determining the bandwidth of the SVDD kernel function is a linear approximation method. When the feature values are small in magnitude and differ significantly, the IFD accuracy is decreased.

The method proposed in this paper shows excellent performance in the IFD of rolling bearing performance degradation, but for the performance degradation of rotating machinery, such as rotor imbalance, rotor rubbing, axis misalignment, surge, oil film whirl and other faults, requires further verification; At the same time, the generalization of the model needs to be further verified and improved by using a large amount of engineering case data.

REFERENCES

[1] L. Cui, J. Huang, F. Zhang, and F. Chu, "HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault," *Mech. Syst. Signal Process.*, vol. 120, pp. 608–629, Apr. 2019, doi: 10.1016/j.ymssp.2018.09.043.

[2] A. Rai and S. H. Upadhyay, "An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence," *Appl. Soft Comput.*, vol. 71, pp. 36–50, Oct. 2018.

[3] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, Dec. 2020, Art. no. 108086, doi: 10.1016/j.measuremen.t.2020.108086.

[4] J. L. Gomez, I. Khelf, A. Bourdon, H. André, and D. Rémond, "Angular modeling of a rotating machine in non-stationary conditions: Application to monitoring bearing defects of wind turbines with instantaneous angular speed," *Mech. Mach. Theory*, vol. 136, pp. 27–51, Jun. 2019.

[5] H. T. Shi and X. T. Bai, "Model-based uneven loading condition monitoring of full ceramic ball bearings in starved lubrication," *Mech. Syst. Signal Process.*, vol. 139, May 2020, Art. no. 106583.

[6] H. Li, R. Yang, C. Wang, and C. He, "Investigation on planetary bearing weak fault diagnosis based on a fault model and improved wavelet ridge," *Energies*, vol. 11, no. 5, p. 1286, May 2018, doi: 10.3390/en11051286.

[7] H. Shi, J. Guo, X. Bai, L. Guo, Z. Liu, and J. Sun, "Research on a nonlinear dynamic incipient fault detection method for rolling bearings," *Appl. Sci.*, vol. 10, no. 7, p. 2443, 2020, doi: 10.3390/app10072443.

[8] F. Jiang, Z. Zhu, and W. Li, "An improved VMD with empirical mode decomposition and its application in incipient fault detection of rolling bearing," *IEEE Access*, vol. 6, pp. 44483–44493, 2018.

[9] M. N. Albezzawy, M. G. A. Nassef, E. S. Elsayed, and A. Elkhatib, "Early rolling bearing fault detection using a Gini index guided adaptive Morlet wavelet filter," in *Proc. IEEE 10th Int. Conf. Mech. Aerosp. Eng. (ICMAE)*, Brussels, Belgium, Jul. 2019, pp. 314–322.

[10] H. Darong, K. Lanyan, M. Bo, Z. Ling, and S. Guoxi, "A new incipient fault diagnosis method combining improved RLS and LMD algorithm for rolling bearings with strong background noise," *IEEE Access*, vol. 6, pp. 26001–26010, 2018.

[11] T. Zan, Z. Liu, H. Wang, M. Wang, X. Gao, and Z. Pang, "Prediction of performance deterioration of rolling bearing based on JADE and PSO-SVM," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vols. 203–210, Aug. 2020, Art. no. 095440622095120, doi: 10.1177/0954406220951209.

[12] A. Rai and S. H. Upadhyay, "Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering," *Mech. Syst. Signal Process.*, vol. 93, pp. 16–29, Sep. 2017.

[13] J. Zhou, F. Wang, C. Zhang, L. Zhang, and P. Li, "Evaluation of rolling bearing performance degradation using wavelet packet energy entropy and RBF neural network," *Symmetry*, vol. 11, no. 8, p. 1064, 2019, doi: 10.3390/sym11081064.

[14] 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," *J. Manuf. Syst.*, vol. 55, pp. 179–198, Apr. 2020.

[15] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1679–1689, Jul. 2018.

[16] W. Yao and Y. Ding, "Smart city landscape design based on improved particle swarm optimization algorithm," *Complexity*, vol. 2020, pp. 1–10, Dec. 2020, doi: 10.1155/2020/6693411.

[17] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.

[18] 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 Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 443–456, Feb. 2020.

[19] C. Yang, W. Mao, Y. Liu, and S. Tian, "Incremental weighted support vector data description method for incipient fault detection of rolling bearing," in *Proc. Prognostics Syst. Health Manage. Conf.*, Qingdao, China, 2019, pp. 1–6.

[20] Y. Zhu, F. Wang, X. Wang, K. Yan, and X. Zhu, "Research on condition assessment method based on projection one-class classifier," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Harbin, China, Jul. 2017, pp. 121–126.

[21] B. Wang, H. Pan, and W. Yang, "Robust bearing degradation assessment method based on improved CVA," *IET Sci., Meas. Technol.*, vol. 11, no. 5, pp. 637–645, Aug. 2017.

[22] L. Zhang, F. Qiao, J. Wang, and X. Zhai, "Equipment health assessment based on improved incremental support vector data description," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Jun. 20, 2019, doi: 10.1109/TSMC.2019.2919468.

[23] C. Zhang, K. Peng, and J. Dong, "An incipient fault detection and self-learning identification method based on robust SVDD and RBM-PNN," *J. Process Control*, vol. 85, pp. 173–183, Jan. 2020.

[24] M. Cha, J. S. Kim, and J.-G. Baek, "Density weighted support vector data description," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3343–3350, Jun. 2014.

[25] D. Tax, A. Ypma, and R. Duin, "Support vector data description applied to machine vibration analysis," in *Proc. 5th Annu. Conf. Adv. School Comput. Imag.*, Heijen, The Netherlands, 1999, pp. 15–23.

[26] D. M. J. Tax and R. P. W. Duin, "Data domain description using support vectors," in *Proc. Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, 1999, pp. 251–256.
[27] P. F. Evangelista, M. J. Embrechts, and B. K. Szymanski, "Some properties of the Gaussian kernel for one class learning," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, Porto, Portugal. Berlin, Germany: Springer, Sep. 2007, pp. 269–278.
[28] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.
[29] H. Wang, L. Zhang, Y. Xiao, and W. Xu, "An approach to choosing Gaussian kernel parameter for one-class SVMs via tightness detecting," in *Proc. 4th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, New York, NY, USA, Aug. 2012, pp. 318–323.
[30] Z. Liu, J. Kang, X. Zhao, M. J. Zuo, Y. Qin, and L. Jia, "Modeling of the safe region based on support vector data description for health assessment of wheelset bearings," *Appl. Math. Model.*, vol. 73, pp. 19–39, Sep. 2019.
[31] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," *SIAM Rev.*, vol. 51, no. 2, pp. 339–360, 2009.
[32] B. G. Gowri and K. P. Soman, "Enhancement of white Gaussian noise affected speech using VMD-l1 trend filter method," *J. Intell. Fuzzy Syst.*, vol. 34, no. 3, pp. 1701–1711, 2018.
[33] S. Selvin, S. G. Ajay, B. G. Gowri, V. Sowmya, and K. P. Soman, "ℓ_1 trend filter for image denoising," *Procedia Comput. Sci.*, vol. 93, pp. 495–502, Jan. 2016.
[34] J. Ottersten, B. Wahlberg, and C. Rojas, "Accurate changing point detection for ℓ_1 mean filtering," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 297–301, Feb. 2016.
[35] Q. F. Wang, J. H. Liu, X. J. Liu, and X. J. Xu, "Data-driven performance degradation trend predicting method for the rotating equipment," *Comput. Integr. Manuf. Syst.*, pp. 1–17, Aug. 2020. [Online]. Available: https://kns.cnki.net/kcms/detail/11.5946.TP.20200817.0847.002.html
[36] Z. He, "Developments and thoughts on operational reliability assessment of mechanical equipment," *J. Mech. Eng.*, vol. 50, no. 2, p. 171, 2014.
[37] X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowl.-Based Syst.*, vol. 163, pp. 450–471, Jan. 2019.
[38] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016.
[39] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 2, Feb. 2005, Art. no. 021906.
[40] M. Costa and J. A. Healey, "Multiscale entropy analysis of complex heart rate dynamics: Discrimination of age and heart failure effects," in *Proc. Comput. Cardiol.*, New York, NY, USA, vol. 30, 2003, pp. 705–708.
[41] H. Azami, E. Kinney-Lang, A. Ebied, A. Fernandez, and J. Escudero, "Multiscale dispersion entropy for the regional analysis of resting-state magnetoencephalogram complexity in Alzheimer's disease," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, New York, NY, USA, Jul. 2017, pp. 3182–3185.
[42] K. Annen, R. J. Hodrick, E. C. Prescott, M. O. Ravn, and C. Zimmermann, "Postwar U.S. Business cycles: An empirical investigation," *J. Money Credit Banking*, vol. 29, no. 1, pp. 1–16, 1997.
[43] F. Bauer and M. A. Lukas, "Comparing parameter choice methods for regularization of ill-posed problems," *Math. Comput. Simul.*, vol. 81, no. 9, pp. 1795–1841, May 2011.
[44] L. Qin and B. C. Xie, "Regularized trend filtering with Lq penalty," *J. Quant. Tech. Econ.*, vol. 31, no. 5, pp. 151–161, 2014.
[45] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vib.*, vol. 289, nos. 4–5, pp. 1066–1090, Feb. 2006.
[46] R. T. Samuel and Y. Cao, "Nonlinear process fault detection and identification using kernel PCA and kernel density estimation," *Syst. Sci. Control Eng.*, vol. 4, no. 1, pp. 165–174, Jan. 2016.
[47] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.
[48] S. Zhu, R. L. Bai, and Q. C. Liu, "Rolling bearing performance degradation assessment based on FOA-WSVDD," *China Mech. Eng.*, vol. 29, no. 5, pp. 602–608, 2018.
[49] H. Wang, Y. Ji, L. B. Zhu, and X. Liu, "Performance degradation evaluation of mechanical equipment based on HDP-CHMM," *J. Vib., Meas. Diagnosis*, vol. 38, no. 4, pp. 733–737, 2018.
[50] Q. Wang, B. Wei, J. Liu, and W. Ma, "Data-driven incipient fault prediction for non-stationary and non-linear rotating systems: Methodology, model construction and application," *IEEE Access*, vol. 8, pp. 197134–197146, 2020.
[51] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 25, pp. 1–27, May 2011.
[52] R. Domingues, M. Filippone, P. Michiardi, and J. Zouaoui, "A comparative evaluation of outlier detection algorithms: Experiments and analyses," *Pattern Recognit.*, vol. 74, pp. 406–421, Feb. 2018.

QINGFENG WANG received the Ph.D. degree from the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, China, in 2011. He is currently an Associate Researcher with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. He is also a security expert of the State Administration of Work Safety of China. His research interests include electromechanical equipment monitoring, diagnosis and maintenance.

XIAOJIN LIU received the B.S. degree from the College of Mechanical Engineering, North China University of Science and Technology, China, in 2019, where he is currently pursuing the M.S. degree. His current research interests include incipient fault warning and health evaluation.

BINGKUN WEI received the B.S. degree from the College of Electrical Engineering, Beijing University of Chemical Technology, China, in 2018, where he is currently pursuing the M.S. degree. His current research interests include fault diagnosis and health evaluation.

WENWU CHEN is currently a Professorate Senior Engineer with the Sinopec Qingdao Research Institute of Safety Engineering, China. He is currently mainly engaged in the research of integrity management, fault diagnosis and prediction technology application in oil refining and chemical equipment.