This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

Updatable Online Learning Successive Difference Mode Decomposition for Rotating Machine Fault Diagnosis

Chao Teng, Zuogang Shang, Xuechun Bai, Ruqiang Yan, *Fellow, IEEE*, and Asoke K. Nandi, *Life Fellow, IEEE*

***Abstract*—Signal processing methods are widely used in fault diagnosis and are known for their strong interpretability. Among them, signal adaptive decomposition algorithms are used to extract the features of fault signals. As an effective adaptive decomposition algorithm, difference mode decomposition divides the signals into three components using spectrum weighting. However, it can only separate mixed fault components and is not suitable for multi-class fault diagnosis tasks. This paper presents a successive difference mode decomposition method. The reference component and concerned components (fault features) are defined based on the differences in faults. Then, the filters corresponding to different components are obtained through iterative convex optimization at each layer. Finally, using these filters, signals are decomposed into multiple fault components corresponding to different fault sources. Furthermore, the white noise replacement module is proposed to solve the gradient vanishing problem introduced by successive decompositions. Also, an updatable online learning framework is proposed for the incremental demand scenario, providing data efficiency and interpretability. The effectiveness of this method is validated on real datasets.**

***Index Terms*—Successive difference mode decomposition, Fault diagnosis, Adaptive mode decomposition**

I. Introduction

ROTATING machinery plays an important role in mechanical equipment but is often challenged by harsh operational conditions. Effective fault diagnosis methods are essential for early failure detection, ensuring equipment safety, and minimizing economic losses [1], [2]. Fault diagnosis methods for rotating machinery typically fall into three categories: signal processing-based [3], AI-based [4], and physics-informed [5] approaches. Among them, signal processing techniques deal with signals that often contain periodic components, fault features, and background noise. In rotating machinery, these signals can be collected from various sources, such as current signals [6], acoustic emissions [7], and vibration signals [8]. Among them, vibration signal analysis is widely used for fault diagnosis due to its sensitivity to mechanical anomalies and ease of measurement. These methods typically involve constructing a filter bank, extracting features, and performing fault diagnosis [9]. They focus on extracting repetitive transient features and impact responses, which are combined with fault mechanisms for thorough analysis.

Signal decomposition is important for fault feature extraction [10], [11]. Compared with traditional signal decomposition method with predefined decomposition basis functions, adaptive mode decomposition algorithms, including empirical mode decomposition (EMD) [12] with its variations [13], [14], variational mode decomposition with its variations [15], [16], [17], and feature mode decomposition with its variations [18], [19], do not need to construct basis functions. They adaptively decompose signals to extract fault components efficiently.

However, these algorithms rely only on the characteristics of the signals themselves to decompose the signals without considering the features of normal and fault signals. According to the practical requirements of fault diagnosis tasks, signals can be divided into fault components, natural periodic components, and noise. Among them, the fault components are the ones that are actually needed for fault diagnosis tasks. To decompose signals according to the difference between normal and fault signals, Wang et al. proposed difference mode decomposition (DMD) [20], which decomposes signals into the concerned component (*CC*, fault component), the reference component (*RC*, periodic component), and the noise component (*Noise*) by weighting spectrum and convex optimization. For the fault diagnosis task, fault feature components are extracted by defining the reference and mixed signal. Considering the presence of noise in practical signals, which can affect the decomposition results, Guo proposed the ensemble difference mode decomposition method [21], [22]. This method effectively weakens the influence of noise by introducing a transfer path elimination technique. It also improves the initialization weights of the difference spectrum by introducing normalized Fourier spectrum differences, thereby enhancing the computational speed. Although the above methods achieve effective separation of fault signals, the spectrums of different *CC*s are still mixed. However, in practical diagnosis tasks, it is necessary to further subdivide the *CC*s and identify specific fault types.

To obtain *CC*s for multi-class fault classification, this paper proposes the successive difference mode decomposition (SDMD). This method first defines *RC* and *CC*s in the signal based on different health conditions. By solving convex optimization for the weights of the spectrum at each decomposition layer, filter indexes (FI) corresponding to *RC* and *CC*s are obtained. Then the signal is filtered into *CC*s corresponding to different fault sources. The white noise replacement module is proposed for

This work was supported in part by The Science Center for Gas Turbine Project (Project No. P2022-DC-I-003–001). *(Corresponding author: Ruqiang Yan)*

Chao Teng, Zuogang Shang, and Xuechun Bai are with the School of Mechanical Engineering and , Xi'an Jiaotong University, Xi'an, 710049, China. (e-mail: Tengchao@stu.xjtu.edu.cn; shangzuogang@stu.xjtu.edu.cn; bxc0524@stu.xjtu.edu.cn).

Ruqiang Yan is with the School of Mechanical Engineering, and also the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, 710049, China. (e-mail: yanruqiang@xjtu.edu.cn)

Asoke K. Nandi is with the Department of Electronic and Electrical Engineering, Brunel University of London, UB8 3PH Uxbridge, U.K., and also with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710021, China (e-mail: asoke.nandi@brunel.ac.uk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

optimizing threshold processing, making the computation of the threshold more stable.

In addition, conventional fault diagnosis may face the situation of insufficient fault data. Normal signals are easy to obtain, and new faults constantly occur during operation, so a model that can be continuously updated online is needed [23], [24]. To tackle the problem, the multivariate Gaussian probability density function of the energy of the sub-signal is used to determine whether it is a new fault. Then the original faults are treated as a whole, and the part of the new fault removed from the filter index is used to achieve updatable online learning. This paper is an extended version of the proceedings paper [25]. The main contributions of this paper are as follows:

(1) SDMD framework is proposed for adaptive multi-class signal decomposition and fault diagnosis. It effectively utilizes the characteristics of available data on various health conditions to decompose signals.

(2) A white noise replacement module is proposed for thresholding. The gradient vanishing problem is overcome by adding white Gaussian noise. It is more suitable for the successive decomposition structure.

(3) An updatable online learning SDMD framework is proposed for the practical scenario where new faults may occur. Experimental validation demonstrates efficient anomaly detection and time-saving performances.

The rest of the paper is organized in the following way. Section II briefly introduces the algorithm of DMD. Section III provides a detailed introduction to the proposed SDMD method, including the overall framework, white noise replacement module, and theoretical characteristics. Section IV provides a variant of a solution for online incremental fault learning, including an overall framework, open-set detection module, and incremental learning module. Sections V and VI present the validation of the effectiveness and superiority of the framework using simulated and real signals, respectively. Finally, the conclusion is summarized in Section VII.

II. Preliminary

In this section, the preliminary works for the proposed method are described. Section A introduces the overall framework of the original difference mode decomposition (DMD) method. Section B introduces the threshold processing module of DMD, which will be improved in Section III B. Section C introduces the multivariate Gaussian probability density function and its functions.

A. Difference mode decomposition (DMD) [20]

As shown in Fig. 1, DMD algorithm assumes that the collected fault signal is a mixed signal (*MIX*), which can be decomposed into three components: concerned component (*CC*), reference component (*RC*), and noise component (*Noise*):

$$MIX = CC + RC + Noise\,. \tag{1}$$

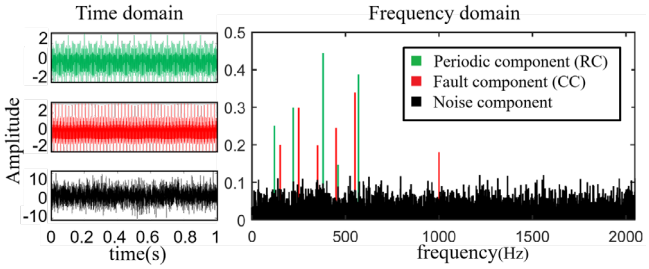

Fig. 1. Different components of a fault signal (*MIX*).

For a specific fault signal, *CC* represents the impulsive features caused by the fault. While *RC* represents the natural periodic features, which are the components present in normal signals. A reference signal (*REF*, usually a normal signal) has two components: *RC* and *Noise*.

For DMD, the Fourier spectrum is weighted to separate the *CC*, which is the fault feature. The specific steps are as follows:

1. Vibration signals are collected from the equipment to obtain the fault signal S_M and the reference signal S_R (normal signal).

2. The Fourier transform is applied to S_M and S_R to obtain the fault signal spectrum FS_M and the reference signal spectrum FS_R. The obtained spectrums are then normalized to obtain the normalized fault signal spectrum NFS_M and the normalized reference signal spectrum NFS_R.

$$NFS_M = \frac{FS_M}{\sum FS_M} \text{ and } NFS_R = \frac{FS_R}{\sum FS_R} \tag{2}$$

3. The spectrum NFS_M and NFS_R are weighted. Convex optimization is used to solve for the optimal difference spectrum ω_{best} [26]. The optimization problem can be formulated as follows:

$$\begin{aligned}
&\omega_{best} = \arg\min L(\omega) = \lambda\left[\omega^T, b\right]\left[\omega^T, b\right]^T \\
&+\frac{1}{2(P+Q)}\sum_{i=1}^{P+Q}\left[-y_i\left[\omega^T, b\right]z_i + \log(1+\exp(\left[\omega^T, b\right]z_i))\right], \\
&z_i = [(NFS_i)^T, 1]^T,
\end{aligned} \tag{3}$$

where λ represents the regular term coefficient, which is set to 0.1. ω and b represent weight and bias. After optimization, ω becomes ω_{best}, called the optimal difference spectrum in this paper. Q and P represent the quantities of *REF* and *MIX*. The binary classification label y_i is set to 1 and 0 when *NFS* equals NFS_M and NFS_R, respectively. NFS_M and NFS_R represent the normalized Fourier spectrum of *MIX* and *REF*, respectively.

4. After obtaining the optimal difference spectrum, it is sorted by the percentile amplitude of each spectrum line. Then, as shown in Fig. 2, two thresholds - th_c, and th_r (see Section II B) are used to separate the difference spectrum into 3 parts, corresponding to *CC*, *RC*, and *Noise*, respectively. The indexes of separated spectrum lines are denoted as I_c, I_r, and I_{no}, as shown in eq. (4). I_c represents the spectrum lines whose amplitude is higher than *thc*, and so forth.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

$$I_c = (\omega_{best} > th_c), I_r = (\omega_{best} < th_r),$$
$$I_{no} = (th_c \le \omega_{best} \le th_r) \quad (4)$$

5. The obtained spectrum line indexes are used as filters to perform filtering on the original signal. The filtered spectrums are denoted as S_c, S_r, and S_{no}, respectively. These filtered spectrums are then subjected to an inverse Fourier transform to complete the decomposition.

$$x_C = ifft(S_c), x_R = ifft(S_R), x_{no} = ifft(S_{no}), \quad (5)$$

where x_C, x_R, and x_{no} are *CC*, *RC*, and *Noise*, respectively.

B. DMD threshold processing module

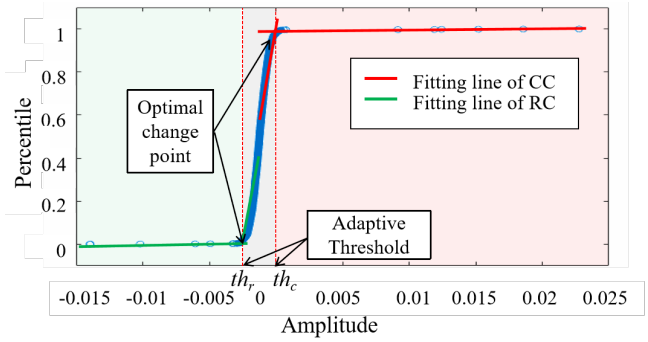

Fig. 2. The optimal difference spectrum in a percentile plot.

In the 4th step of DMD, the amplitude percentile of the decomposed optimal difference spectrum ω_{best} is calculated and sorted. Fig. 2 illustrates the plot of amplitude versus percentile, similar to an empirical cumulative distribution function plot. The coordinates of each optimal change point are calculated by minimizing the error between two fitted line segments. The specific calculation formula is as follows:

$$T = \begin{cases} \alpha_1 + \beta_1 p_n, & n = 1,2,...,\tau, \text{ and} \\ \alpha_2 + \beta_2 p_n, & n = (\tau+1),(\tau+2),...,N_P \end{cases}$$
$$E_1(\tau) = \sum_{n=1}^{\tau} (T_n - (\alpha_1 + \beta_1 p_n))^2, \quad (6)$$
$$E_2(\tau) = \sum_{n=1}^{\tau} (T_n - (\alpha_2 + \beta_2 p_n))^2,$$

$$\tau = \arg\min(E_1(\tau) + E_2(\tau)), \quad (7)$$

where α and β are the coefficients of two straight lines. The τ-th point is the optional change point. p_n is the n-th percentile point. N_P represents the maximum number of points. Taking the search for change points on the *RC* side, it is typically chosen as the bottom 16.7% of the total number of points in the difference spectrum. In Eq. (7), the optimal value of the change point τ is obtained by minimizing the error $E(\tau)$ between two straight lines. The amplitude value corresponding to the τ value serves as the threshold th_r. Similarly, taking the search for change points on the *CC* side, it is typically chosen as the bottom 16.7% of the total number of points in the difference spectrum. Then following Eq. (7), the amplitude value corresponding to the τ value serves as the threshold, th_c.

C. Multivariate Gaussian probability density function (PDF)

The formulation of d-dimension PDF is as follows [27]:

$$\text{PDF} = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right), \quad (8)$$

where μ and Σ represent mean vector and covariance matrix, respectively.

The PDF is a multivariate extension of the univariate normal distribution. It is used to describe and analyze the overall distribution of multiple random variables. In this study, PDF is used as a quantitative measure to assess whether a sample belongs to one of the learned faults. The larger the value of PDF, the higher the likelihood that the data belongs to that distribution.

III. Successive Difference Mode Decomposition

A. The overall SDMD framework

In the basic DMD algorithm, regardless of the number of faults, the signal is always decomposed into three components: *CC*, *RC*, and *Noise*. However, for the component *CC*, which contains features of all fault types, it can be challenging to analyze individual faults and achieve precise fault classification. With the prior difference of faults from historical data, further decomposing the *CC* to obtain the frequency spectrum corresponding to each fault is possible. The proposed structure in this study is as follows:

$$MIX = RC + \sum_{i=1}^{T} (CC_i + Noise_i). \quad (9)$$

Different from Eq. (1), this paper considers that the *CC* is composed of T types of faults. In each level of decomposition, a filtering component for noise is included, resulting in a multilevel decomposition of the noise component. The adaptive decomposition generates individual *CC* components corresponding to the fault components with prior label information.

As shown in Fig. 3, taking three types of faults as an example, the content of the successive decomposition framework is as follows:

1. The *MIX* signal consists of three fault types: Fault_1, Fault_2, and Fault_3. The normal signal is considered the reference signal (*REF*). The dataset is constructed using these signals.

2. The normalized Fourier spectrum NFS_M and NFS_R of the *MIX* and *REF* signals are calculated respectively. DMD is then performed to obtain the optimal difference spectrum ω_{best}:

$$\omega_{best} = \text{DMD}(NFS_M, NFS_R), \quad (10)$$

The *RC* spectrum line indexes I_r obtained from Eq. (4) represent the spectrum components extracted at the current level of decomposition. These indexes indicate the specific spectrum lines corresponding to the *RC* component, which need to be saved. Then, the dataset is updated by removing the *REF*. The remaining fault spectrum is then filtered using

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

the *CC* indexes. This involves applying the *CC* index as a filter to the spectrum of the other faults, followed by inverse transformation and re-normalization of the filtered spectrum.

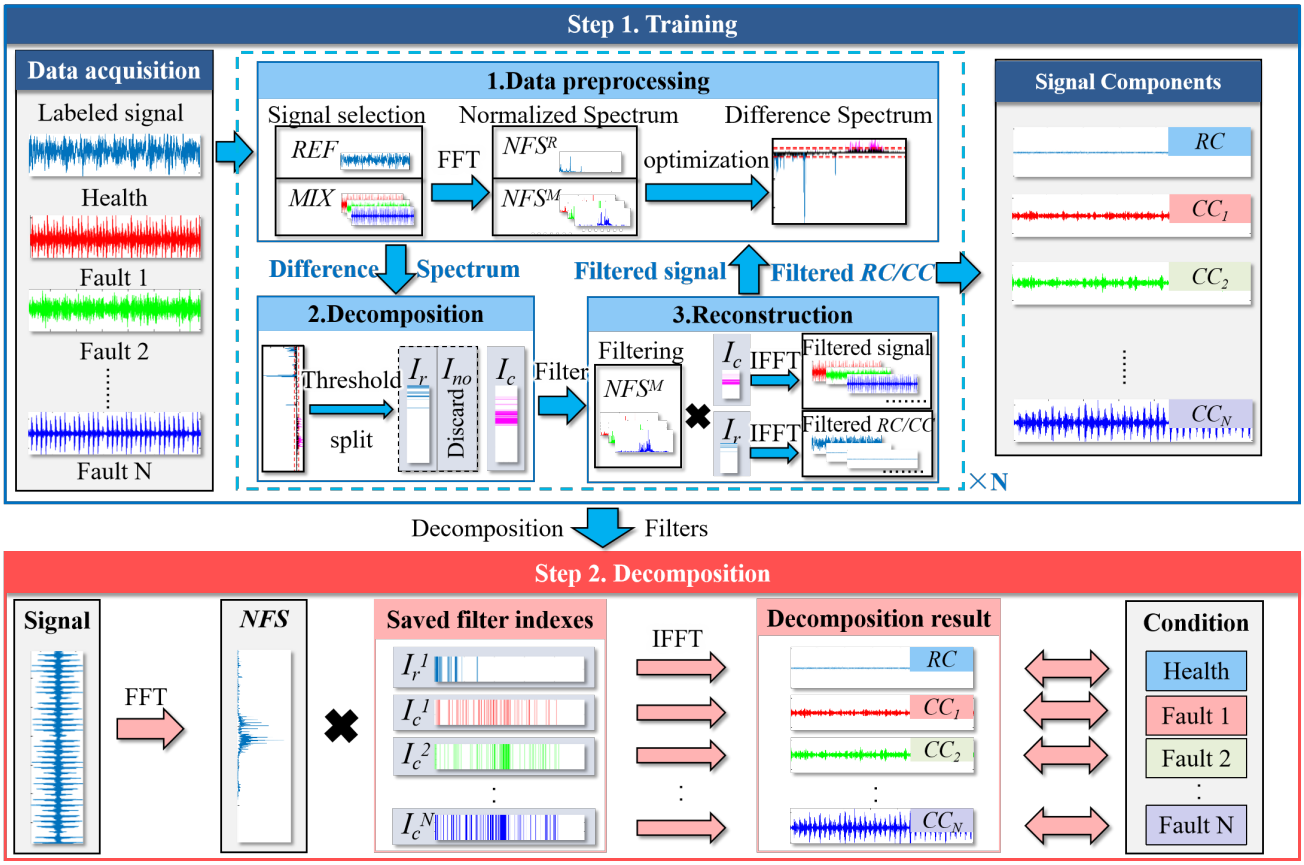

Fig. 3. SDMD framework.

$$S_c = NFS_M \odot I_c$$
$$x = ifft(S_c) \tag{11}$$

where the operator ($\odot$) represents the dot product.

3. One of the fault components is randomly selected as *REF*. The remaining two fault components are used as *MIX*. The process of the 2nd step is repeated for the new *MIX* and *REF*.

4. When only two fault components remain after repeated decomposition, the last decomposition is performed. The *CC* index obtained from this final decomposition is retained as the final spectrum components, representing the last fault-specific features. As shown in Fig. 3, 1 *RC* (signal corresponding to blue filter index) and 3 *CCs* are extracted.

5. After all *CC* and *RC* are separated from the original signal, the residual component is *Noise*. This is more efficient and memory-saving compared to saving the indexes of noise at each level of decomposition.

In the case of having T faults, the optimized final model can be represented as follows:

$$MIX = RC + \sum_{i=1}^{T}(CC_i) + Noise\,. \tag{12}$$

The complete SDMD framework is shown in Algorithm 1.

Algorithm 1 Successive Difference Mode Decomposition

Input: Dataset *signal* with $N \times L$; The label of data *label* with $1 \times L$.

Output: Filter index matrix F; Best difference spectral ω_{best}.

Procedure:

Extract all unique labels $l = unique(label)$ with $1 \times C$

Initialize

$F = zeros(\frac{N}{2}+1,\ C)$ and $\omega_{best} = zeros(\frac{N}{2}+1,\ C-1)$;

Cumulative increasing value $num = 1$; Copy of unique labels $mixlabel = l$.

for $ii = l(1:end-1)$ **do**

Discard the *REF* value of the previous layer of *mixlabel*

Get *REF* from *signal* which $label = ii$

Get *MIX* from *signal* which $label \in mixlabel$

Calculate Normalized Fourier Spectral NFS_M, NFS_R

Make NFS_M, NFS_R the same size (Data Enhancement).

Calculate ω using **Eq. (3).**

Threshold processing to get filter index I_c, I_R.

Saving $F_{I_r,num} = 1, \omega_{best_{num}} = \omega$.

Update *signal* using **Eq. (5).**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

Add white Gaussian *Noise* to *signal*.
end for
Saving $F_{I_c,num}=1$
Return: Filter index matrix F; Best difference spectral ω_{best}.

B. White noise replacement module for adaptive threshold learning

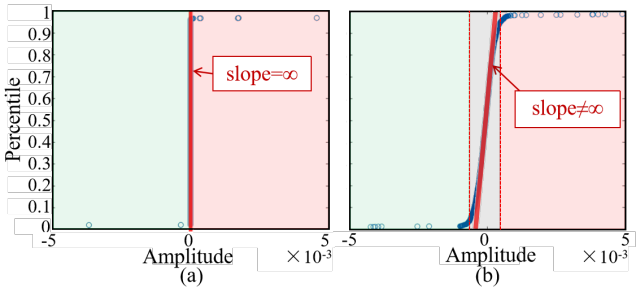

Fig. 4. Optimal difference spectrum with amplitude limited between ±5e-3 (a) The result without introducing noise has an infinite slope. (b) The result of introducing noise is a finite slope.

During successive decomposition, it is difficult to compute th_c and th_r effectively, which makes the separation of *CC*, *RC,* and *Noise* impossible. As shown in Fig.4 (a), after each level of decomposition, the *RC* and *Noise* are filtered and set to zero. Since there are too many zero-valued components, one of the lines fitted becomes a vertical line perpendicular to the coordinate axis. This results in an infinite slope, making it impossible to calculate the error and leading to a threshold of 0, introducing significant errors.

Therefore, white noise is added to soften the zero-valued spectrum lines caused by noise. In the re-normalization process of the 2nd step, a suitable amplitude of white noise is introduced to change the slope of the noise percentile fitting, preventing it from becoming infinite. The added noise is shown in Eq. (13).

$$x_c^{New}=x_c+A\times randn(N), \tag{13}$$

where A represents the amplitude of the introduced noise, and its value is set to 0.2 times the standard deviation of x_c [28]. As shown in Fig. 4 (b), with added white noise, the fitted curve obtained the slope.

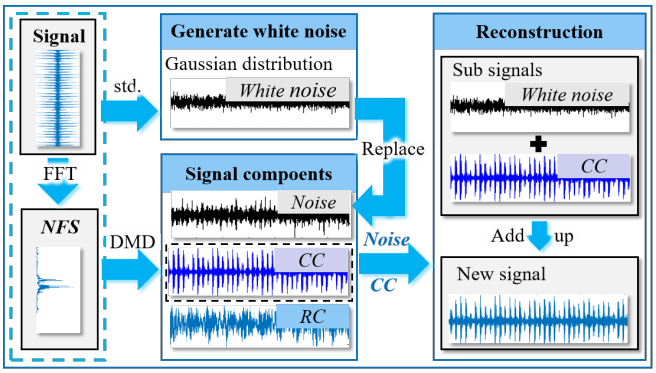

Fig. 5. The white noise replacement module.

The white noise replacement module is shown in Fig. 5. Before decomposition, white noise with standard amplitude and variance is used to replace *Noise* in the original signal, without changing the characteristics. It also improves the thresholding module of the next layer, making the training process of this framework more stable.

C. Theoretical Characteristics of SDMD framework

The proposed framework possesses the following characteristics:

1) Order-Invariance of Decomposition

Assuming that all significant fault spectrum lines that reflect fault characteristics are located in different positions, the decomposition order is independent of the presence of noise that may obscure the fault spectrum lines:

$$NFS_i=\left[a_1^{(i)},a_2^{(i)},...,a_n^{(i)}\right] \tag{14}$$

When the j-th spectrum line is neither an *RC* nor a *CC* spectrum line, $a_j^{(i)}=0$. Taking three types of faults as examples when $n=5$. When $i=1, 2, 3$, the three types of faults are:

$$\begin{aligned}NFS_{M1}&=\left[A_1^{(1)},\ a_2^{(1)},\ a_3^{(1)},\ a_4^{(1)},\ a_5^{(1)}\right]\\ NFS_{M2}&=\left[a_1^{(2)},\ A_2^{(2)},\ a_3^{(2)},\ a_4^{(2)},\ a_5^{(2)}\right]\\ NFS_{M3}&=\left[a_1^{(3)},\ a_2^{(3)},\ A_3^{(3)},\ a_4^{(3)},\ a_5^{(3)}\right]\end{aligned} \tag{15}$$

In this case, the distinguishable fault frequencies are represented by the uppercase letter A. This means that the amplitude A^i percentage in the current *NFS* at that position is much higher compared to the other faults at the same position. Since the essence of the DMD algorithm is spectrum weighting, the difference spectrum corresponding to the higher amplitude percentage has a larger weight. The results obtained after thresholding are more likely to be retained compared to the smaller weight spectra. Taking NFS_{M1} as *REF* and the rest as *MIX*, the first layer of decomposition yields a filtered index focused on fault 1:

$$\begin{aligned}I_{r1}&=\left[1,\ 0,\ 0,\ i_{r4}^{(1)},\ i_{r5}^{(1)}\right]\\ I_{c1}&=\left[0,\ 1,\ 1,\ i_{c4}^{(1)},\ i_{c5}^{(1)}\right]\end{aligned}, \tag{16}$$

$$i_r=1-i_c\in\{0,1\}, \tag{17}$$

where i_r, i_c represent the spectrum lines of *RC* and *CC* at the corresponding position, respectively. In the corresponding positions of the spectrum lines, the results here are not important because they do not represent distinguishable significant spectrum lines. Through the first layer of decomposition, it can be observed that the important spectrum lines focused on the fault are preserved at their corresponding positions. At this stage, reconstruction is performed to obtain the *NFS* required for the second layer of decomposition, as shown in Eq. (18):

$$\begin{aligned}NFS_{M2}&=\left[0,\ A_2^{(2)},\ a_3^{(2)},\ i_{c4}^{(1)}\times a_4^{(2)},\ i_{c5}^{(1)}\times a_5^{(2)}\right]\\ NFS_{M3}&=\left[0,\ a_2^{(3)},\ A_3^{(3)},\ i_{c4}^{(1)}\times a_4^{(3)},\ i_{c5}^{(1)}\times a_5^{(3)}\right]\end{aligned} \tag{18}$$

Performing the next layer of decomposition, the filter index obtained is:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

$$I_r^{(2)} = \left[0,\ 1,\ 0,\ i_{r4}^{(2)},\ i_{r5}^{(2)}\right]$$
$$I_c^{(2)} = \left[0,\ 0,\ 1,\ i_{c4}^{(2)},\ i_{c5}^{(2)}\right] \tag{19}$$

Based on all the decomposed filtered indices, signal filtering is performed at the corresponding layer, resulting in the following results:

$$NFS_{M1}^{filter} = \left[A_1^{(1)},\ 0,\ 0,\ i_{r4}^{(1)} \times a_4^{(1)},\ i_{r5}^{(1)} \times a_5^{(1)}\right]$$
$$NFS_{M2}^{filter} = \left[0,\ A_2^{(2)},\ 0,\ i_{c4}^{(1)} \times i_{r4}^{(2)} \times a_4^{(2)},\ i_{c5}^{(1)} \times i_{r5}^{(2)} \times a_5^{(2)}\right]$$
$$NFS_{M2}^{filter} = \left[0,\ 0,\ A_3^{(3)},\ i_{c4}^{(1)} \times i_{c4}^{(2)} \times a_4^{(3)},\ i_{c5}^{(1)} \times i_{c5}^{(2)} \times a_5^{(3)}\right] \tag{20}$$

In the calculations from Eq. (16) to Eq. (20), changing the order of the three types of faults essentially changes the position of the important fault spectrum line *A* at each layer. In the spectrum weighting algorithm, the position of the spectrum is discrete, and its order does not affect the calculation results. The change in order primarily affects the position of the non-distinguishable important spectrum line *a*, which is irrelevant to the diagnostic task aiming to extract results. Additionally, the non-important spectrum line *a* is mostly noise, with a small portion being the aliasing result of high-frequency modulation. The former will be filtered by the thresholding module, while the latter will be added as the *CC* component to the *REF* index of the current layer, possibly resulting in a weak modal superposition phenomenon. Overall, changing the decomposition labels, i.e., changing the decomposition order, does not have any effect on the results of the important spectrum lines of interest, thus reducing the issue of selecting hyperparameters. Eq. (21) summarizes the order-invariance under three-layer decomposition. This can be extended to four layers, five layers, or even higher levels of decomposition.

$$SDMD(1,\ 2,\ 3) = SDMD(1,\ 3,\ 2)$$
$$= SDMD(2,\ 1,\ 3) = ... = SDMD(3,\ 2,\ 1) \tag{21}$$

2) Orthogonal Completeness

In the iterative decomposition process with a total of N layers, for all the *RC* and the final residual component *CC* indices obtained from the decomposition:

$$I_{ri} = \{a_i^{(1)},\ a_i^{(2)},\ ...,\ a_i^{(n)}\}$$
$$I_{rc} = \{b^{(1)},\ b^{(2)},\ ...,\ b^{(n)}\} \tag{22}$$

where *i* represents the decomposition layer, *n* represents the point index, $a_i^k, b^k \in \{0,1\}, k \in 1,2,...,n$. The *RC* and *Noise* spectrum lines obtained from each layer are set to zero and will not appear in the subsequent layers. This means their product is zero, as shown in Eq. (23).

$$b^k \prod_{i=0}^{N} a_i^k = 0 \tag{23}$$

The inner product of all spectrum lines is then given by:

$$\left\langle I_{r1},\ I_{r2},...,\ I_{rN}, I_c \right\rangle = \sum_{k=1}^{N+1} b^k \prod_{i=1}^{n} a_i^k = 0 \tag{24}$$

The *Noise* spectrum lines include all spectrum lines except S_r and S_c. By performing inverse transformation and linear weighting on all the spectrum lines obtained from the decomposition, the result is as follows:

$$NFS = \sum_{i=1}^{N} l_i \times S_{ri} + l_{N+1} \times S_c + l_{N+2} \times S_{no} \tag{25}$$

It can be seen that $l_i = 1$ represents the weight of the spectrum line. Therefore, the frequency domain orthogonality completeness of the signal is proven.

Furthermore, from Eq. (17) and Eq. (20), it can be concluded that:

$$\begin{aligned}
&\left\langle NFS_{M1}^{filter}, NFS_{M2}^{filter}, NFS_{M3}^{filter} \right\rangle \\
&= (A_1^{(1)} \times 0 \times 0) + (0 \times A_2^{(2)} \times 0) + (0 \times A_3^{(3)} \times 0) + \\
&[i_{r4}^{(1)} \times (i_{c4}^{(1)} \times i_{r4}^{(2)}) \times (i_{c4}^{(1)} \times i_{c4}^{(2)})] + \\
&[i_{r5}^{(1)} \times (i_{c5}^{(1)} \times i_{r5}^{(2)}) \times (i_{c5}^{(1)} \times i_{c5}^{(2)})] \\
&= i_r^{(1)} \times (i_c^{(1)} \times i_r^{(2)}) \times (i_c^{(1)} \times i_c^{(2)}) \\
&= i_r^{(1)} \times ((1 - i_r^{(1)}) \times i_r^{(2)}) \times ((1 - i_r^{(1)}) \times (1 - i_r^{(2)})) \\
&= 0
\end{aligned} \tag{26}$$

Without considering the presence of noise filtering, there is only one value among $i_r^1,\ i_c^1 \times i_r^2$, and $i_c^1 \times i_c^2$ that is equal to 1. This means that in the resulting *NFS* after the three filters, each corresponding position of the spectrum line has only one value. It demonstrates the orthogonality of the proposed framework again.

3) Sparsity

The sparsity property facilitates the extraction of distinct periodic fault characteristic frequencies, enabling interpretable fault diagnosis [29], [30]. As mentioned in Section III, with an increase in the number of layers, the number of non-zero frequencies decreases. Sparsity refers to the concentration of coefficients representing fault characteristics on a small subset of the difference spectrum. Therefore, it is only necessary to perform Fourier transformation on the signal and filter only the filter indexes obtained, achieving compressed storage of the signals of interest (such as fault components). The removal of noise components can also be achieved. Eq. (27) provides evidence for this.

$$\left\|I_{r1}\right\|_0 + \left\|I_{r2}\right\|_0 + ... + \left\|I_{rL}\right\|_0 + \left\|I_c\right\|_0 < \left\|I_{signal}\right\|_0 = N, \tag{27}$$

where *L* represents the total number of layers. I_{signal} represents all spectrums of the original signal, and *N* is the total number of discrete spectrum lines.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

IV. Updatable online learning SDMD (USDMD)

A. The overall USDMD framework

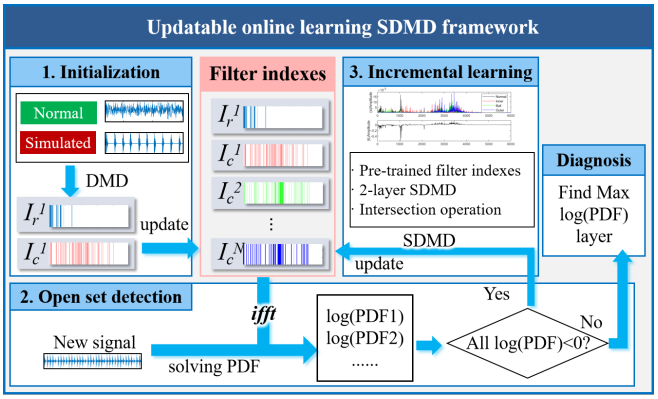

Fig. 6. The updatable online learning SDMD framework, includes the initialization module, the open-set detection module, and the incremental learning module.

In practical production scenarios, normal signals are more common, and obtaining fault signals can be challenging. What is more, various anomalies may occur, making it challenging for our trained models to comprehensively address all possible faults. Therefore, the problem addressed in this section is how to detect signals that do not belong to the current sample library and extend the existing SDMD layers accordingly. An updatable online learning fault diagnosis architecture is proposed, which includes open-set detection for new signals and incremental learning. The overall training process is illustrated in Fig. 6 and is outlined as follows:

1) Weights initialization. If fault signals exist, use them as *MIX*; if not, generate a simulated signal arbitrarily. Use normal signals as *REF*. Perform SDMD to obtain the difference spectrum and filtering indexes for each layer, which serve as the initialization values for the framework.

2) Open-set detection. Input the new signal and perform SDMD. Calculate the energy for each decomposition layer and compute the logarithmic PDF values between each energy vector and known faults. If the value is below a given threshold, it is considered a new fault and proceeds to step 3. If it does not meet the requirement, diagnose the corresponding fault based on the maximum PDF value.

3) Incremental learning. Label the new samples and input them as new faults for incremental learning within the SDMD framework.

The details of each submodule are further elaborated in the remaining subsections of this section.

B. Weight initialization module

In the case where one fault and normal signals are available, performing SDMD according to Eq. (10) will yield the initial difference spectrum and filtering indexes.

However, in practical production scenarios where pre-designed experiments are not conducted, the signals collected are often normal, and obtaining fault signals can be challenging. Therefore, obtaining the *MIX* signal becomes a challenge in initializing the difference spectrum weights. Considering that simulated signal models are easy to establish, in the absence of faults, a simulated signal can be created based on the fault mode to serve as Fault_1 [31]. SDMD can then be performed to obtain the initial difference spectrum.

In the subsequent open-set detection module, when a fault is detected, the detected fault is used to replace the simulated signal, updating the initial optimal difference spectrum.

C. Open-set detection module

To detect whether a new signal belongs to existing faults, a simple open-set detection module is proposed in this section. The specific steps are as follows:

1. The existing difference spectrum and filtering indexes are used to perform SDMD on the new signal according to Eq. (11) to obtain the sub-signals for each layer. Calculate the energy of each layer's signal. The energy E of a discrete signal with length N is calculated as follows:

$$E = \sum_{i=1}^{N} x_i^2 \tag{28}$$

2. Randomly select typical signals with known fault labels and calculate the energy of each fault type in each layer. Calculate the mean and self-covariance of the energies using Eq. (29).

$$\begin{aligned} mean(E) &= \frac{1}{N}\sum_{i=1}^{N} E_i \ , \\ \text{cov}(E) &= \frac{\sum_{i=1}^{N}(E_i - mean(E))^2}{N-1}, \end{aligned} \tag{29}$$

where E_i represents the energy of signals with the same label and in the same layer. For a sample set with T different fault conditions, the mean(E) has a length of T, and the cov(E) matrix has a size of $T \times T$.

3. Use Eq. (8) and Eq. (30) to calculate the logarithmic PDF (LPDF) values of the new signal relative to each existing fault, which is used for similarity judgment. If all the LPDF values are less than 1, it is considered a new fault sample, and the incremental learning module is executed. If there are values above 1, diagnose the fault type T_{signal} as the fault layer corresponding to the maximum LPDF value, as shown in Eq. (31).

$$\text{LPDF} = \log(\text{PDF}) \tag{30}$$

$$T_{signal} = \begin{cases} \text{New\_fault} & \forall \text{LPDF}<1 \\ \text{Normal} & \exists \text{LPDF} > 1 \,\&\, \arg\max(\text{LPDF}_i) = 0 \\ \text{Fault\_}i & \exists \text{LPDF} > 1 \,\&\, \arg\max(\text{LPDF}_i) = i \\ & i = 0,1,...,T \end{cases} \tag{31}$$

Considering that there may be some errors in the operation of the equipment and signal acquisition process when determining a fault that does not belong to known labels, it is necessary to have a continuous number of signals with all LPDF values less than 1. This confirms the occurrence of a new fault. The number of examples can be arbitrarily chosen, and in this paper, it is set to 15.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

D. Incremental learning module

After obtaining new fault samples, it is necessary to update the filtering indexes of each layer and train the decomposition layer corresponding to the new fault to achieve incremental learning. The process is as follows:

1. The existing fault data with known labels are randomly selected as typical signals. The number of data for each type should be kept as similar as possible. In this paper, the number of examples for each fault condition is set to 5. The continuous 15 fault examples collected in the open-set detection module are given a new fault label to establish a new dataset. However, when performing SDMD, the label should be given as follows: The health datasets are 0, and the known faults are all given 1, mixed as one type of fault. And the new fault is separately assigned to 2. The correspondence between the original and USDMD labels is shown in TABLE I.

TABLE I. LABEL CORRESPONDENCE TABLE

Method	Health	Known fault	New fault
Fault label	0	1~*N*-1	*N*
USDMD label	0	1	2

2. Then, a two-loop SDMD is performed to obtain filter indexes of each health condition. The first loop is to obtain FIs of health conditions, and the second loop is to separate the new fault from other known faults. So the filter indexes of health and new fault are saved for the next calculation, while others corresponding to fault 1 to (N-1) are discarded.

3. Finally, the intersection with the new fault FIs from the FIs of known faults is removed, which is to satisfy the orthogonality of FIs. The specific process of this step is shown in Algorithm 2.

Algorithm 2 Updatable SDMD

Input: Dataset *signal* with $N \times L$, The label of data *label* with $1 \times L$, filter index matrix of known health condition F

Output: Updated filter index matrix F.

Reset labels using TABLE I. ;

Procedure:

Get a new filter index matrix F^{new} by performing SDMD using Algorithm 1.

Replace the filter index of health by:

$$F_{n,1} = F_{n,1}^{new},\ n = 1, 2, ..., (\frac{N}{2}+1)$$

Add filter index of new fault by:

$$F_{n.N} = F_{n,3}^{new},\ n = 1, 2, ..., (\frac{N}{2}+1)$$

Update filter indexes of known faults by:

$$F_{n,num} = F_{n,num} - F_{n,num} \cap F_{n,3}^{new}$$

$$n = 1, 2, ..., (\frac{N}{2}+1), \text{ and } num \in [2, N-1]$$

Return: Updated filter index matrix F.

The above incremental learning module considers the filter indexes of known faults as a whole, only requires two loops to separate the new fault, and removes the part of the filter indexes of known faults containing the new fault based on orthogonality. This process can meet the requirement that no matter how many faults exist in the established fault library, any new fault can be learned in just two loops, greatly reducing the time for retraining and achieving efficient incremental learning.

V. SIMULATION SIGNAL CASE ANALYSIS

In this section, a comparative analysis using simulated signals is performed to validate the effectiveness of SDMD framework. The decomposition performance and noise reduction effect of the threshold processing module are included.

In practical scenarios, the characteristic frequencies of multi-class fault components may not be distributed in separate frequency bands. Similar to the approach in [20], the following simulated signals are constructed:

$$\begin{aligned} x_R(t) &= \sum_{i=1}^{N_R} A_{Ri} \cos(2\pi f_{Ri} t) \\ x_C^{(j)}(t) &= \sum_{i=1}^{N_C} A_{Ci}^{(j)} \cos(2\pi f_{Ci}^{(j)} t) \end{aligned}, \tag{32}$$

where t represents time, $x_R(t)$ represents RC, $x_c^{(j)}(t)$ represents the j-th CC, A_{Ri} and f_{Ri} represent the amplitude and frequency of the i-th components in RC, respectively. $A_{Ci}^{(j)}$ and $f_{Ci}^{(j)}$ represent the amplitude and frequency of the i-th component in the j-th CC, respectively. N_R and N_C represent the total number of frequencies in RC and CC, respectively.

Considering the various fault situation in fault diagnosis, three fault components are defined with their amplitudes and frequencies respectively, i.e. $f_C^{(1)}$ = [150, 250, 350, 540, 550, 1000]^T, $f_C^{(2)}$ = [155, 245, 355, 455, 540, 900]^T, $f_C^{(3)}$ = [160, 255, 360, 440, 560, 950]^T, $A_C^{(1)}$ = [0.4, 0.6, 0.4, 0.5, 0.7, 0.4]^T, $A_C^{(2)}$ = [0.5, 0.4, 0.3, 0.6, 0.5, 0.4]^T, $A_C^{(3)}$ = [0.45, 0.45, 0.5, 0.6, 0.7, 0.5]^T. The amplitude and frequency of the normal signal are f_R = [120, 220, 380, 460, 570]^T, and A_R = [0.5, 0.6, 0.9, 0.3, 0.8]^T.

By substituting values into Eq. (5), the results of all components can be constructed. At low frequencies, all fault components closely resemble the spectrum of the normal signal, with slight differences in amplitude. At high frequencies, only the fault spectrums are present. The spectrums of different faults are very close, with only a difference of around 10 Hz. Conventional decomposition methods using bandpass filters make it hard to effectively extract these individual fault features.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

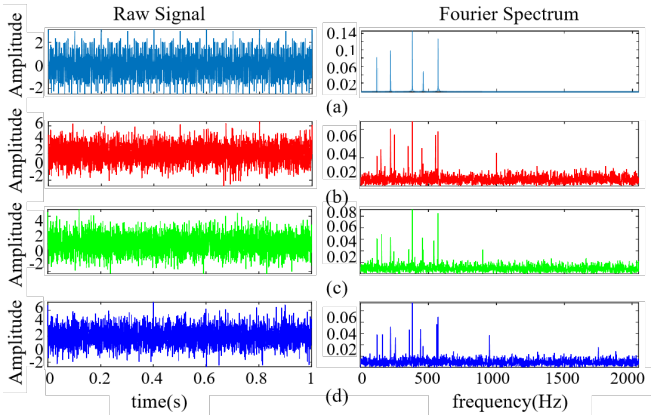

Fig. 7. Time and frequency domain of simulated normal and mixed signal. (a) *REF*, (b) MIX^1, (c) MIX^2, and (d) MIX^3.

In this simulation case, all defined fault signals are considered as single-class faults, which means there is no coupling of fault frequencies. The definition is as follows:

$$x_{MIX}^{(j)} = x_R(t) + x_C^{(j)}(t)\ , \qquad x_{REF} = x_R(t)\ , \tag{33}$$

where $x_{MIX}^{(j)}$ and x_{REF} represent the j-th mixed signal and normal signal, respectively. Fig. 7 shows the time and frequency domain of each signal with slight noise. The simulated signal with introduced slight noise is decomposed using the SDMD framework. Fig. 8 shows the different spectrums and thresholds obtained at each level of decomposition. The *RC* spectrums correspond to negative values, while the remaining *CC* spectrum is represented by positive values [20]. As the number of decomposition levels increases, the number of non-zero spectrums decreases, which aligns with the sparsity. In the last level of decomposition, the positive best difference spectrums perfectly correspond to *CC*3, and the *RC* spectrum obtained at each level does not overlap, satisfying the orthogonality.

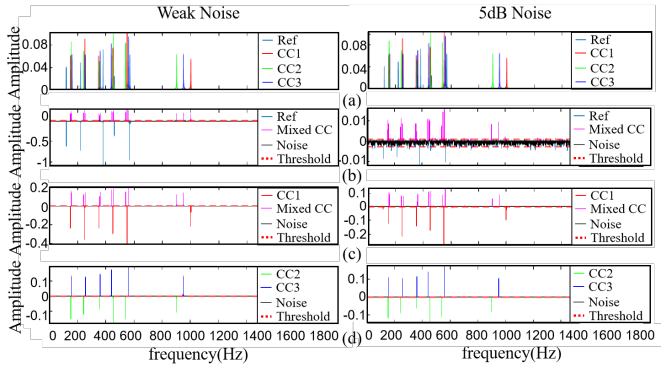

Fig. 8. (a) The comparison between the spectrum of the simulated signal at different noise levels. (b), (c), and (d) are the optimal difference spectrum obtained from the first, second, and third levels of decomposition, respectively.

The individual components obtained after inverse Fourier transformation are compared with the simulated signal in Fig. 9. In each simulated fault signal, the corresponding fault component is successfully extracted and appears at the corresponding level. The signals in the remaining levels are mostly represented by a straight line, with fluctuations due to the presence of noise components. Samples MIX^1 (a), MIX^2 (b), MIX^3 (c), and MIX^4 (d) were separately extracted and then compared individually with the simulated signal. The results are presented in Fig. 10. The blue and red line represents the extracted and simulated signal, respectively. The two lines almost completely overlap, with only slight differences caused by noise during the signal construction.

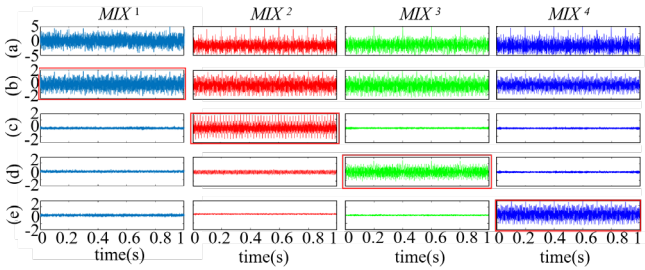

Fig. 9. The SDMD results of simulated signal *MIX*1~4. (a) Original signal, (b) *REF* component, (c) *CC*1 component, (d) *CC*2 component, and (e) *CC*3 component

For the proposed method, the threshold processing part still obtains a few noise spectrum lines. The experiment uses mean square error (MSE), signal-to-noise ratio (SNR), and peak error (PE) as evaluation metrics. The results of the proposed method and two classical noise elimination methods: wavelet packet [32] and mean filtering denoising are compared. The results are shown in TABLE II. , where *CC*1~3 of the SDMD scheme are the decomposition components obtained through the framework. Denoising signal *CC*1~3 of WPT and mean filter are using original *CC* with the noise of the same SNR. It can be observed that the MSE of the proposed framework is much smaller than the two comparison methods, the SNR is the highest, and the PE is the smallest. Therefore, the SDMD framework performs well in signal denoising.

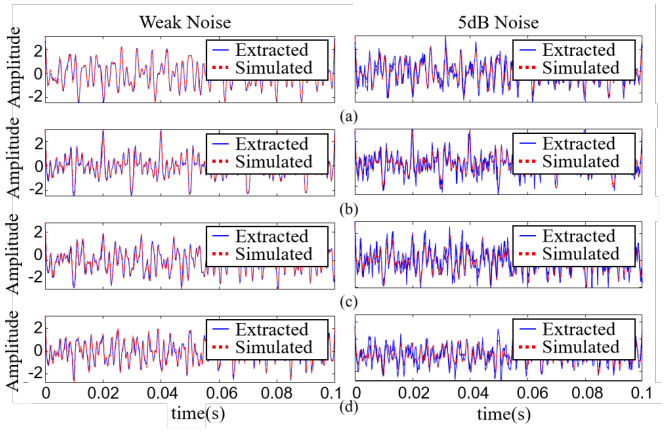

Fig. 10. The comparison among every extracted and simulated component with the different noise levels. (a) *REF*, (b) *CC*1, (c) *CC*2, and (d) *CC*3.

TABLE II. THE COMPARISON AMONG SDMD AND CLASSIC METHODS OF DENOISING EFFECT

Method	MSE			SNR			PE		
	CC1	*CC2*	*CC3*	*CC1*	*CC2*	*CC3*	*CC1*	*CC2*	*CC3*
SDMD	0.41	0.34	0.72	9.50	9.53	8.21	0.59	0.43	0.65
WPT	6.81	6.74	5.55	1.15	1.09	2.24	2.37	2.20	2.11
Mean filter	1.59	1.40	1.73	9.42	9.26	9.28	0.60	0.60	0.67

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

VI. SDMD STRUCTURE FOR FAULT DIAGNOSIS ON REAL DATASETS

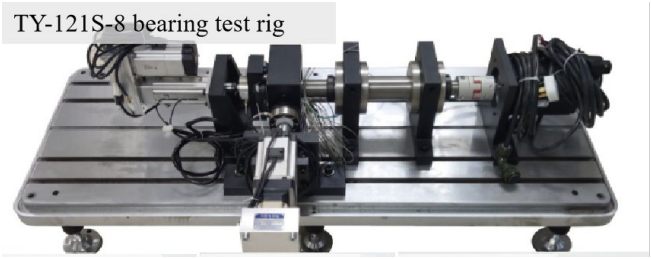

Fig. 11. TY121S-8 bearing test rig.

In this section, the framework proposed in this study is applied to performing successive decomposition and fault diagnosis on the data collected from the TY121S-8 bearing test rig. The experimental setup of this dataset is shown in Fig. 11. Three types of faults are considered: inner race fault, outer race fault, and ball fault. The bearing speeds are set to 1,800, and each signal is sampled at a frequency of 20k Hz. Each sample is divided into 4,096 data points. Considering that fault signals are relatively scarce in various scenarios, only 25 samples are collected for each type of fault. Fig. 12 shows the time and frequency domain plots of each type of fault.

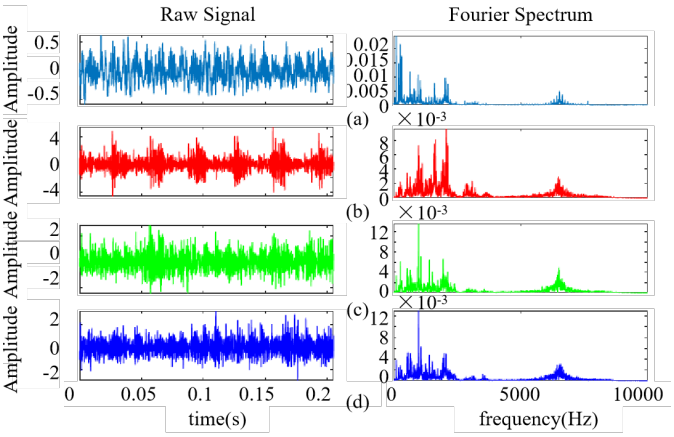

Fig. 12. Time and frequency domain plots of normal and fault signal of TY bearing datasets. (a) Normal signal, (b) Inner fault, (c) Outer fault, and (d) Ball fault.

For the bearings in this dataset, the fault frequencies of the fundamental frequency f_r, inner race fault frequency f_i, outer race fault frequency f_o, and ball fault frequency f_b are shown in TABLE III.

TABLE III. THE CHARACTERISTIC FREQUENCY OF TY DATASET

f_r (Hz)	f_i (Hz)	f_b (Hz)	f_o (Hz)
30	165	63	104.4

Due to the order-invariance property, the order of faults does not affect the decomposition results. Assuming that the normal signal is collected first, followed by the inner race fault, then the ball fault, and finally the outer race fault. The order of the fault samples inputted is shown in Fig. 13.

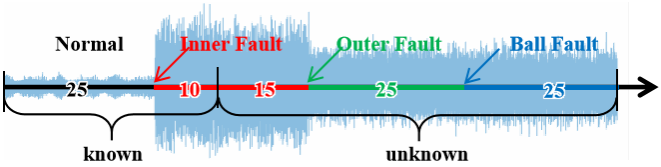

Fig. 13. The sample input order. The normal and partial inner race faults are known initially, and as time progresses, new faults are added to the samples. In this experiment, a maximum of 25 samples are available for each type of fault.

The online updating framework is executed, and the energy vector is visualized during the continuous open-set detection. The results are shown in Fig. 14. It can be observed that as the samples are continuously inputted into the USDMD framework, the energy is dispersed among the existing decomposition layers. The faults are well-clustered in the visualization plot, and the distances between different faults are quite distinct. For new samples whose fault is already learned, their energy is distributed to the corresponding fault class in the visualization plot, and the LPDF value in the corresponding layer is the highest. For faults that have not been learned, their energy distribution in the visualization plot is far from the existing signal classes, and all their LPDF values are less than 1. After ten consecutive samples are identified as new faults, it is confirmed that a new fault has indeed occurred. At this point, the best difference spectrum and filtering index are updated. After all the signals have been inputted, the overall execution of USDMD is completed.

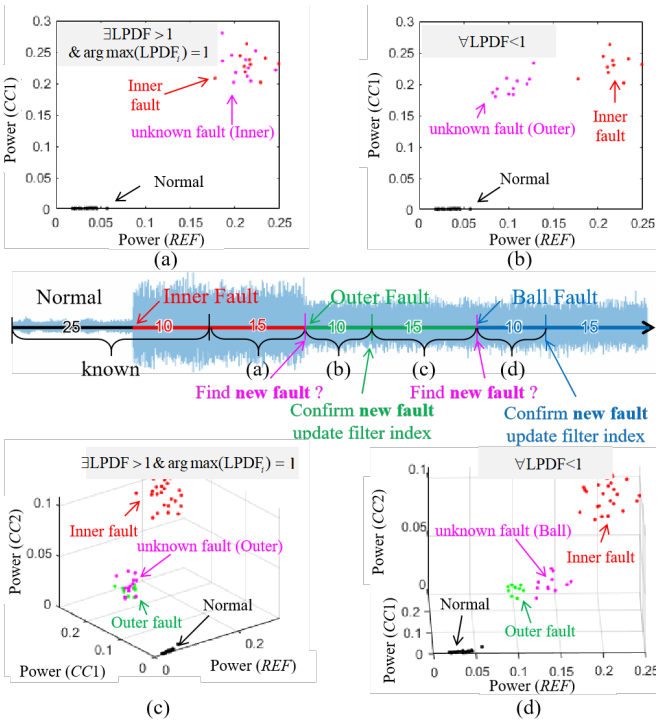

Fig. 14. Visualization of sample energy of TY dataset at each stage by USDMD. In the first stage, (a) USDMD detects unknown learned faults (Inner fault), and (b) USDMD detects unknown and unlearned faults (Outer fault). In the second stage, (c) USDMD detects unknown learned faults (Outer fault), and (d) USDMD detects unknown unlearned faults (Ball fault).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TIM.2025.3601249, IEEE Transactions on Instrumentation and Measurement

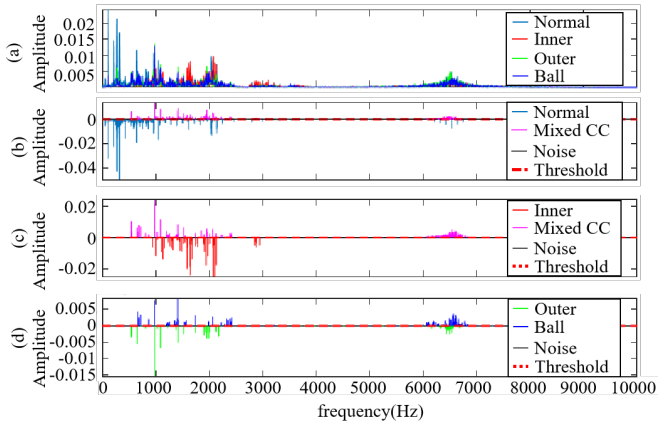

Fig. 15. (a) The comparison between *NFS* of normal and all fault signals, and optimal difference spectrum of (b) layer 1, (c) layer 2, (d) layer 3 of TY datasets.

As shown in Fig. 15, it can be observed that the low-frequency components of the signal are mainly decomposed into normal reference periodic components, while the high-frequency modulation components are continuously divided. In the TY bearing data, the fault components are almost in the same frequency band at high frequencies, and the spectrum lines have a high degree of similarity. In Fig. 15 (c) and (d), it can be observed that the number of filter indexes decreases continuously, gradually becoming sparse, and all of them extract the corresponding modulation frequencies at the high-frequency modulation. As shown in Fig. 16, by applying the obtained optimal difference spectrum, an inverse transform is performed to obtain the decomposition results of the normal and every fault signal. The amplitude of waveforms is more pronounced at the corresponding fault decomposition components, while the other components have very small amplitudes. It indicates that the components of other decomposition layers are very few.

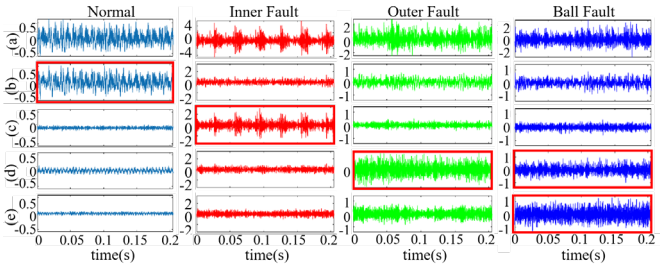

Fig. 16. SDMD results of time domain TY datasets. (a) Original signal, and the components of (b) Normal, (c) Inner fault, (d) Outer fault, and (e) Ball fault.

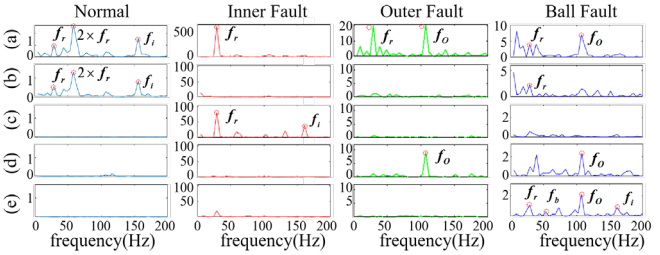

Fig. 17. SDMD results of the square envelope of TY datasets. (a) Original signal, and the components of (b) Normal, (c) Inner fault, (d) Outer fault, and (e) Ball fault.

However, for the ball fault, the amplitudes of the inner race fault and outer race fault components are only slightly smaller than the ball fault decomposition layer. Research has shown that when diagnosing ball faults, features of both inner race and outer race faults are often extracted. This is due to a slippage phenomenon of the rolling elements in the cage [33]. In such cases, the average slippage of the bearing is likely to have self-adjusted and locked onto the subharmonics of the dominant frequency (such as shaft speed). As shown in Fig. 17, the mean square envelope spectrum features of the decomposed signals for the normal, inner race fault, and outer race fault highly coincide with the calculated results. On the other hand, the ball fault frequencies are more prominent at 1.5 and 2 times the fundamental frequency, which correspond to subharmonic frequencies with a 0.5 interval. This demonstrates that the successive decomposition framework performs well in extracting fault components even in the presence of high-frequency modulated signals. Furthermore, due to the orthogonality, even if there are mixed modes (such as the slippage phenomenon in ball faults), they would be separated into different components, which is advantageous for fault diagnosis.

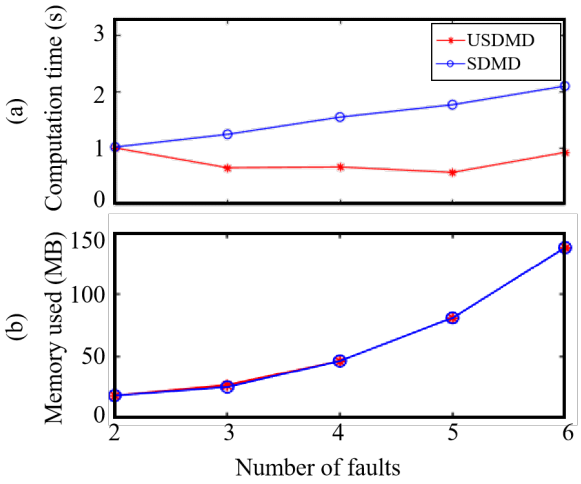

Fig. 18. Comparison of (a) computation time, and (b) memory used between USDMD and SDMD

In addition, the different severity levels of faults in the TY dataset were treated as different faults for comparing the computation time of USDMD and SDMD. As shown in Fig. 18 (a), when the fault number is 2, the USDMD and SDMD computation time are the same. As the number of faults increases, the SDMD calculation time keeps increasing, while the USDMD calculation time fluctuates around the initial value. This is because the training progress of USDMD is fixed at 2 layers of decomposition, while the number of layers of SDMD decomposition is affected by the number of faults. For memory complexity, they are the same, because the successive framework does not change, and the size of the target - ω_{best} of each layer is same. The comparison is shown in Fig. 18 (b). This experiment demonstrates the time efficiency of USDMD in incremental learning.

VII. Conclusion

Building on the DMD, this paper proposes the SDMD framework for extracting multi-class fault features. SDMD

adaptively decomposes fault signals with prior fault knowledge, requiring minimal training data and enabling real-time fault diagnosis with fast Fourier transform and inverse transform. Experimental validation confirms the orthogonal and noise removal capabilities of SDMD, providing frequency-domain interpretability and time-domain discrimination of different faults. Furthermore, an updatable online learning framework USDMD has been proposed for the incremental demand scenario when a new fault occurs, providing data efficiency and interpretability. Experimental validation demonstrates the effectiveness and time efficiency. However, the proposed method can not deal with variable operating conditions. Future work will focus on the improvements for variable operating conditions scenarios.

REFERENCES

[1] E. Zio, "Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice," *Reliab. Eng. Syst. Saf.*, vol. 218, p. 108119, Feb. 2022, doi: 10.1016/j.ress.2021.108119.

[2] P. Kumar, S. Khalid, and H. S. Kim, "Prognostics and health management of rotating machinery of industrial robot with deep learning applications—A review," *Mathematics*, vol. 11, no. 13, Art. no. 13, Jan. 2023, doi: 10.3390/math11133008.

[3] C. Zhang *et al.*, "Time-Domain Sparsity-Based Bearing Fault Diagnosis Methods Using Pulse Signal-to-Noise Ratio," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–4, 2024, doi: 10.1109/TIM.2024.3375978.

[4] O. Matania, I. Dattner, J. Bortman, R. S. Kenett, and Y. Parmet, "A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges," *J. Sound Vib.*, vol. 590, p. 118562, Nov. 2024, doi: 10.1016/j.jsv.2024.118562.

[5] Z. Wang, Z. Zhou, W. Xu, C. Sun, and R. Yan, "Physics informed neural networks for fault severity identification of axial piston pumps," *J. Manuf. Syst.*, vol. 71, pp. 421–437, Dec. 2023, doi: 10.1016/j.jmsy.2023.10.002.

[6] G. Niu, X. Dong, and Y. Chen, "Motor fault diagnostics based on current signatures: A review," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–19, 2023, doi: 10.1109/TIM.2023.3285999.

[7] L. Tang, X. Wu, D. Wang, and X. Liu, "A comparative experimental study of vibration and acoustic emission on fault diagnosis of low-speed bearing," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023, doi: 10.1109/TIM.2023.3312761.

[8] H. Ahmed and A. K. Nandi, *Condition monitoring with vibration signals: Compressive sampling and learning algorithms for rotating machines*. John Wiley & Sons, 2020.

[9] Z. Shang, Z. Zhao, and R. Yan, "Denoising fault-aware wavelet network: A signal processing informed neural network for fault diagnosis," *Chin. J. Mech. Eng.*, vol. 36, no. 1, p. 9, Jan. 2023, doi: 10.1186/s10033-023-00838-0.

[10] M. Civera and C. Surace, "A Comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark," *Sensors*, vol. 21, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/s21051825.

[11] J. Li, W. Luo, and M. Bai, "Review of research on signal decomposition and fault diagnosis of rolling bearing based on vibration signal," *Meas. Sci. Technol.*, vol. 35, no. 9, p. 092001, Jun. 2024, doi: 10.1088/1361-6501/ad4eff.

[12] N. E. Huang, Shen, and Long, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

[13] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, no. 1, pp. 108–126, Feb. 2013, doi: 10.1016/j.ymssp.2012.09.015.

[14] S. Sun, J. Yuan, Q. Zhao, H. Jiang, and J. Zhu, "Endogenous noise-expanded multivariate empirical mode decomposition and its application to mechanical compound fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–11, 2025, doi: 10.1109/TIM.2025.3544739.

[15] Dragomiretskiy and Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.

[16] Nazari and Sakhaei, "Successive variational mode decomposition," *Signal Process.*, vol. 174, p. 107610, Sep. 2020, doi: 10.1016/j.sigpro.2020.107610.

[17] M. Chen *et al.*, "Optimized variational mode decomposition algorithm based on adaptive thresholding method and improved whale optimization algorithm for denoising magnetocardiography signal," *Biomed. Signal Process. Control*, vol. 88, p. 105681, Feb. 2024, doi: 10.1016/j.bspc.2023.105681.

[18] Y. Miao, B. Zhang, C. Li, J. Lin, and D. Zhang, "Feature mode decomposition: New decomposition theory for rotating machinery fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 1949–1960, Feb. 2023, doi: 10.1109/TIE.2022.3156156.

[19] S. Chauhan, G. Vashishtha, R. Kumar, R. Zimroz, M. K. Gupta, and P. Kundu, "An adaptive feature mode decomposition based on a novel health indicator for bearing fault diagnosis," *Measurement*, vol. 226, p. 114191, Feb. 2024, doi: 10.1016/j.measurement.2024.114191.

[20] D. Wang, "Difference mode decomposition for adaptive signal decomposition," *Mech. Syst. Signal Process.*, vol. 191, p. 110203, May 2023, doi: 10.1016/j.ymssp.2023.110203.

[21] J. Guo, Y. Liu, R. Yang, W. Sun, and J. Xiang, "Ensemble difference mode decomposition based on transmission path elimination technology for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 212, p. 111330, Apr. 2024, doi: 10.1016/j.ymssp.2024.111330.

[22] J. Guo, Y. Liu, R. Yang, W. Sun, and J. Xiang, "A simulation-driven difference mode decomposition method for fault diagnosis in axial piston pumps," *Adv. Eng. Inform.*, vol. 62, p. 102624, Oct. 2024, doi: 10.1016/j.aei.2024.102624.

[23] D.-W. Zhou, Q.-W. Wang, Z.-H. Qi, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Class-incremental learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9851–9873, Dec. 2024, doi: 10.1109/TPAMI.2024.3429383.

[24] S. Yan, H. Shao, X. Wang, and J. Wang, "Few-shot class-incremental learning for system-level fault diagnosis of wind turbine," *IEEEASME Trans. Mechatron.*, pp. 1–10, 2024, doi: 10.1109/TMECH.2024.3490733.

[25] C. Teng, Z. Shang, X. Bai, and R. Yan, "Successive difference mode decomposition for rotating machine fault diagnosis," in *2024 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, Oct. 2024, pp. 1–7. doi: 10.1109/ICSMD64214.2024.10920554.

[26] Hou, "Interpretable online updated weights: Optimized square envelope spectrum for machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 169, p. 108779, Apr. 2022, doi: 10.1016/j.ymssp.2021.108779.

[27] B. W. Silverman, "On a Gaussian process related to multivariate probability density estimation," *Math. Proc. Camb. Philos. Soc.*, vol. 80, no. 1, pp. 135–144, 1976, doi: 10.1017/S0305004100052762.

[28] M. Li, H. Wang, and G. Tang, "An improved method based on CEEMD for fault diagnosis of rolling bearing," *Adv. Mech. Eng.*, 2014, doi: 10.1155/2014/676205.

[29] T. Yan, D. Wang, and Y. Wang, "Discrimination- and sparsity-driven weight-oriented optimization model for interpretable initial fault detection and fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024, doi: 10.1109/TIM.2023.3335512.

[30] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features," *Mech. Syst. Signal Process.*, vol. 99, pp. 459–477, Jan. 2018, doi: 10.1016/j.ymssp.2017.06.027.

[31] Z. Zhao, S. Wang, B. An, Y. Guo, and X. Chen, "Hierarchical hyper-Laplacian prior for weak fault feature enhancement," *ISA Trans.*, vol. 96, pp. 429–443, Jan. 2020, doi: 10.1016/j.isatra.2019.06.007.

[32] G. Han, J. Shen, Z. Wang, Y. Zhu, and Y. Xie, "A novel dual-domain adversarial method for vibration signal denoising in bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025, doi: 10.1109/TIM.2025.3551836.

[33] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64–65, pp. 100–131, Dec. 2015, doi: 10.1016/j.ymssp.2015.04.021.