Mechanical Systems and Signal Processing 202 (2023) 110664

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network

Xiaoan Yan [a,b,**], Wang-Ji Yan [a,c,*], Yadong Xu [d], Ka-Veng Yuen [a,c]

[a] State Key Laboratory of Internet of Things for Smart City and Department of Civil and Environmental Engineering, University of Macau, Macau SAR 999078, China
[b] School of Mechatronics Engineering, Nanjing Forestry University, Nanjing 210037, China
[c] Guangdong–Hong Kong-Macau Joint Laboratory for Smart Cities, University of Macau, Macau SAR 999078, China
[d] School of Mechanical Engineering, Southeast University, Nanjing 211189, China

ARTICLE INFO

Keywords:
Multivariate feature mode decomposition
Multi-attention fusion residual convolutional neural network
Multi-sensor data
Machinery fault diagnosis

ABSTRACT

Due to the complex and rugged working environment of real machinery equipment, the resulting fault information is easily submerged by severe noise interference. Additionally, some informative features may be omitted if the feature learning concerns only a single sensor of machinery vibration data. Therefore, to mine more comprehensive fault information and achieve more robust fault diagnosis results, this study proposes a machinery multi-sensor fault diagnosis method based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network. As an extension of the feature mode decomposition (FMD), the adaptive multivariate feature mode decomposition (AMFMD) with the improved whale optimization algorithm (IWOA) is firstly presented to automatically decompose the collected multi-sensor vibration data into a group of multichannel mode components, which both inherit the anti-noise robustness of the original FMD and overcome the obstacles of artificial parameter selection of FMD. Subsequently, multichannel mode components containing the most abundant fault information are selected via an impulse sensitive measure hailed as multichannel comprehensive index (MCI), and the frequency slice wavelet transform (FSWT) of the selected multichannel mode components is further calculated and organically fused to generate the colored multichannel time–frequency representation (MTFR) containing multi-sensor important signatures. Finally, by integrating the advantages of feature learning of residual network (ResNet) and convolutional neural network (CNN), a multi-attention fusion residual convolutional neural network (MAFResCNN) with squeeze-excitation module (SEM) and convolutional block attention module (CBAM) is constructed to simultaneously capture global and local feature information from the fused multichannel time–frequency representation and implement automatic discrimination of machinery fault states, which can both enhance machinery fault information and whittle down the useless information, even promote the feature learning performance without significantly increasing the computational burden of the model. The validity of the proposed approach is verified by a diagnosis case of a real wind turbine, demonstrating that the proposed approach has superiority in machinery fault identification compared with some similar techniques.

---

* Corresponding author.
** Co-corresponding author.
E-mail addresses: yanxiaoan89@sina.com (X. Yan), civilyanwj@gmail.com (W.-J. Yan).

https://doi.org/10.1016/j.ymssp.2023.110664
Received 25 May 2023; Accepted 3 August 2023
Available online 19 August 2023
0888-3270/© 2023 Elsevier Ltd. All rights reserved.

1. Introduction

As one of the key components of machinery equipment, the running state of the bearing, gear or rotor directly affects the performance of whole machinery equipment. Due to the complex structure of machinery equipment and harsh working environment, these key parts (e.g., bearing, gear, or rotor) often have various faults, resulting in direct or indirect economic losses. When a bearing, gear, or rotor is damaged, mechanical vibration signal collected often presents non-stationary characteristics and will inevitably be contaminated by strong noise in practical engineering, which indicates that it is very difficult to effectively identify machinery fault states [1–5]. Therefore, it is very necessary to carry out high-efficiency feature extraction and intelligent fault diagnosis for rolling bearings.

Even to this day, non-stationary signal processing algorithms have been widely used in the field of mechanical fault diagnosis, such as short-time Fourier transform (STFT), Wigner-Ville distribution (WVD), and continuous wavelet transform (CWT). However, these time–frequency analysis methods need to select appropriate window functions or wavelet basis functions [6,7]. To address this issue, based on empirical mode decomposition (EMD) [8], a series of adaptive signal decomposition methods are presented and rapidly expanded, such as local mean decomposition (LMD) [9], local characteristic-scale decomposition (LCD) [10], symplectic geometry mode decomposition (SGMD) [11], variational mode decomposition (VMD) [12–15], and fault information-guided variational mode decomposition (FIVMD) [16]. Nevertheless, these methods generally have shortcomings such as mode aliasing, and end effect, etc. Based on the above background, to avoid the shortcomings of existing signal decomposition algorithms and improve fault diagnosis performance, inspired by adaptive signal decomposition theory and deconvolution principle, Miao et al. [17] proposed a new non-stationary signal decomposition theory entitled feature mode decomposition (FMD). The FMD method can decompose the non-stationary multi-component signal into several mono-components by iteratively updating the filter coefficients, establishing the adaptive FIR filter bank, and estimating the fault period of the measured signal based on the autocorrelation function. Therefore, FMD not only considers the periodicity and impact of signals, but also has certain robustness to noise interference signals and remove the dependence on the prior knowledge of the signal fault period. Nevertheless, like VMD, FMD does not have parameter self-adaptability. Two important parameters (i.e., mode number and filter size) must be manually set in advance. Specifically, if the parameters of FMD are set unreasonably, their performance and efficiency will be significantly affected. Additionally, FMD cannot process multi-channel signal synchronously at present. Regarding these facts, inspired by the success of previous multi-channel data analysis methods (e.g., multivariate time series model [18], multivariate empirical mode decomposition (MEMD) [19], multivariate local characteristic-scale decomposition (MLCD) [20] and multivariate variational mode decomposition (MVMD) [21]), a new multichannel signal analysis method defined as adaptive multivariate feature mode decomposition (AMFMD) is proposed to process multi-sensor signal. Similar to MEMD and MVMD, AMFMD essentially works as a dyadic filtering bank on each channel of multi-sensor input signal. In other words, AMFMD amounts to an extended version of FMD, which can fill a gap of FMD in multivariate signal processing and help resolve the manual parameter selection issue in the original FMD method via the guidance of parameter optimizer. Moreover, different from FMD, AMFMD can align the corresponding mode component from different channels across the same frequency range.

Although using the non-stationary signal processing methods can obtain rich feature information in the multichannel time domain, frequency domain, or time–frequency domain, it is very dependent on the use of expert knowledge for judging fault category, that is, it does not have intelligence. To remedy this deficiency, many scholars adopt various intelligent classifiers to deal with those results obtained by the representative non-stationary signal processing methods, which are crucial for intelligent fault identification in real engineering. For instance, Zuo et al. [22] adopted LMD to decompose the raw vibration signal and extract six dimensionless features as the import of a spiking neural network (SNN) to identify bearing fault categories. Yao et al. [23] first utilized LCD for signal - decomposition and extracted multi-scale permutation entropy (MPE) as a feature vector, and then introduced an extreme learning machine (ELM) to achieve fault classification. Jin et al. [24] adopted the optimized VMD to decompose and reconstruct the signal, then extract features of the reconstructed signal, and finally, an adaptive mutation particle swarm optimization random forest (RF) is used for bearing fault pattern identification. Zhang et al. [25] used an improved particle swarm optimization variational mode decomposition (PSO-VMD) to decompose the raw vibration signal and calculate multi-scale information entropy and applied the support vector machine (SVM) for bearing intelligent fault diagnosis. Chen et al. [26] applied VMD to decompose the raw vibration signal into several mode components and calculated the energy entropy of the first four mode component as the input features of SVM to achieve fault classification. Nevertheless, the above classification models (e.g., ANN, SVM, ELM and RF) belong to shallow architecture, which indicates that they are difficult to learn the fault information covered by heavy noise. Besides, fault identification ability of these shallow models strongly depends on the expert knowledge and artificial-extracted features. That is, their robustness and generality are insufficient when facing strong noise and data imbalance.

To overcome these weaknesses, the deep learning or transfer learning-based intelligent classification model is arousing more and more interest in fault diagnosis aspects and a large number of achievements are presented [27–30]. Wang et al. [31] first used a short-time Fourier transform (STFT) to generate a time–frequency map, then CNN is employed to further learn deeply fault information from the corresponding time–frequency map and perform motor fault diagnosis automatically. Cheng et al. [32] first calculated continuous wavelet transform (CWT) of bearing vibration signal to obtain time–frequency representation, and then the acquired time–frequency representation is input into CNN to finish intelligent fault identification. Kumar et al. [33] used analytical wavelet transform (AWT) to get the greyscale images of the collected signal, and then the obtained greyscale images are fed into the improved CNN to identify defects in the centrifugal pump. Islam et al. [34] applied wavelet packet transform (WPT) to extract two-dimensional representation information and combined CNN to automatically identify bearing fault types. Gu et al. [35] first used VMD to preprocess and reconstruct bearing vibration data, then adopted CWT to obtain a two-dimensional time–frequency image of the reconstructed data, and finally constructed a CNN model to perform model training and fault classification. Liang et al. [36] firstly average descent rate singular value decomposition (ADR-SVD) to preprocess mechanical vibration signal, then the noise-reducing signal is converted to the two-dimensional image as the input of residual neural network for fault classification. Guo et al. [37] adopted the Hilbert-Huang transform (HHT) to extract time–frequency energy images as the input of CNN to identify fault categories of power distribution systems. Wang et al. [38] adopted morphological erosion operation to convert the collected time-domain vibration signals into images, then applied CNN to obtain fault classification results. Ke et al. [39] calculated the synchrosqueezing transform (SST) of the original signals to obtain time–frequency representations as the input of CNN to further learn the underlying features and conduct intelligent fault diagnosis. Chen et al. [40] first adopted Cyclic Spectral Coherence (CSCoh) to construct two-dimensional feature maps, and then used convolutional neural networks (CNN) to learn the helpful feature representation and achieve fault identification. Kim et al. [41] embedded a health-adaptive time-scale representation (HTSR) into CNN to learn fault-related information and accomplish high-efficiency gearbox fault diagnostics. Zhu et al. [42] used WT to convert the sample data into time–frequency representation (TFR) and presented a multiscale convolutional neural network (MSCNN) to learn feature information hidden in bearing faults. Liu et al. [43] embedded residual learning into multiscale kernel CNN to construct a multiscale kernel-based residual convolutional neural network (MK-ResCNN) for automatic feature learning and motor fault identification. Zhao et al. [44] first employed compressed sensing to reconstruct the collected signal and converted the reconstructed signal into a two-dimensional signal image, and then designed a multiscale inverted residual convolutional neural network (MIRCNN) to identify bearing faults. Ravikumar et al. [45] first combined CNN and residual learning to extract fault features, and then presented a stacked long short-term memory for gearbox fault identification. Wei et al. [46] first used continuous wavelet transform (CWT) to convert the original vibration signal into time–frequency images, and then the residual convolutional neural network (ResCNN) with extreme learning machine is applied to extract features and execute fault discrimination. Unfortunately, there are some shortcomings in the above-reported research. On the one hand, the above-mentioned classification model only focuses on feature learning of vibration signals collected at a single sensor, whereas the channel information of other sensors is ignored, this easily leads to unsatisfactory diagnostic accuracy due to the acquisition of insufficient information. On the other hand, most of the existing deep learning models are constructed via stacking a large number of convolution layers, which easily trap into local optimization and even generate a certain degree of performance degradation problems due to the limitations of current training methods. Additionally, as an improved version of CNN, although ResCNN has the advantages that the gradient information is easier to be propagated and optimized in the process of back-propagation and the problem of gradient vanishing caused by adding a network layer can be alleviated, the gradient exploding phenomenon sometimes appear in ResCNN with direct identity mapping, thus easily affecting the normal training of network. Also, the ability of traditional ResCNN for mining global and local feature information is not strong. Consequently, to address these challenges, a new model named multi-attention fusion residual convolutional neural network (MAFResCNN) is formulated by conducting the integration of the squeeze-excitation module (SEM) and convolutional block attention module (CBAM) in the inside of ResCNN architecture, which can fully utilize global and local information of the input data and avoid performance degradation problem. Popularly speaking, MAFResCNN can both enhance the differentiated feature information and weaken the useless feature information contained in the input data.

In a capsule, against the above background, this study proposes a new approach based on adaptive multivariate feature mode decomposition (AMFMD) and multi-attention fusion residual convolutional neural network (MAFResCNN) for machinery multi-sensor fault diagnosis. The overall ideology of the pretended approach is to generate the fused multichannel time–frequency representation (MTFR) based on AMFMD that includes multi-sensor important information, while automatically learning the discriminant features from the attainable MTFR and using the constructed MAFResCNN model to achieve a high-quality diagnostic result. Specifically, our proposed method can make full use of the joint feature information between multi-sensor signals by considering the collected multi-sensor data located in multiple channels simultaneously. Additionally, AMFMD is equipped with mode alignment property. In other words, multichannel mode components obtained by AMFMD have a similar pattern of oscillation at each channel. Taken overall, the contributions and novelties of our study are summarized as follows:

- Adaptive multivariate feature mode decomposition (AMFMD) with the improved whale optimization algorithm (IWOA) is proposed for multi-sensor non-stationary data analysis, which can not only extend FMD to multivariate signal analysis but also avoid the empirical parameter setting of FMD in signal decomposition.
- Multi-attention fusion residual convolutional neural network (MAFResCNN) is presented to handle and refine multichannel time–frequency representation, which can not only inherit the powerful self-learning ability of ResCNN but also consider global and local feature information in channel and spatial.
- A machinery multi-sensor intelligent diagnosis framework based on AMFMD and MAFResCNN is proposed by considering multi-sensor information fusion, which can realize the automatic identification of machinery faults and promote recognition accuracy without increasing the number of samples.

The remainder of this study is organized as follows. Section 2 introduces the theoretical foundation of AMFMD and compares MFMD, MEMD and MVMD using multivariate simulation signal. Section 3 constructs a multi-attention fusion residual convolutional neural network (MAFResCNN) via reviewing the squeeze-excitation module (SEM) and convolutional block attention module (CBAM). Section 4 depicts the overall technology roadmap of the proposed fault diagnosis framework. Section 5 validates the effectiveness and superiority of the presented approach through case studies and contrastive analysis, and some discussion and research perspectives are also given. The conclusion part is finally displayed in Section 6.

2. Adaptive multivariate feature mode decomposition

2.1. Feature mode decomposition

Feature mode decomposition (FMD) is a non-stationary signal decomposition approach, which is composed of several steps (e.g., adaptive FIR filter bank design, filtering coefficients updating, fault period estimation and mode selection). FMD provides different components simultaneously via initializing an FIR filter bank and updating the filtering coefficients. The fundamental theory of FMD is briefly summarized as follows [17]:

(1) Load one-dimensional original time series × and preset the parameters of FMD, such as the required mode number K, the filter size L, and the max iteration number I.
(2) Initialize FIR filter banks by utilizing M Hanning windows and set the initial iteration $i = 1$.
(3) Take advantage of the formula $u_m^i = x \otimes f_m^i, m = 1, 2, \cdots, M$ to acquire the decomposed mode component, where $\otimes$ denotes the convolution operator.
(4) According to the input one-dimensional time series ×, the decomposed mode components u_m^i and the estimated fault period T_m^i, the filter coefficients are updated, where T_m^i is the time delay corresponding to the local maximum value of auto-correlation spectrum R_{mc}^i after the first zero-crossing point.
(5) Judge if the current iteration number achieves the required maximum iteration number I. If not, go back to step 3 to continue the iteration process. Otherwise, go to step 6.

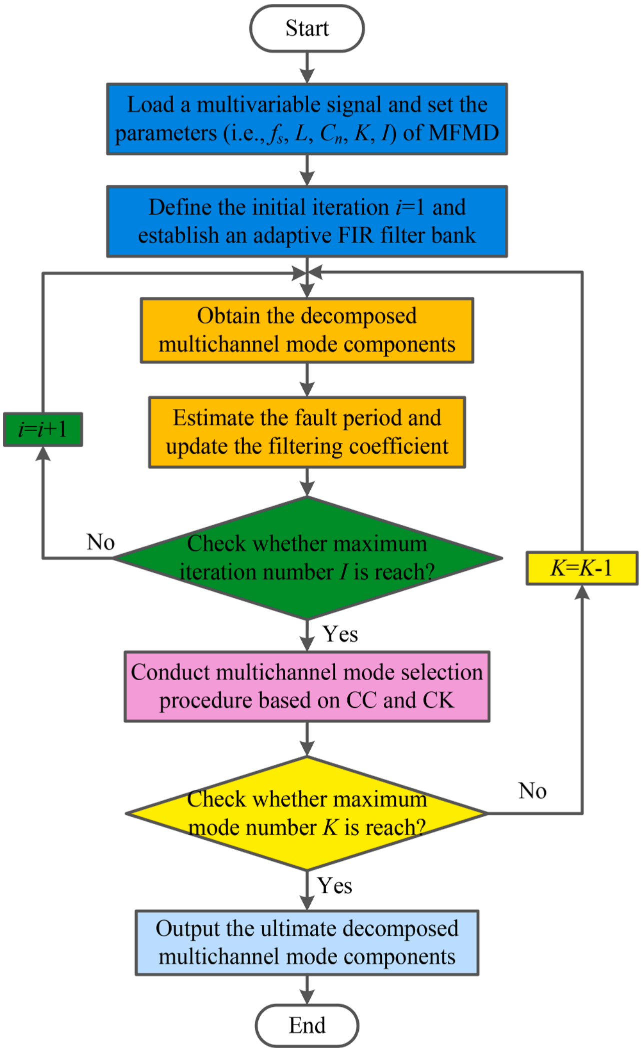


Fig. 1. Flowchart of MFMD method.

(6) The correlation coefficient matrix $CC_{(M\times M)}$ with $M \times M$ is first obtained by calculating the correlation coefficient between two adjacent mode components. Then, two adjacent mode components with the maximum correlation coefficient are selected. Meanwhile, based on the estimated fault period T^i_m, the correlation kurtosis of the selected mode components is estimated. Finally, one mode component with the larger correlation kurtosis (CK) is treated as the mode component of FMD and define $M = M-1$.

(7) Judge if the current mode number M reaches the required mode number K. If not, go back to step 3. Otherwise, terminate the decomposition process and output the final decomposition results. Details of FMD are available in [17].

2.2. Multivariate feature mode decomposition

In this section, to extend the original FMD to process multivariate signals, a new approach termed as multivariate feature mode decomposition (MFMD) without extra parameters is presented by integrating parallel multichannel data analysis into FMD. In other words, MFMD is essentially a supplement of FMD, which has mode alignment properties in multivariate signal processing and can obtain meaningful joint feature information related to multichannel mode components. The primary target of MFMD is to naturally extract multichannel mode components directly in the multidimensional space of the input multivariate signal. Fig. 1 plots the flowchart of MFMD. Implementation details of MFMD are given below:

(1) Load a multivariable signal $X(n) = [x_1(n), x_2(n), \cdots, x_o(n)], n = 1, 2, \cdots, N$ with o channels and set the parameters (i.e.,f_s, L, C_n, K, I) of MFMD, where $x_o(n) = [x(1), x(2), \cdots, x(N)]^T, o = 1, 2, \cdots$ represents the o-th channel signal, $x(N)$ means the N-th data point of $x_o(t)$,f_s denotes the sampling frequency, L represents the filter size, C_n denotes the cut number of the whole frequency band of the o-th channel signal, K means the extracted maximum mode number, and I means maximum iteration number.

(2) Define the initial iteration $i = 1$ and establish an adaptive FIR filter bank. Specifically, for each mono-variate signal, use M Hanning windows to construct a group of uniformly distributed adaptive FIR filters.

(3) Obtain the decomposed multichannel mode components. Specifically, use $\widehat{u}^i_{mo} = X_o \otimes \widehat{f}^i_{mo}$ to simultaneously acquire the decomposed multichannel mode components, where $\widehat{u}^i_{mo}$ is the decomposed m-th mode components of the o-th channel signal at the i-th iteration, X_o means the matrix which is composed of M row segments with the same filtering size L, $\widehat{f}^i_{mo}$ is the m-th FIR filter of the o-th channel signal at the i-th iteration, $m = 1, 2, \cdots, M$ and $\otimes$ implies the convolution operator.

(4) Estimate the fault period and update the filtering coefficient. Specifically, for each mono-variate signal, use the auto-correlation spectrum R^i_{mo} of multichannel mode components to determine the fault period T^i_{mo}, and then update the filtering coefficient based on the estimated fault period and the above-decomposed mode components $\widehat{u}^i_{mo}$. More exactly, $R^i_{mo}(\tau) = \int \widehat{u}^i_{mo}(\widehat{n}_o)\widehat{u}^i_{mo}(\widehat{n}_o + \tau)d\widehat{n}, \widehat{n}_o = 1, 2, \cdots, N-L+1$ represents the auto-correlation spectrum of $\widehat{u}^i_{mo}$, T^i_{mo} is the estimated fault period of $\widehat{u}^i_{mo}$ which can be selected as the time delay corresponding to the local maximum value of the auto-correlation spectrum R^i_{mo} after the first zero-crossing point.

(5) Examine whether maximum iteration number I is reach. If $i = I$, terminate the iteration procedure and continue the next step (5). Otherwise, set $i = i + 1$ and go back to step (3) to run again.

(6) Conduct multichannel mode selection. Specifically, for each mono-variate signal, according to the following Eq. (1), calculate the correlation coefficient (CC) between all two adjacent multichannel mode components to construct multichannel correlation coefficient matrix $MCC_{(M\times M\times o)}$ and lock two multichannel mode components with the largest CC, and then calculate the correlated kurtosis (CK) of the locked two multichannel mode components and remove multichannel mode components with smaller CK and define $M = M-1$.

$$CC_{m\times(m-1)\times o} = \frac{\sum(\widehat{u}^i_{mo} - mean(\widehat{u}^i_{mo}))(\widehat{u}^i_{(m-1)o} - mean(\widehat{u}^i_{(m-1)o}))}{\sqrt{\sum\left(\widehat{u}^i_{mo} - mean\left(\widehat{u}^i_{mo}\right)\right)^2}\sqrt{\sum\left(\widehat{u}^i_{(m-1)o} - mean\left(\widehat{u}^i_{(m-1)o}\right)\right)^2}} \tag{1}$$

where $CC_{m\times(m-1)\times o}$ represents the correlation coefficient between the m-th and (m-1)-th mode components between at the o-th channel, $mean(\widehat{u}^i_{mo})$ and $mean(\widehat{u}^i_{(m-1)o})$ represent the mean value of the m-th and (m-1)-th mode components at the o-th channel, respectively.

(7) Check if the stopping condition is satisfied. Specifically, if the maximum number of modes K is reached, terminate the iteration procedure and output the ultimate multivariable signal decomposition results. In other words, by doing the above step, K-order multichannel mode components $U^{(k)}(n)$ of the input multivariable signal $X(n)$ can be obtained synchronously, where $U^{(k)}(n) = [\widehat{u}^{(k)}_1(n), \widehat{u}^{(k)}_2(n), \cdots, \widehat{u}^{(k)}_o(n)], k = 1, 2, \cdots, K$ and $\widehat{u}^{(k)}_o(n)$ denotes the k-th decomposed mode component at the o-th channel.

2.3. Adaptive parameter selection of MFMD

In a similar fashion as FMD, when MFMD is applied to handle multi-sensor data, its two key parameters (i.e., mode number K and filter size L) are also subject to manual selection. In other words, MFMD is not equipped with parameter adaptation. To tackle this issue, inspired by the inertia weight of particle swarm optimization (PSO), the improved whale optimization algorithm (IWOA) with nonlinear convergence factor is presented to determine automatically the important parameters of MFMD, which can effectively balance the global search and local search capabilities of the original whale optimization algorithm (WOA) [47]. Meanwhile, a new health index named the sparsity-complexity integration measure (SCIM) is proposed as the fitness function of IWOA to assist the parameter optimization process of MFMD. Fig. 2 shows the flowchart of adaptive parameter selection of MFMD with IWOA. The specific parameter selection process of MFMD can be described as follows:

(1) Initialize the population and preset the parameters of IWOA. Concretely, define the population size $N = 50$, the maximum number of iterations $T = 30$. Considering that two important parameters (i.e., mode number K and filter size L) of MFMD need to be artificially selected, a vector $X_i = [K, L]$ is utilized to depict the position of whales, where K implies the decomposed mode number of MFMD, L denotes the filter size of MFMD. The upper and lower bound of the vector X_i are fixed at [6, 100] and [2, 10], respectively.

(2) Estimate the fitness value of whales and determine the current optimal location of whales. At present, many indices of mechanical health monitoring are published, such as kurtosis [48], cyclic kurtosis [49], Gini index [50], smoothness index [51], harmonic-to-noise-ratio (HNR) [52], and Shannon entropy [53] and so on. Although mostly sparsity measures (e.g., kurtosis, cyclic kurtosis, Gini index, smoothness index, and HNR) can be effectively used for characterizing the impulsiveness and sparseness of repetitive transients related to bearing or gear faults, the complexity or uncertainty of mechanical vibration signal is too often neglected, whereas Shannon entropy provides an effective way to solve the problem of quantitative of complexity or uncertainty of information contained in a signal. Therefore, in this step, inspired by cyclic kurtosis and Shannon entropy, a novel health indicator nominated sparsity-complexity integration measure (SCIM) is presented as the fitness value to guide the parameter optimization process of MFMD, which can consider simultaneously the measurement of the sparsity and complexity of mechanical vibration signals. For one signal with a period T, the SCIM index can be formulated by Eq. (2).

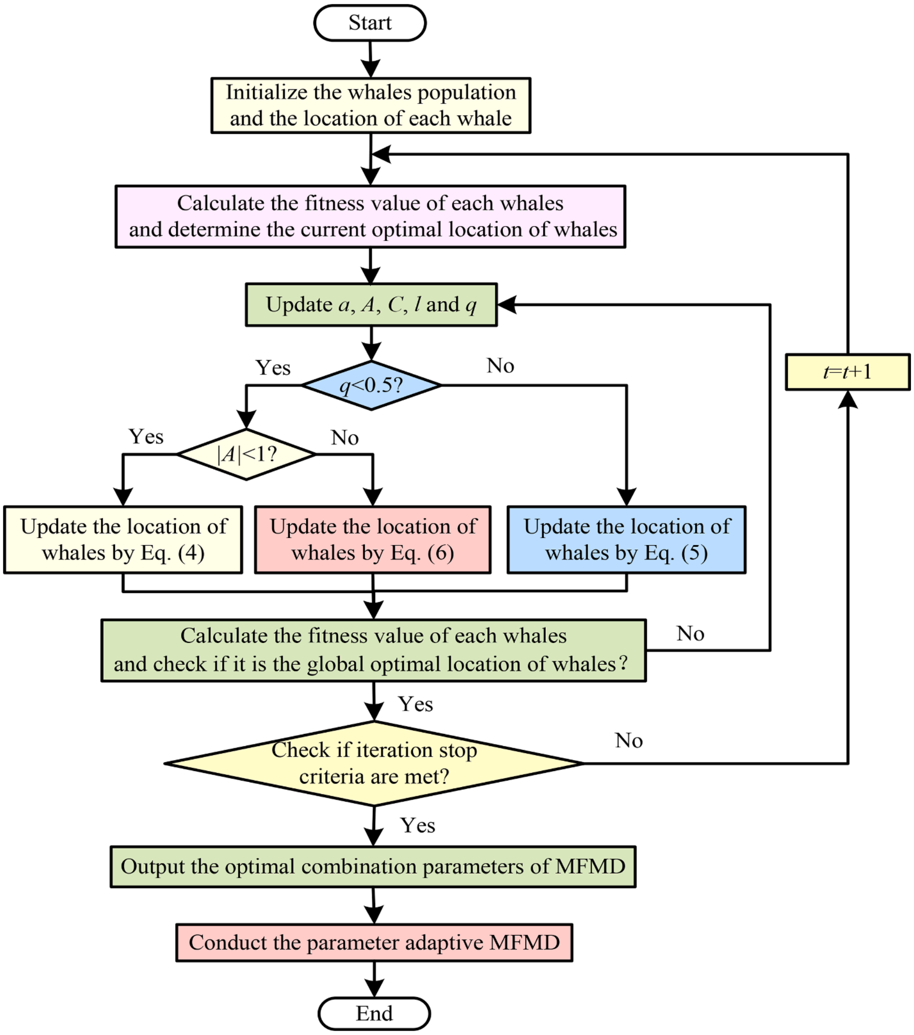


Fig. 2. The flowchart of adaptive parameter selection of MFMD with IWOA.

$$\begin{cases} SCIM(x,T) = -\sum_{i=1}^{M}(p_i(x,T)\log_2 p_i(x,T)) = -\sum_{i=1}^{M}\left(\frac{R_{SE(x)}(iT)}{\sum_{i=1}^{M}R_{SE(x)}(iT)}\log_2\frac{R_{SE(x)}(iT)}{\sum_{i=1}^{M}R_{SE(x)}(iT)}\right)\propto\frac{\sum_{i=1}^{M}R_{SE(x)}(iT)}{R_{SE(x)}(iT)} \\ \qquad = \frac{\sum_{i=1}^{M}R_{SE(x)}(iT)/[R_x(0)]^2}{R_{SE(x)}(iT)/[R_x(0)]^2} = \frac{CK(x,T)}{R_{SE(x)}(iT)/[R_x(0)]^2}\propto CK(x,T) \\ \qquad \sum_{i=1}^{M}p_i(x,T) = \sum_{i=1}^{M}\frac{R_{SE(x)}(iT)}{\sum_{i=1}^{M}R_{SE(x)}(iT)} = \frac{\sum_{i=1}^{M}R_{SE(x)}(iT)}{\sum_{i=1}^{M}R_{SE(x)}(iT)} = 1 \end{cases} \tag{2}$$

where $R_{SE(x)}(iT)$ represents the amplitude of unbiased autocorrelation of the squared envelope of the signal $x(n)$ at the i-th period; $\sum_{i=1}^{M}R_{SE(x)}(iT)$ denotes the overall amplitude of unbiased autocorrelation of squared envelope of the signal $x(n)$ at all periods; $p_i(x,T)$ is the probability of $R_{SE(x)}(iT)$ versus $\sum_{i=1}^{M}R_{SE(x)}(iT)$;$SE(x)$ indicates the squared envelope signal of the signal $x(n)$ and satisfy $SE(x) = |x|^2 = |x + j\cdot Hilbert(x)|^2$; $Hilbert(\cdot)$ signifies the Hilbert transformation operator; $R_{SE(x)}$ means the autocorrelation function of $SE(x)$ and satisfies $R_{SE(x)} = \int SE(x)SE(x+\tau)dt$; $iT(i = 1, 2, \cdots, M)$ implies the time delay corresponding to the i-th period of the signal $x(n)$; and $CK(x,T)$ represents cyclic kurtosis of the signal $x(n)$. Due to the larger time delay iT usually corresponds to the smaller amplitude of the autocorrelation function, so M in this paper is set as 3. Formally speaking, the formulated SCIM can be considered as the Shannon entropy of unbiased autocorrelation of the squared envelope of the signal $x(n)$. More intuitively, the SCIM is proportional to the cyclic kurtosis $CK(x,T)$ of the signal $x(n)$. In other words, the larger the SCIM is, the richer the obtained impulsive information related to mechanical faults is, the less the complexity of the obtained signal is, and the better the signal decomposition performance of MFMD will be. In other words, the parameter selection procedure of MFMD can be interpreted as the optimization problem of searching the largest SCIM. Consequently, the fitness function of IWOA in adaptive parameter selection of MFMD is represented by Eq. (3).

$$\underset{X_i=(K,L)}{\operatorname{argmax}}\{SCIM_{X_i}\},\ subject\ to\ K\in[2,6]\ and\ L\in[10,100] \tag{3}$$

where $SCIM_{X_i}$ indicates the SCIM of the decomposed multichannel mode components under different combination parameters $X_i = [K,L]$. To demonstrate the effectiveness of the formulated SCIM in evaluating impulsive features, four simulated signals (see Fig. 3) are generated. Meanwhile, four health indicators (i.e., kurtosis, cyclic kurtosis, Shannon entropy and SCIM) of each simulated signal are calculated to compare their performance in evaluating fault feature information and the calculated results are plotted in Fig. 4. As seen from Fig. 4, for signal 4 (periodic impulses), three health indicators (i.e., kurtosis, cyclic kurtosis and SCIM) can obtain the largest normalized amplitude, which shows that they are more sensitive to periodic impulses related to faults compared with Shannon entropy. For signal 2 (stochastic noise) and signal 3 (single impulse), the normalized amplitude of Shannon entropy is the smallest, which indicates that Shannon entropy is less susceptible to stochastic noise and the single impulse. Meanwhile, for signal 3 (single impulse), the normalized amplitude of SCIM is smaller than kurtosis and cyclic kurtosis, which means that SCIM is less sensitive to single abnormal impulse than kurtosis and cyclic kurtosis. For signal 1 (harmonic signal), the normalized amplitude of three health indicators (i.e., kurtosis, cyclic kurtosis and SCIM) are less than that of Shannon entropy. In other words, three health indicators (i.e., kurtosis, cyclic kurtosis and SCIM) have better harmonic resistance ability. Therefore, according to this comparison, we can conclude that the SCIM synthesizes the feature evaluation ability of cyclic kurtosis and Shannon entropy, which is propitious to the assessment of fault-related features and is able to measure impulsive information from sparsity and complexity simultaneously.

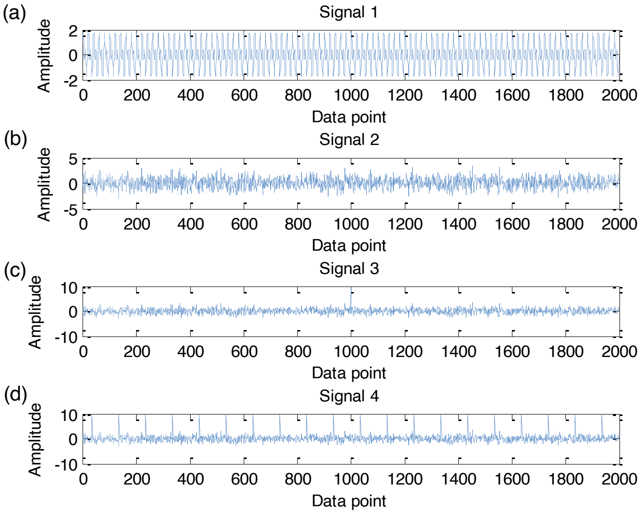


Fig. 3. Four simulated signals: (a) Harmonic signal, (b) stochastic noise, (c) single impulse and (d) periodic impulses.

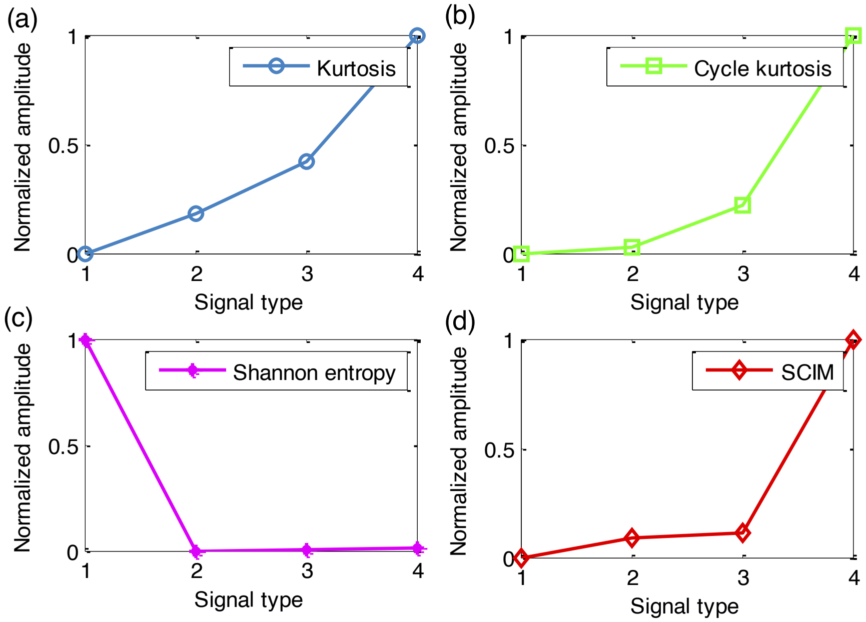


Fig. 4. Health indicator comparison of different simulated signals: (a) kurtosis, (b) cyclic kurtosis, (c) Shannon entropy and (d) SCIM.

(3) For each iteration, update the parameters a, A, C, l and q. If $q < 0.5$, we will select the location updating formula of shrinking encircling mechanism of whales. If not, we will select the location updating formula of spiral model of whales. In other words, there is the same probability of choosing a shrinking encircling mechanism or the spiral model to update the location of whales. Specifically, if $q < 0.5$ and $|A|\langle 1$, Eq. (4) will be applied to update the location of the current whale. If $q \geq 0.5$, Eq. (5) is utilized to update the location of whales. Additionally, if $q < 0.5$ and $|A| \geq 1$, Eq. (6) is introduced to update the location of whales and it can be viewed as a random search process.

$$X(t+1) = X^{*}(t) - A|C \times X^{*}(t) - X(t)| \text{ if } q < 0.5 \text{ and } |A| < 1 \tag{4}$$

$$X(t+1) = |C \times X^{*}(t) - X(t)|\, e^{bl}\cos(2\pi l) + X^{*}(t) \text{ if } q \geq 0.5 \tag{5}$$

$$\begin{cases} X(t+1) = X_{\text{rand}}(t) - A \times D \text{ if } q < 0.5 \text{ and } \|A\| \geq 1 \\ D = |C \times X_{\text{rand}}(t) - X(t)| \end{cases} \tag{6}$$

where X denotes a location vector for all whales, t means the iteration metrics, X^* represents the current optimal solution, A and C are the coefficient vector and satisfy $A = 2a \cdot r - a$ and $C = 2 \cdot r$. X_{rand} means the location vector of the randomly selected whales, D represents the distance of the i-th whale to prey. Note that, to balance the global and local search ability of WOA, a nonlinear convergence factor $a^{'}$ is adopted here to replace the original convergence factor a with linear changing to update the location of whales, which can be represented by $a^{'} = 2 - 2 \times \sin(\pi t/2T)$, where t is the current iteration number. Besides, r is the random vector; and q is the random number between 0 and 1; b implies a constant term; and l denotes the random number between -1 and 1, which satisfies the uniform distribution.

(4) Estimate the fitness value of each whale and determine the global optimal location of whales. If $\widehat{X}^{i}$ is superior to X^{i}, $\widehat{X}^{i}$ is selected as the global optimal location of whales. If not, X^{i} is retained to continue to conduct the iterative operations.

(5) Determine whether the stopping criterion is reached. Specifically, examine if the largest SCIM or maximum number of iterations is obtained. If we obtained the largest SCIM or a maximum number of iterations, the optimized results of IWOA will be achieved, that is, the optimal combination parameters of MFMD are obtained. If not, set $t = t + 1$, we will keep on running steps (3)-(4) until the stopping criterion is satisfied.

2.4. Simulation signal analysis

To shed further light on the decomposition performance of the proposed MFMD method, according to the following Eq. (7), the artificial simulated multivariate signal $X(t) = [x_1(n), x_2(t), x_3(t)]$ containing three channels is generated to model multi-sensor data under bearing local faults [54], which includes four parts (i.e., periodic impulses, random impulses, harmonic interference, and noise).

$$\begin{cases} x_1(t) = \underbrace{\sum_i e^{-\beta_1(t-kT_a-\tau_i)}\sin\left(2\pi f_{g1}(t-kT_a-\tau_i)\right)}_{Periodic\ impulses} + \underbrace{\sum_j^{J_1} e^{-\beta_2\left(t-\varepsilon_j\right)}\sin\left(2\pi f_{g2}\left(t-\varepsilon_j\right)\right)}_{Random\ impulses} + \underbrace{\sum_k^{J_2}\sin(2\pi f_k t+\varphi_k)}_{harmonic\ \mathrm{interference}} + \underbrace{n(t)}_{Noise} \\ x_2(t) = \underbrace{1.2\sum_i e^{-\beta_1(t-kT_a-\tau_i)}\sin\left(2\pi f_{g1}(t-kT_a-\tau_i)\right)}_{Periodic\ impulses} + 0.7\underbrace{\sum_j^{J_1} e^{-\beta_2\left(t-\varepsilon_j\right)}\sin\left(2\pi f_{g2}\left(t-\varepsilon_j\right)\right)}_{Random\ impulses} + \underbrace{1.1\sum_k^{J_2}\sin(2\pi f_k t+\varphi_k)}_{harmonic\ \mathrm{interference}} + \underbrace{n(t)}_{Noise} \\ x_3(t) = \underbrace{1.1\sum_i e^{-\beta_1(t-kT_a-\tau_i)}\sin\left(2\pi f_{g1}(t-kT_a-\tau_i)\right)}_{Periodic\ impulses} + 4\underbrace{\sum_j^{J_1} e^{-\beta_2\left(t-\varepsilon_j\right)}\sin\left(2\pi f_{g2}\left(t-\varepsilon_j\right)\right)}_{Random\ impulses} + 2\underbrace{\sum_k^{J_2}\sin(2\pi f_k t+\varphi_k)}_{harmonic\ \mathrm{interference}} + \underbrace{n(t)}_{Noise} \end{cases} \tag{7}$$

where$\beta_1 = 930$ rad/s is the damping coefficients of periodic impulses, β_2=680 rad/s is the damping coefficients of random impulses, $T_a = 1/f_a$ is the time interval of adjacent impulses and$f_a = 100$ Hz is the simulated bearing fault frequencies; τ_i is defined as 1 $\sim$ 2% of the fault period T_a, which is used to simulate the random slip of rolling elements or the time jitter caused by speed fluctuation. In addition, ε_j is the occurrence time of the j-th random impulses, f_{g1}=2600 Hz is the resonant frequency of periodic impulses, f_{g2}=4800 Hz is the resonant frequency of random impulses, $J_1 = 3$ is the number of random impulses, $J_2 = 3$ is the number of harmonic interference, f_k represents the frequency of the k-th harmonic interference, and φ_k represents the phase of the k-th harmonic interference. Specifically, three sinusoidal signals with the frequency (i.e.,$f_1 = 15$ Hz,$f_2 = 70$ Hz, and$f_3 = 180$ Hz) and the phase, (i.e., φ_1=0, φ_2=π/6, and$\varphi_3 = \pi/3$) are superimposed to construct the whole harmonic interference. $n(t)$ indicates the background noise with a standard deviation of 0.933. The sampling frequency and sampling time are set as 25.6 kHz and 1 s, respectively. Fig. 5 shows the simulated multivariate signal and its corresponding envelope spectrum. From the envelope spectrum, there is an obvious harmonic interference-related coupling frequency (i.e., $f_2 - f_1$, $f_3 - f_2$ and $f_3 - f_1$) at different channels, but bearing fault frequencies cannot be observed.

The proposed MFMD and two other methods (i.e., MVMD and MEMD) are compared to decompose the simulated multivariate signal. For fair comparison, the parameters of MFMD and MVMD are all determined by the IWOA method. Specifically, the mode number and filter size of MFMD are optimized as 4 and 30, respectively. The mode number and balance factor of MVMD are optimized as 4 and 500, respectively. Figs. 6 to 8 respectively show the analysis results obtained by three methods (i.e., MFMD, MVMD and MEMD) for simulated multivariate signal, including the decomposed multichannel mode components and their corresponding envelope spectrum. It can be found from Fig. 6(b) that bearing fault frequencies f_a are effectively extracted at the envelope spectrum of 2-order multichannel mode components, which indicate the availability of MFMD in multivariate signal processing. In Fig. 7(b), although the harmonic interference-related coupling frequencies are separated, bearing fault frequencies f_a are almost invisible. Besides, in Fig. 8(b), only the envelope spectrum of the second mode component of the 2nd channels can extract bearing fault frequencies f_a, impulse fault information is basically not found for other channels. Therefore, from this comparison, it can be intuitively reflected that the decomposition performance of MFMD is better than MVMD and MEMD for multivariate signal.

To evaluate the performance of each algorithm quantitatively, here a multichannel comprehensive index (MCI) (see Eq. (8)) is formulated, which is formally defined as the product between the average kurtosis and average feature energy ratio (FER) at each order multichannel mode components.

$$MCI = \left(\frac{1}{N_o}\sum_{o=1}^{N_o} Kur_o^{(k)}\right) \times \left(\frac{1}{N_o}\sum_{o=1}^{N_o} FER_o^{(k)}\right), k = 1, 2, \cdots, K \tag{8}$$

$$Kur_o^{(k)} = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(\widehat{u}_o^{(k)}(n) - \overline{\widehat{u}}_o^{(k)}(n)\right)^4}{\sigma^4}, FER_o^{(k)} = \frac{A(f)+A(2f)+A(3f)}{A_{total}} \tag{9}$$

where N_o denotes the total number of channels of the input multivariate signal. In Eq. (9), $Kur_o^{(k)}$ and $FER_o^{(k)}$ are the kurtosis and

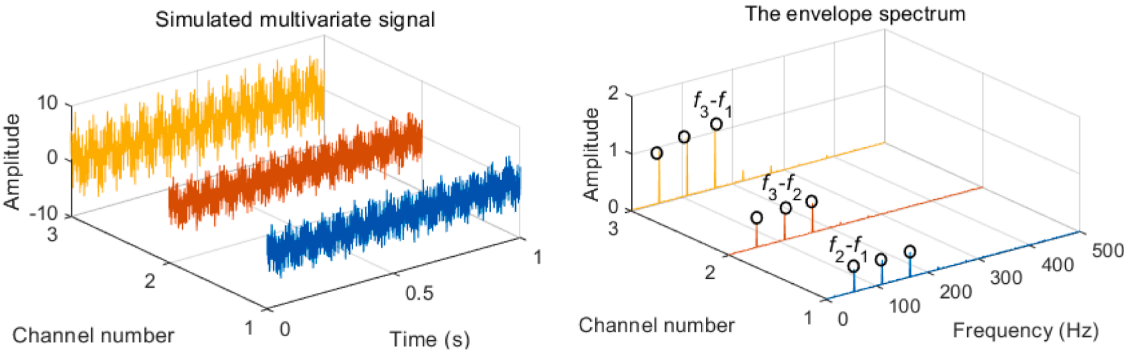


Fig. 5. Simulated multivariate signal and the corresponding envelope spectrum.

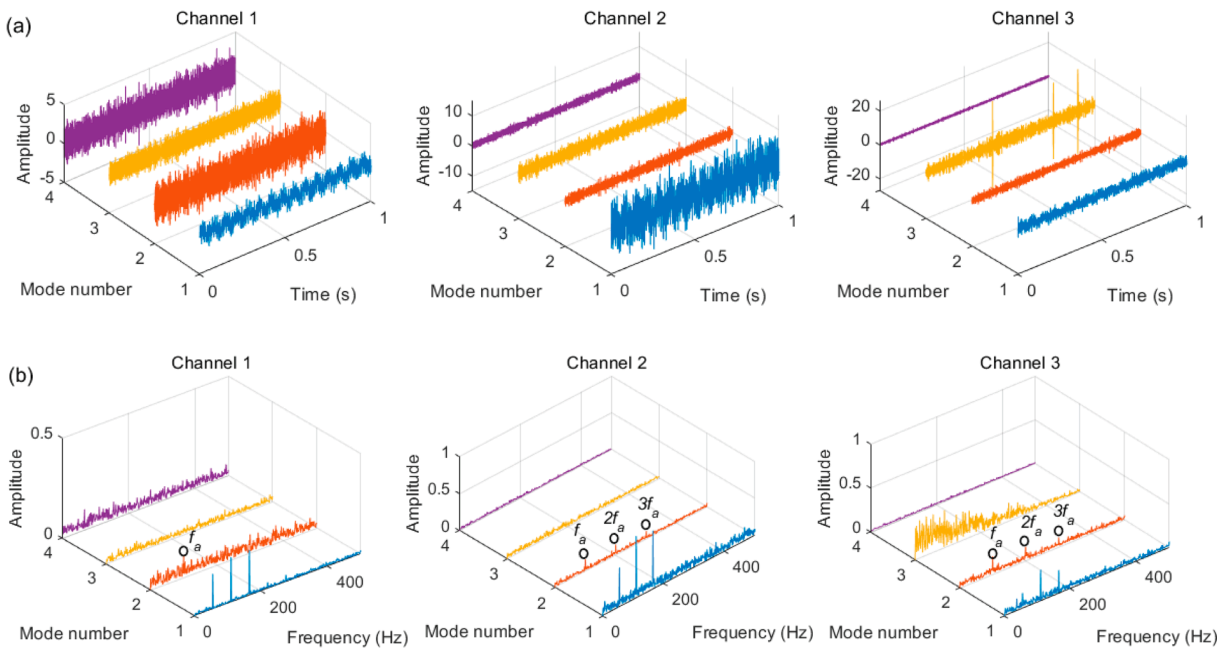


Fig. 6. Results of MFMD in simulated multivariate signal: (a) multichannel mode components and (b) their envelope spectrum.

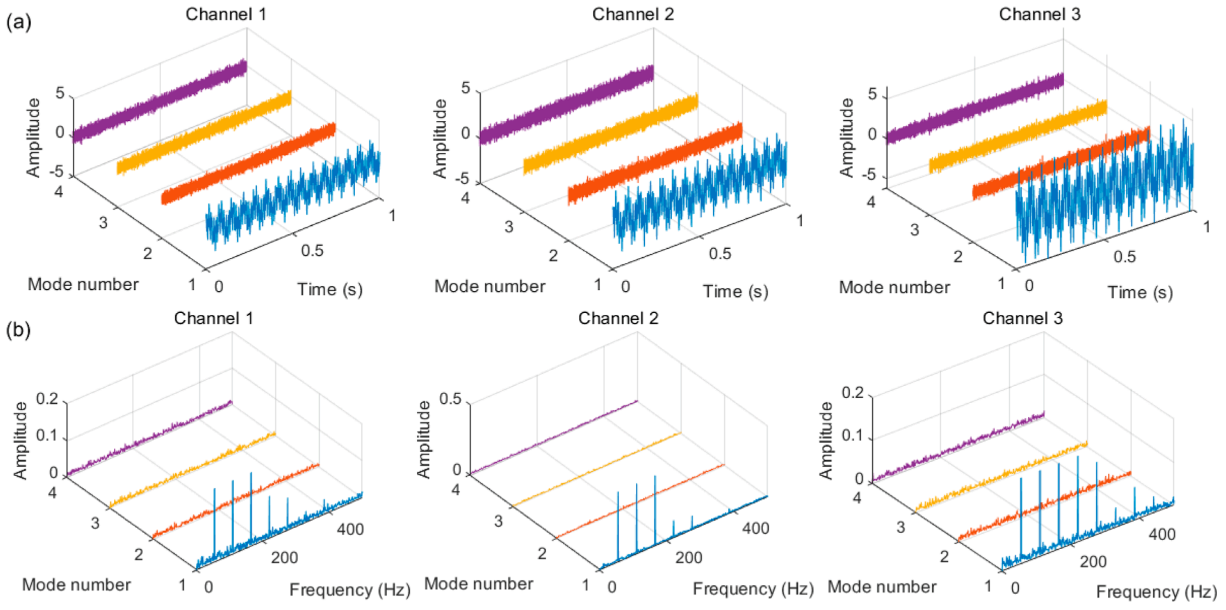


Fig. 7. Results of MVMD in simulated multivariate signal: (a) multichannel mode components and (b) their envelope spectrum.

feature energy ratio of the k-th mode component $\widehat{u}_o^{(k)}(n)$ at the o-th channel, respectively. N represents the number of data points of multivariate signal; σ is the standard deviation of the k-th mode component $\widehat{u}_o^{(k)}(n)$ at the o-th channel; $\overline{\widehat{u}_o^{(k)}}(n)$ represents the mean value of $\widehat{u}_o^{(k)}(n)$; f denotes the fault frequency; $A(f)$,$A(2f)$ and $A(3f)$ are, respectively, the amplitude corresponding to the first three order fault frequencies at Hilbert envelope spectrum; and A_{total} is the total amplitude at Hilbert envelope spectrum. Since kurtosis can depict the impulsiveness of bearing fault signals from time-domain, while feature energy ratio (FER) can characterize the cyclostationarity of bearing fault signals from Hilbert envelope spectrum domain [55], the designed MCI index can simultaneously measure the impulsiveness and cyclostationarity of multichannel mode components obtained by various methods, which is suitable for evaluating the performance of different decomposition algorithms. Concretely, the larger the MCI value represents the better the feature extraction performance. Table 1 lists the MCI value and CPU time of three methods (i.e., MFMD, MVMD and MEMD) for the simulated

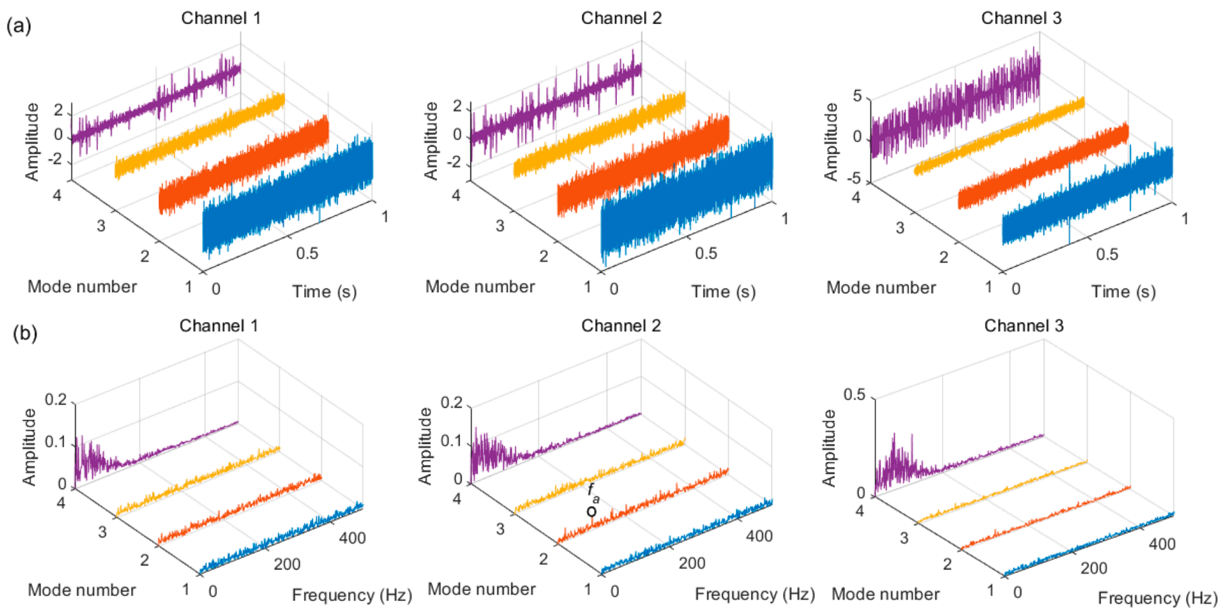


Fig. 8. Results of MEMD in simulated multivariate signal: (a) multichannel mode components and (b) their envelope spectrum.

multivariate signal. Apparently, compared with MVMD and MEMD, the MCI value of MFMD is basically the largest in each order multichannel mode components, except for 4-order multichannel mode components. This again demonstrates the excellent feature extraction performance of MFMD. Besides, in terms of running time, MEMD runs the fastest, but CPU time of MFMD is almost half that of MVMD. This further indicates the potential of MFMD in multi-sensor signal analysis.

3. Multi-attention fusion residual convolutional neural network

3.1. Squeeze-excitation module

Squeeze-excitation module (SEM) presented by Hu et al. [56] is mainly composed of squeeze operation F_{sq} and excitation operation F_{ex}, which can strengthen the connection between each channel and improve the performance of the network model by automatically obtaining the importance of each channel. Fig. 9 shows the structure diagram of SEM. As shown in Fig. 9, in SEM, the input feature maps $X = [x_1, x_2, \cdots, x_c] \in R^{H \times W \times C}, c = 1, 2, \cdots, C$ is firstly performed the convolution operation $F_{c'}$ to obtain the feature maps $U = [u_1, u_2, \cdots, u_{c'}] \in R^{H' \times W' \times C'}, c' = 1, 2, \cdots, C'$, then the squeeze operation F_{sq} and excitation operation F_{ex} of the obtained feature maps U are executed orderly to generate the extracted results $y_{k'}$, and finally the scale operation $F_{scale}(u_{c'}, y_{c'})$ between the extracted results $y_{c'}$ and the feature maps U to obtain the final output result $X' = [x'_1, x'_2, \cdots, x'_{c'}]$ of SEM. More explicitly, in SEM, main idea of the squeeze operation F_{sq} is to conduct global average pooling of the input feature maps U, and compress each feature map into a real number with a global receptive field, which can be represented by Eq. (10).

$$F_{sq}(u_{c'}) = \frac{1}{H' \times W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} u_{c'}(i,j) \tag{10}$$

where $c' = 1, 2 \cdots, C'$, C' is the maximum number of channels of the input feature maps U, and $u_{c'}$ is the input k'-th feature map with the size of $H' \times W'$. Besides, the excitation operation F_{ex} is composed of two fully-connected layers and two sigmoid activation functions, which can contribute to capture channel correlations and generate weights for corresponding channels. Specifically, the excitation operation F_{ex}can be represented by Eq. (11).

$$y_{k'} = F_{ex}\left(F_{sq}(u_{c'}), W\right) = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot F_{sq}(u_{c'})\right)\right) \tag{11}$$

Table 1
MCI value of each order multichannel mode components acquired via various approaches.

Different approaches	1st mode	2nd mode	3rd mode	4th mode	CPU time (s)
MFMD	0.0052	0.0403	0.0321	0.0087	7.5367
MVMD	0.0043	0.0089	0.0052	0.0039	16.1296
MEMD	0.0020	0.0122	0.0109	0.0285	1.4098

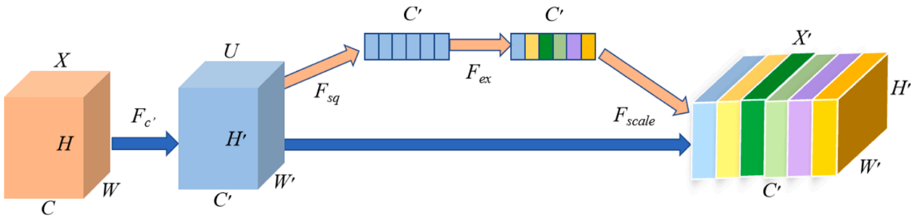


Fig. 9. The structure diagram of SEM.

where W_1 and W_2 respectively are the first and second fully-connected layer operation, $\sigma(\cdot)$ and $\delta(\cdot)$ are the sigmoid activation function and ReLU activation function, respectively.

3.2. Convolutional block attention module

The convolutional block attention module (CBAM) presented by Woo et al. [57] is a kind of lightweight attention module considering spatial and channel information simultaneously, which can enhance the useful feature information in the feature maps and suppress the useless feature information, so as to achieve better results in practical application. Fig. 10 shows the structure diagram of CBAM. Specifically, CBAM includes two calculation stages (i.e., channel attention module and spatial attention module). In the first stage, the input feature F is first fed into the channel attention module (see the green block in Fig. 10) to adjust the input feature maps and obtain channel attention matrix $M_c(F)$ with global information. Then, it is multiplied by the original feature maps F to enlarge the important channel information. That said, the output features of the first stage can be expressed by $F^{'} = M_c(F) \otimes F$. In the second stage, the output features $F^{'}$ of the first stage are regarded as the input of spatial attention module (see purple block in Fig. 10) to further adjust the input feature maps and obtain the spatial attention matrix $M_s(F)$ with global information, and multiply by the output features $F^{'}$ of the first stage to enlarge the important spatial information. In other words, the output features of the second stage are regarded as the final output feature of CBAM, which can be expressed by $F^{''} = M_s(F^{'}) \otimes F^{'}$. More explicitly, the output of the channel attention module and spatial attention module can be described by Eqs. (12) and (13), respectively.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{\max}))) \tag{12}$$

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F);MaxPool(F)])) = \sigma\left(f^{7\times7}\left(\left[F^s_{avg};F^s_{\max}\right]\right)\right) \tag{13}$$

where F denotes the input features; $\sigma$is the sigmoid activation function; $MLP(\cdot)$ is multilayer perceptron operation; $AvgPool(\cdot)$ denotes the average pooling operation; $MaxPool(\cdot)$represents maximum pooling operation; $W_1$and W_0 are the weight learned by $MLP(\cdot)$; $F^c_{avg}$and $F^c_{\max}$ are, respectively, the features obtained by $AvgPool(\cdot)$ and $MaxPool(\cdot)$ in channel attention module; $F^s_{avg}$and $F^s_{\max}$ are, respectively, the features obtained by $AvgPool(\cdot)$ and $MaxPool(\cdot)$ in spatial attention module; and $f^{7\times7}(\cdot)$ means the convolution kernel with the size of 7×7.

3.3. Multi-attention fusion residual convolutional neural network

Residual convolutional neural network (ResCNN) is a class of advanced CNN consisting of convolution, pooling and residual connection and so on, which can solve the problem of gradient vanishing existing in traditional CNN by using jump-over connection with the identity mapping. That is, the over-fitting problem of traditional CNN can be alleviated and the network performance degradation can be prevented when ResCNN is adopted to automatically learn feature information. Nevertheless, traditional ResCNN with direct identity mapping block occasionally leads to gradient exploding problems and even not working properly network training. Furthermore, learning performance of traditional ResCNN with direct identity mapping block is insufficient in excavating synchronously global and local feature information. Therefore, to address these issues, a multi-attention fusion residual convolutional neural

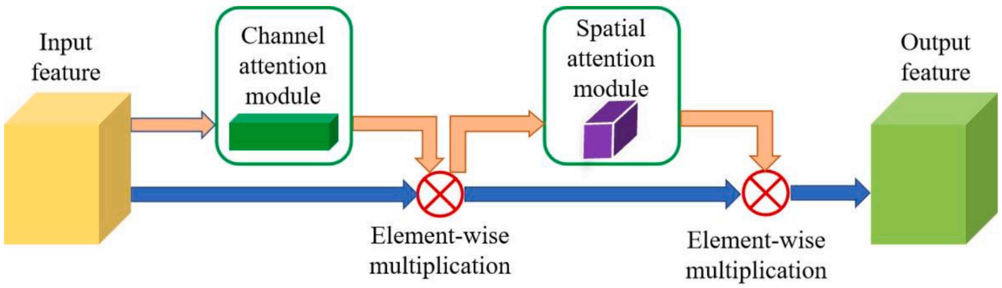


Fig. 10. The structure diagram of CBAM.

network (MAFResCNN) with two attention blocks (i.e., SEM and CBAM) is constructed to capture more discriminative and richer fault feature information, and promote fault feature learning performance of deep network model. Different from traditional ResCNN, the MAFResCNN adopts SEM suggested by Hu et al. [56] instead of direct identity mapping to construct a residual block and prevent gradient exploding along with network instability phenomenon. Also, by doing this, the MAFResCNN can optimize network performance from feature channel without significantly increasing the complexity and computational burden of the model. Besides, to further improve the network learning performance, the MAFResCNN add a CBAM suggested by Woo et al. [57] into the output of ResCNN to simultaneously capture more comprehensively feature information in spatial and channel. Fig. 11 shows the architecture of the MAFResCNN model. Specifically, in the MAFResCNN model, the input feature maps $X = [x_1, x_2, \cdots, x_c] \in R^{H\times W\times C}$ are firstly fed into three residual blocks with SEM to sequentially obtain the corresponding global and local channel features $F_c = [f_c^1, f_c^2, f_c^3]$ in the way of stack learning of layer to layer, where $f_c^i, i = 1, 2, 3$ represents the features learned by the i-th residual blocks with SEM and formally satisfy the following formula $f_c^i = F_{scale}^i(u_c, y_c) \oplus \delta(BN(w_c^i \otimes x_c + b_c^i))$, where $F_{scale}^i(u_c, y_c)$ indicates the output results of the i-th SEM, $\oplus$ denotes the element-wise summation operator, $\delta(\cdot)$ represents ReLU activation function, $BN(\cdot)$ implies batch normalization operator, $\otimes$ means the convolution operator, w_c^i and b_c^i are the weight and bias of the i-th convolution operator at the c-th channel, respectively. Subsequently, according to Eqs. (12) and (13), CBAM is adopted to further enhance the obtained channel features while connecting two fully-connected layers and a dropout layer which is used to prevent the over-fitting issue and improve model generalization ability, and finally implementing fault categorization tasks based on softmax function.

4. Proposed fault diagnosis framework

To obtain more discriminative feature information and improve the identification accuracy, this paper proposes a novel method based on AMFMD and MAFResCNN for machinery multi-sensor fault diagnosis, which is divided into four main stages (i.e., vibration data collection, AMFMD-based data processing, MAFResCNN-based model training, and intelligent fault diagnosis). Fig. 12 shows the whole flowchart of the proposed fault diagnosis framework. The specific steps of our method can be summarized as follows:

(1) *Vibration data collection.* Machinery multi-sensor vibration data are firstly collected via installing multiple accelerometers, then the collected vibration data are randomly divided into the training samples and testing samples.

(2) *AMFMD-based data processing.* In this step, AMFMD is first utilized to decompose all sample data into a group of multichannel mode components, where the IWOA is adopted to automatically select the key parameters (i.e., mode number K and filter size L) of MFMD. Subsequently, the above-reported multichannel comprehensive index (MCI) that has proven to be effective is directly adopted to guide the selection of multichannel mode components containing the most abundant fault information, which can save a certain amount of time cost brought about by the weighted average signal reconstruction. Finally, considering that frequency slice wavelet transform (FSWT) [58] not only has the superior time–frequency analysis ability, but also can preferably adapt to non-stationary vibration signal compared with traditional time–frequency tools (e.g., STFT and WT). Besides, to achieve multi-sensor information fusion without increasing the number of samples, FSWT of the selected multichannel mode components is further calculated and organically fused to generate the colored multichannel time–frequency representation (MTFR) containing multi-sensor important signatures, which can both depict the fault-related information and further reduce the computational burden for subsequent network training. To facilitate the deep understanding of whole data processing, Fig. 13 shows the procedure of obtaining the fused MTFR, which can be understood as a simplest direct concatenation fusion way of data-level fusion. Specifically, take a multivariable signal $X(n) = [x_1(n), x_2(n), x_3(n)], n = 1, 2, \cdots, N$ with three channels as an example, the k-th multichannel mode components obtained by using AMFMD are assumed to be $U^{(k)}(n) = [\widehat{u}_1^{(k)}(n), \widehat{u}_2^{(k)}(n), \widehat{u}_3^{(k)}(n)], k = 1, 2, \cdots, K$, which are selected as multichannel mode components containing the most abundant fault information with the help of MCI. Based on this, according to Eqs. (14) and (15), the FSWT of the selected multichannel mode components $U^{(k)}$ is calculated and skillfully fused to obtain the colored multichannel time–frequency representation (MTFR) which can be figuratively interpreted as three-channel red–greenblue (RGB) images containing multi-sensor important signatures.

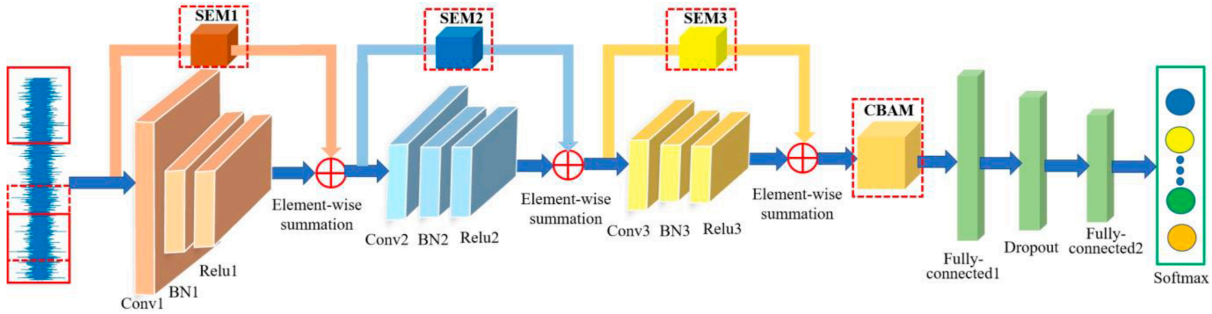


Fig. 11. The architecture of the MAFResCNN model.

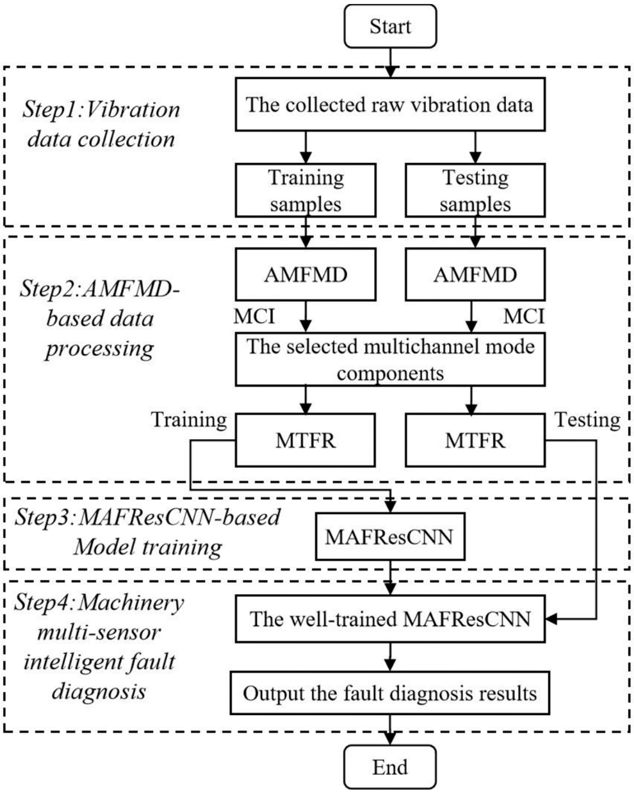


Fig. 12. Flowchart of the proposed fault diagnosis framework.

$$FSWT_o(t,\omega,\sigma)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\widehat{U}^{(k)}(\omega')\widehat{p}^{*}\left(\frac{\omega'-\omega}{\sigma}\right)e^{i\omega t}d\omega',o=1,2,3 \tag{14}$$

$$\begin{cases} GFSWT_o=rgb2gray(FSWT_o) \\ MTFR=cat(3,GFSWT_1,GFSWT_2,GFSWT_3) \end{cases} \tag{15}$$

where $FSWT_o$ represents the FSWT at the o-th channel, σ is a scale factor that is usually constant, $\widehat{U}^{(k)}(\cdot)$ denotes the Fourier transformation operator of the selected multichannel mode components $U^{(k)}$, * represents the conjugate symbol, $\widehat{p}(\cdot)$ means the frequency slice function; t, ω and ω' are the observed time, observed frequency and evaluation frequency, respectively. $rgb2gray(\cdot)$ and $cat(\cdot)$ are the grayscale converting operation and the concatenation operation in MATLAB toolbox, respectively.

(3) *MAFResCNN-based model training.* The GTFR obtained by training samples is fed into MAFResCNN to conduct a model training procedure, where the adaptive moment estimation (Adam) and L2 regularizer are selected to optimize the training process and alleviate the over-fitting problem in model training.
(4) *Machinery multi-sensor intelligent fault diagnosis.* The GTFR obtained by testing samples is fed into the well-trained MAFResCNN to achieve automatic fault identification of machinery faults and output the ultimate diagnosis results.

5. Case studies

In this section, two cases from laboratory and engineering are adopted to verify the effectiveness of the proposed method in machinery multi-sensor fault diagnosis. Additionally, comparison analysis with existing popular methods is launched to validate the advantages of the proposed method. Some discussions and prospects are given in the end.

5.1. Case 1: Bearing multi-sensor data from machinery fault simulator

5.1.1. Description of experimental facilities and bearing data

The artificial bearing fault experiment is conducted on a machinery fault simulator from Southeast University (SEU) to collect bearing multi-sensor vibration data to demonstrate the availability of the presented approach. Fig. 14 displays the experimental facilities and their corresponding structure schematic drawing, which is mainly composed of a loading system, a bearing testing module,

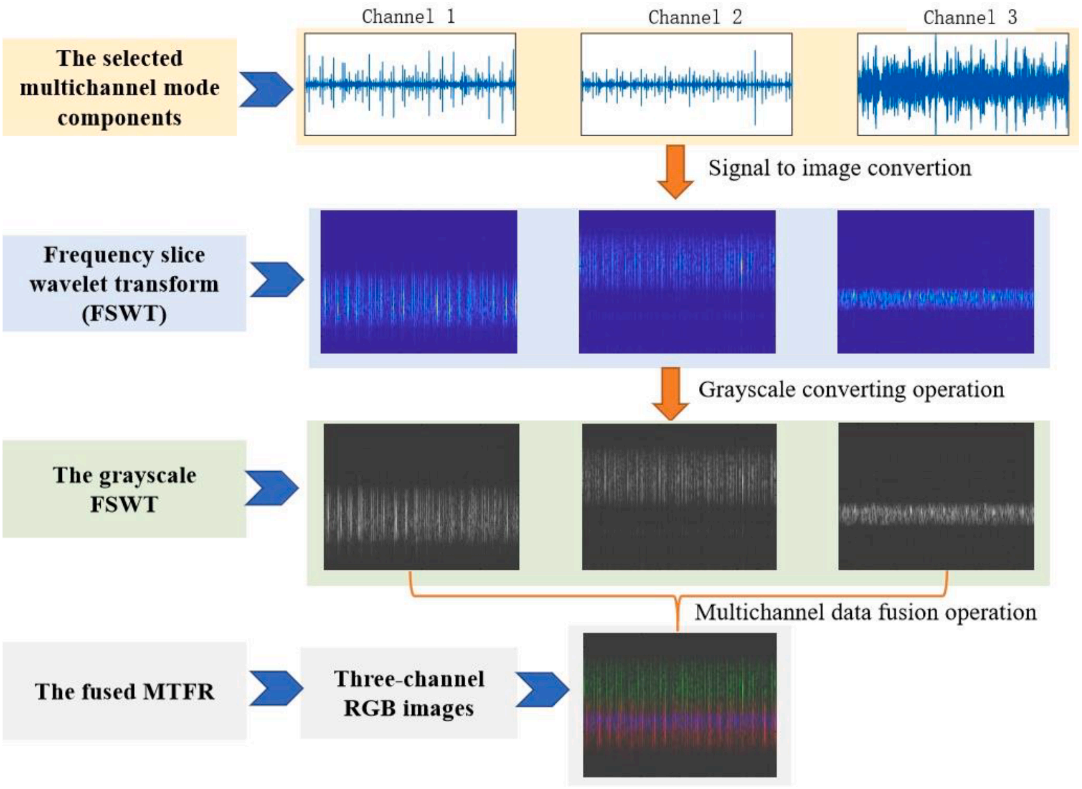
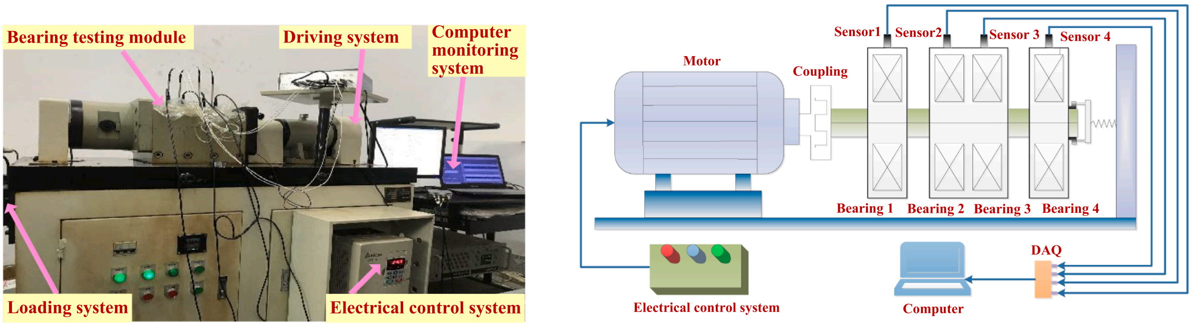


Fig. 13. The procedure of obtaining the fused MTFR.

a driving system, an electrical control system and a computer monitoring system [59,60]. During the experiment, the electric discharge technology was used to implant five fault states (i.e., outer race fault (ORF), inner race fault (IRF), ball fault (BF), outer-inner race compound fault (OIRCF) and outer race-ball compound fault (ORBCF)) on testing bearing (i.e., bearing 1), respectively. Fig. 15 shows the picture of different damaged bearings. Table 2 lists the size of the testing bearing. Four PCB accelerometers with a sensitivity of 100 mV/g are respectively mounted on the vertical direction of four bearings at the bearing testing module to collect bearing multi-sensor data. The motor speed, sampling frequency and sampling time are set as 1050 rpm (i.e., rotating frequency $f_r = 17.5$ Hz), 10240 Hz and 1 s, respectively. Bearing fault frequencies of the outer race, inner race and ball are $f_o = 62.73$ Hz, $f_i = 94.76$ Hz and $f_b = 41.24$ Hz, respectively. In this experiment, although the actual collected bearing data have four channels, here only bearing multi-sensor data with three channels (i.e., sensor 1, sensor 2, and sensor 3) are considered since they are closer to the testing bearing. Fig. 16 shows the time domain waveform of bearing multi-sensor signal under different states. As shown in Fig. 16, it is not easy to intelligently and effectively distinguish the running state of bearings by observing the collected multi-sensor signal directly, especially for sensor 3. In this situation, to facilitate intelligent identification, 100 samples of each bearing state are acquired in this experiment, where 50 samples of each bearing state are selected as training samples and the remainder is regarded as testing samples. Table 3 lists the



(a) (b)

Fig 14. (a) Experimental facilities and (b) its structural schematic drawing [59].

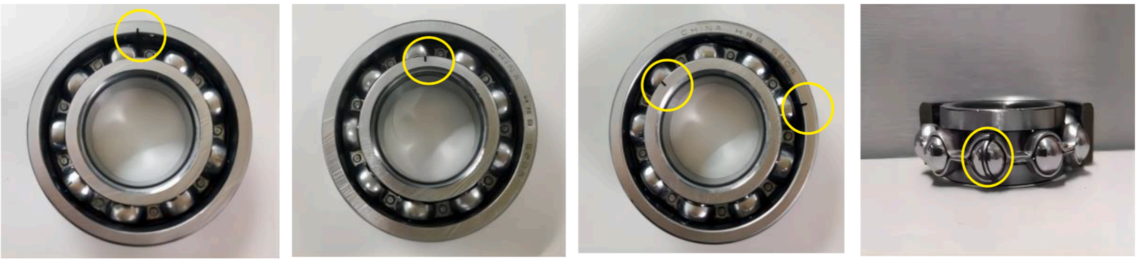

(a) (b) (c) (d)

Fig 15. The picture of damaged bearings [60].

Table 2
The size of the testing bearing.

Bearing type	Roller diameter	Pitch diameter	Number of balls	Contact angle
HRB6205	7.94 mm	39.04 mm	9	0^o

detailed description of bearing data samples.

5.1.2. Results and comparison

The proposed method is adopted to analyze the collected bearing multi-sensor data. According to the diagnosis flowchart, the AMFMD is firstly used to decompose bearing multi-sensor data into a group of multichannel mode components, where the mode number K and filter size L of MFMD are respectively selected as 4 and 20 via the IWOA method. Under this set of parameters, Figs. 17 and 18 show the decomposed multichannel mode components and their corresponding envelope spectrum, respectively. Seen from Figs. 17 and 18, bearing outer race fault frequencies $f_o = 62.73$ Hz and inner race fault frequencies $f_i = 94.76$ Hz can be extracted obviously, and their frequency doubling ingredients can also be found in the envelope spectrum. This preliminarily indicates that the

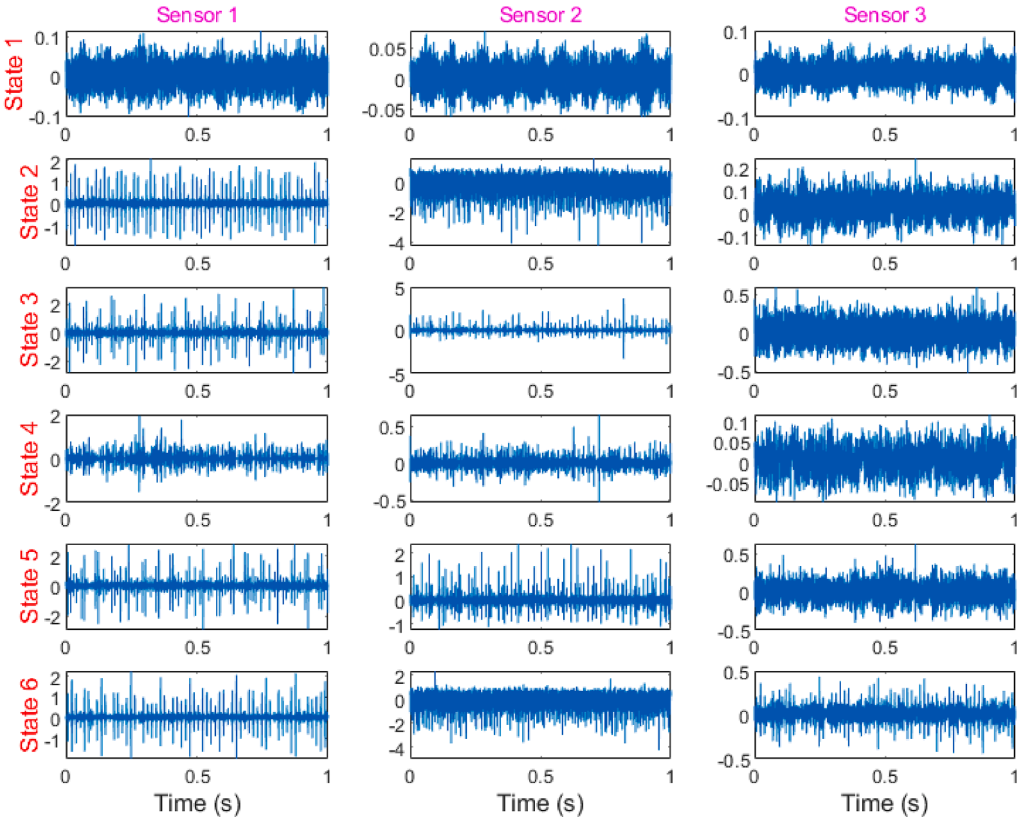


Fig. 16. Time domain waveform of bearing multi-sensor signal under different states.

Table 3
Description of bearing datasets.

State abbreviation	State category	Number of training samples	Number of testing samples	State labels
State 1	Normal	50	50	1
State 2	Outer race fault (ORF)	50	50	2
State 3	Inner race fault (IRF)	50	50	3
State 4	Ball fault (BF)	50	50	4
State 5	Outer-inner race compound fault (OIRCF)	50	50	5
State 6	Outer race-ball compound fault (ORBCF)	50	50	6

presented AMFMD method is effective in detecting fault features of bearing outer race or inner race, but the related domain knowledge and expert experience are required high through the direct human observation of fault frequency of envelope spectrum to diagnose mechanical fault. That said, it may be not suit for some ordinary workers. In view of these factors, multichannel mode components containing the most abundant fault information are automatically selected with the help of MCI. To save space, take bearing multi-sensor IRF signal as an example, Fig. 19(a) and (b) give the MCI and the selected multichannel mode components, respectively. It can be clearly found from Fig. 19 that the third multichannel mode components are selected due to its largest MCI, which are consistent with the facts that the third multichannel mode components have more fault frequency information than other multichannel mode components in the envelope spectrum results of AMFMD. Meanwhile, FSWT of the selected multichannel mode components are further calculated and fused to obtain the colored MTFR containing multi-sensor information, which is plotted in Fig. 20. Seen from Fig. 20, the transient impact information under various fault states is relatively clear, enable us to obtain the complementary multi-channel fault information which is very helpful for fault identification. Finally, the fused MTFR is fed into the designed MAFResCNN model to conduct training procedures and identify automatically bearing fault states. Table 4 lists the details of key parameters of the constructed MAFResCNN model. Additionally, to highlight the advantages of the MAFResCNN model, several representative deep network models (i.e., MK-ResCNN, ResCNN, MSCNN, CNN and long short-term memory (LSTM) [61]) are also adopted to deal with the same MTFR. Fig. 21(a) to (f) show the confusion matrix of our proposed approach and other representative deep network models, respectively. It is obviously in Fig. 21 that the proposed approach can obtain a recognition accuracy of 100% (300/300), whereas the recognition accuracy of the other five approaches (i.e., MK-ResCNN, ResCNN, MSCNN, CNN and LSTM) are 98.33% (295/300), 98% (294/300), 97.33% (292/300), 96.33% (289/300) and 95.33% (286/300), respectively. That said, the proposed approach has a higher recognition accuracy compared with the other five representative approaches, which primitively verified the advantages of MAFResCNN in bearing fault state identification.

To illustrate the effectiveness of MFMD used in the proposed approach, MVMD and MEMD are also integrated with MAFResCNN to analyze the same bearing multi-sensor data. In addition, 10 trials of each approach are executed to obtain an objective and impartial comparison results. Fig. 22 plots the recognition accuracy of different combination approaches in 10 trials, and the detailed statistical results are listed in Table 5. As seen from Fig. 22 and Table 5, the average recognition accuracy (99.97%) of the proposed approach is the largest. Besides, the standard deviation (0.1044) of the proposed approach is smaller than that of the other two combination approaches (i.e., MVMD and MAFResCNN, MEMD and MAFResCNN), which indicates that the proposed approach not only has better

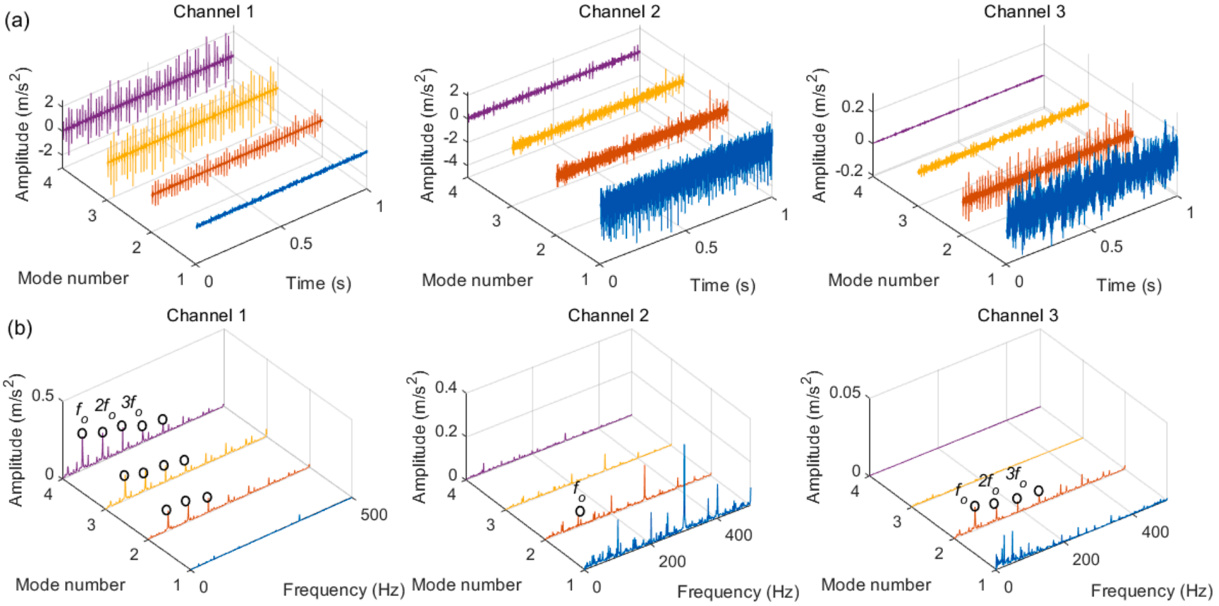


Fig. 17. Results of AMFMD in bearing multi-sensor ORF signal: (a) multichannel mode components and (b) their envelope spectrum.

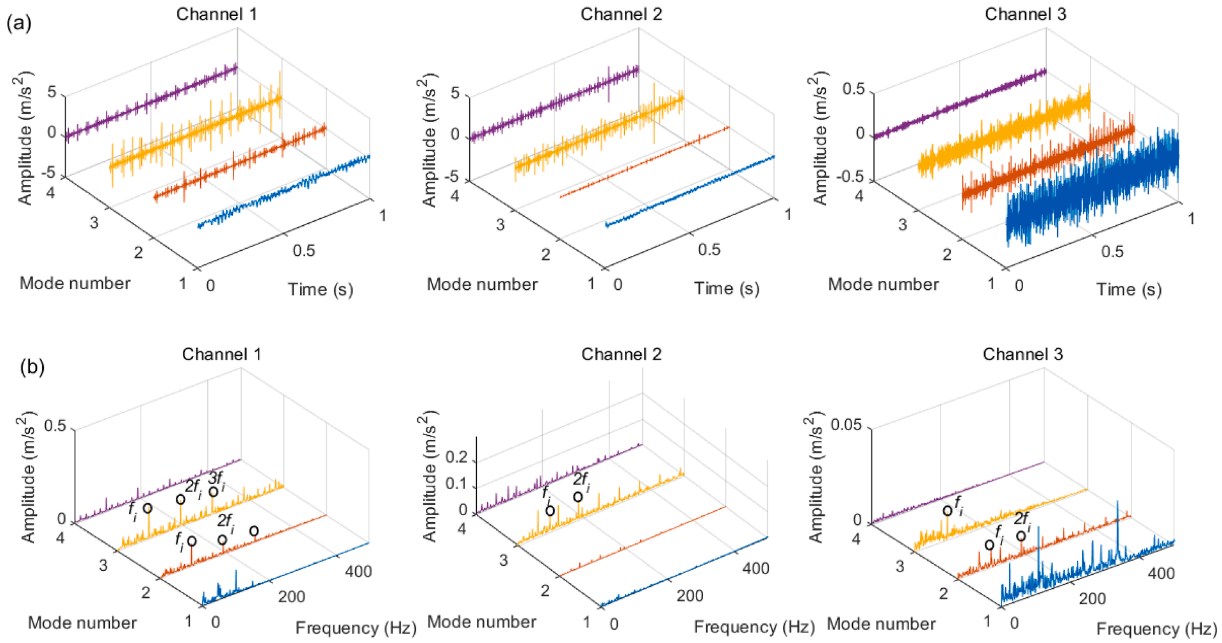


Fig. 18. Results of AMFMD in bearing multi-sensor IRF signal: (a) multichannel mode components and (b) their envelope spectrum.

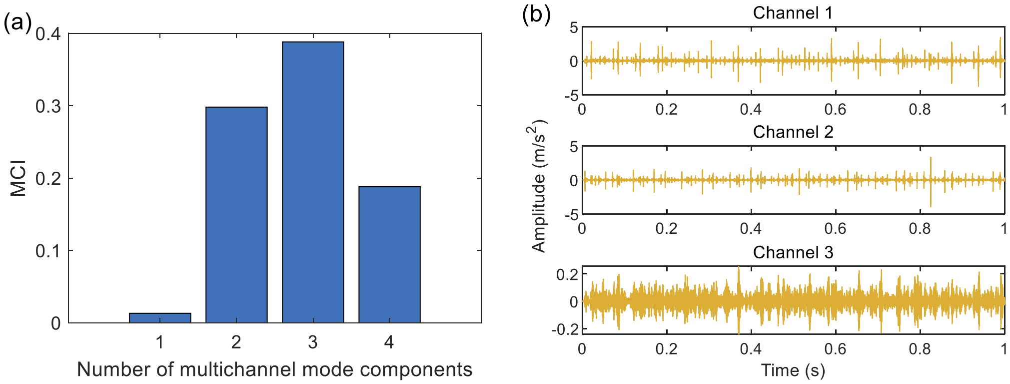


Fig. 19. Mode selection results of bearing multi-sensor IRF signal: (a) MCI and (b) the selected multichannel mode components.

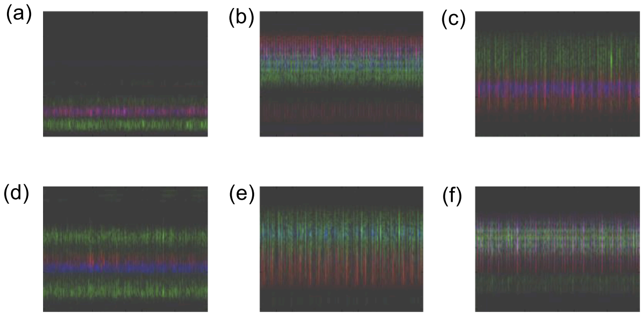
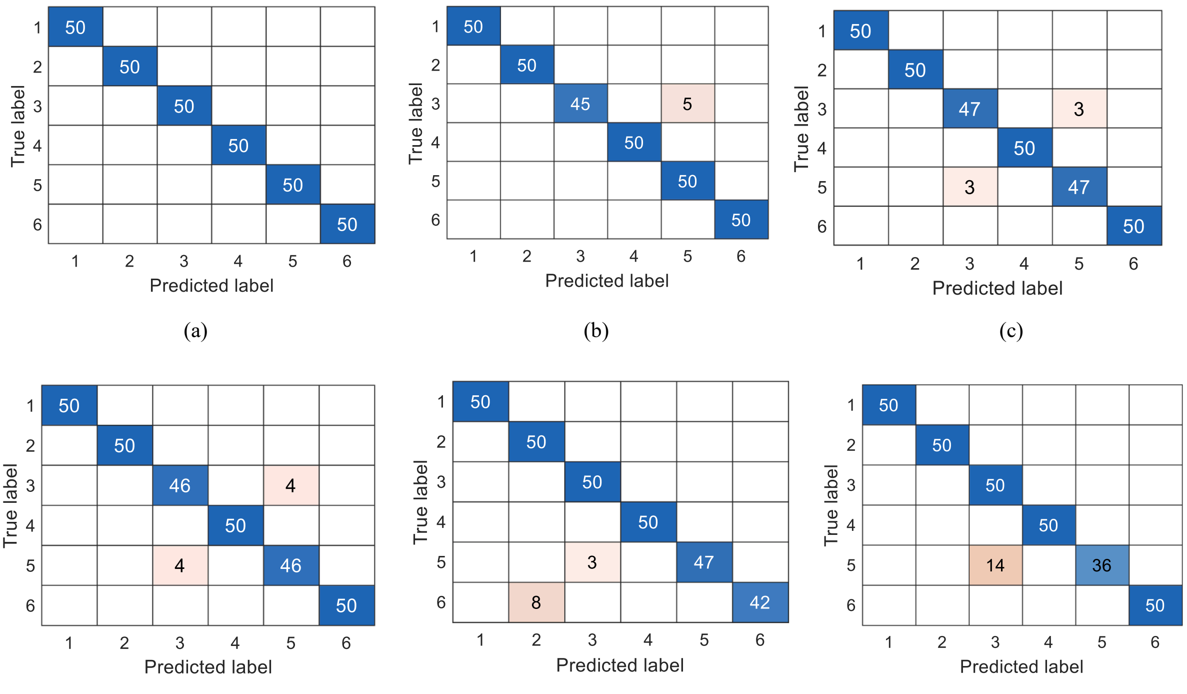


Fig. 20. MTFR of different bearing operation states: (a) Normal, (b) ORF, (c) IRF, (d) BF, (e) OIRCF and (f) ORBCF.

Table 4
The details of parameters of the constructed MAFResCNN model.

No.	Layers	No. of Filters	Kernel size	Stride	Output Size	Padding
1	Input layer	—	—	—	64 × 64 × 3	—
2	Conv1	32	3 × 3	2	32 × 32 × 32	Yes
3	BN1	—	—	—	32 × 32 × 32	—
4	Relu1	—	—	—	32 × 32 × 32	—
5	Conv2	32	2 × 2	1	32 × 32 × 32	Yes
6	BN2	—	—	—	32 × 32 × 32	—
7	Relu2	—	—	—	32 × 32 × 32	—
8	Conv3	32	2 × 2	2	16 × 16 × 32	Yes
9	BN3	—	—	—	16 × 16 × 32	—
10	Relu3	—	—	—	16 × 16 × 32	—
11	Fully-connected 1	—	—	—	1 × 1 × 128	—
12	Dropout (rate = 0.5)	—	—	—	1 × 1 × 128	—
13	Fully-connected 2	—	—	—	1 × 1 × 6	—
14	Softmax	—	—	—	—	—



(a) (b) (c) (d) (e) (f)

Fig. 21. Identification results of different approaches in bearing multi-sensor signal: (a) Proposed approach, (b) MK-ResCNN, (c) ResCNN, (d) MSCNN, (e) CNN and (f) LSTM.

recognition performance, but also has better stability. That said, MFMD used in the proposed approach is more effective and reasonable for bearing fault identification compared with MVMD and MEMD.

To further demonstrate the effectiveness and superiority of the proposed method, seven existing machinery single-sensor fault diagnosis methods (i.e., PSO-VMD + SVM [25], CWT + CNN [32], HHT + CNN [37], SST + CNN [39], CSCoh + CNN [40], FSWT + CNN, CWT + ResCNN [46]) and one machinery multi-sensor fault diagnosis method (i.e., MSCNN + BiLSTM [62]) are also employed to handle the above bearing vibration data. Here some key details of these contrasting methods are described. Specifically, in method 1 (PSO-VMD + SVM), the PSO-VMD is firstly adopted to decompose the collected bearing vibration data into several mode components and multi-scale entropy (MSE) of the obtained mode components is calculated to establish the eigenvector as the input of SVM to identify bearing fault states. Additionally, according to Ref. [25], the mode number of VMD is optimized as 4, the embedded dimension and scale factor of MSE are set as 6 and 12, respectively. In method 2 (CWT + CNN), CWT is firstly used to convert the original bearing vibration data into time–frequency image and then CNN is adopted to learn features from the time–frequency image and implement fault classification, where CWT adopts complex Morlet (cmor3-3) as wavelet basis function, and the network architecture and model parameters of CNN are set the same as Ref. [32]. In method 3 ~ 5 (i.e., HHT + CNN, SST + CNN and CSCoh + CNN), SST, HHT and

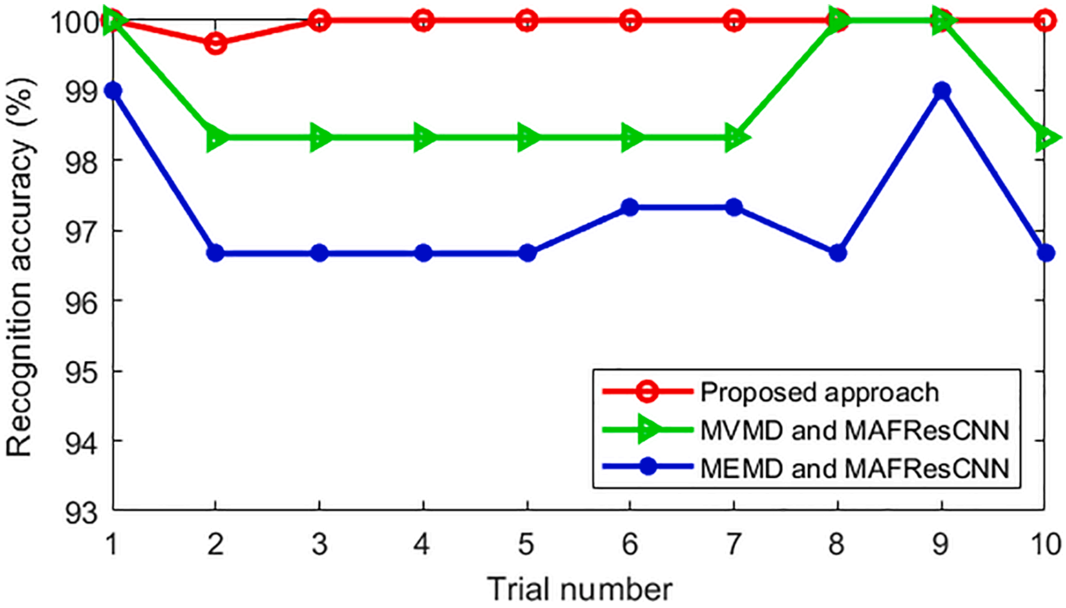


Fig. 22. Recognition accuracy of 10 trials of different combination approaches in case 1.

Table 5

Recognition results of 10 trials of different combination approach in case 1.

Different approaches	Maximum	Minimum	Mean	Standard deviation
The proposed approach	100%	99.67%	99.97%	0.1044
MVMD and MAFResCNN	100%	98.33%	98.83%	0.8067
MEMD and MAFResCNN	99%	96.67%	97.27%	0.9518

CSCoh are firstly applied to convert the original bearing vibration data into time–frequency representation and then CNN is utilized for conducting automatic feature learning and fault classification tasks. Note that, the network architecture and model parameters of CNN in method 3 ~ 5 are set the same as Ref. [37,39] and [40], respectively. In method 6 (FSWT + CNN), FSWT is combined with CNN to learn sensitive feature information and identify bearing fault patterns, where the network architecture and model parameters of CNN is the same as that of method 3 (SST + CNN). In method 7 (CWT + ResCNN), CWT containing complex Morlet is combined with ResCNN for intelligent fault identification, where the network architecture and model parameters of ResCNN are set the same as Ref. [46]. In method 8 (MSCNN + BiLSTM), multi-scale convolutional neural network (MSCNN) is combined with bidirectional long short term memory (BiLSTM) for multi-sensor fault diagnosis. It should be noted that the model parameters of MSCNN and BiLSTM can be found in Ref. [62]. In addition, it is important to emphasize that method 1 ~ 7 belongs to machinery single-sensor fault diagnosis method, whereas method 8 is essentially a machinery multi-sensor fault diagnosis method. Due to sensor 1 being closer to testing bearing

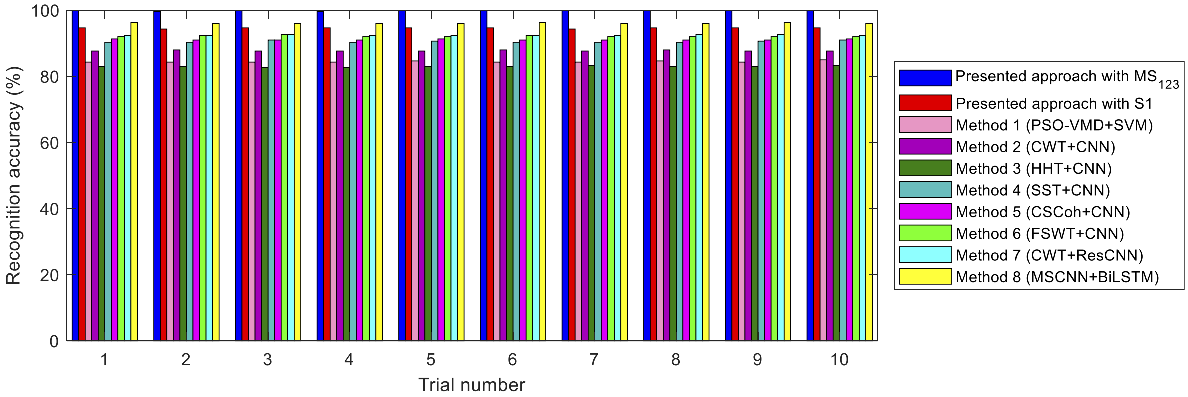


Fig. 23. Recognition accuracy of 10 trials of different comparison approaches in case 1.

compared with other sensors, so methods 1 ∼ 7 are all adopted for analyzing single channel data obtained by sensor 1, while method 8 is used to multi-sensor data with three channels for a fair comparison. Fig. 23 shows the recognition accuracy of different approaches in 10 trials, and the specific recognition results are summarized in Table 6. As seen from Fig. 23 and Table 6, the presented method with multi-sensor information fusion can achieve an average recognition accuracy of 99.97%, which is higher than that of other comparison methods (method 1 ∼ 8). Additionally, from the perspective of single sensor data, when the presented approach is adopted to analyze the data of sensor 1, the obtained average recognition accuracy is 94.40% (2832/3000), which is also larger than that of methods 1 ∼ 7, which are 84.47% (2534/3000), 87.76% (2633/3000), 83.00% (2490/3000), 90.53% (2716/3000), 91.10% (2733/3000), 92.13% (2764/3000) and 92.43% (2773/3000), respectively. This further verifies the effectiveness and superiority of the presented approach containing multi-sensor information fusion in identifying bearing fault states. Additionally, we can see from Fig. 23 and Table 6 that the proposed method has the smallest standard deviation (i.e., best stability), and the CPU time during model training is also very competitive compared with other approaches. The main reason is that the proposed method can organically fuse important information of multi-sensor and effectively finish fault diagnosis tasks without increasing the number of samples. That said, the input data volume of the proposed method is the same as methods 1 ∼ 7, but less than that of method 8.

To show the effectiveness and superiority of MCI-based mode selection used in the proposed method, we conducted a comparison between MCI-based mode selection and squared envelope spectrum Gini index (SESGI)-based signal reconstruction [63]. Specifically, the original collected multi-sensor data are artificially mixed with white Gaussian noise (WGN) containing a signal-to-noise ratio (SNR) of different levels (i.e., 5 dB, 0 dB and −5 dB) to generate the noisy multi-sensor data, and then the MCI-based mode selection and SESGI-based signal reconstruction are applied to analyze the same noisy multi-sensor data. It should be specially explained that, compared with MCI-based mode selection, the only difference of SESGI-based signal reconstruction is that multichannel mode components obtained by AMFMD are regrouped via the weighted average algorithm with SESGI, and FSWT of the regrouped multichannel signal is further calculated to generate MTFR which is regarded as the input of MAFResCNN for completing fault identification. Table 7 lists the specific comparison results. As listed in Table 7, although the identification accuracy of MCI-based mode selection and SESGI-based signal reconstruction are very close in the original signal, the identification accuracy of MCI-based mode selection for the noisy signal is obviously larger than that of SESGI-based signal reconstruction. Additionally, computational cost of MCI-based mode selection is less than that of SESGI-based signal reconstruction. Therefore, the above results verify that MCI-based mode selection used in the proposed method is feasible and more favorable.

To verify the effectiveness and necessity of the fused MTFR used in the proposed method, we conducted a comparison among the proposed method with and without information fusion. Specifically, the proposed method containing the fused MTFR is compared with the proposed method containing the independent FSWT. Note that, in the proposed method containing the independent FSWT, the most sensitive multichannel mode components obtained by AMFMD are firstly selected by MCI, and then FSWT of the selected multichannel mode components are directly input into the same MAFResCNN for automatic feature learning and fault recognition. Table 8 shows the comparative result. From the diagnosis results of Table 8, we can see that the presented approach containing the fused MTFR can achieve higher identification accuracy than the presented approach containing the independent FSWT. Additionally, the presented approach containing the fused MTFR has less model training time compared with the presented approach containing the independent FSWT. This comparative results demonstrate the effectiveness and superiority of data fusion used in the proposed method. Here some reasons are as follows: (1) The addition operations are omitted when the colored FSWT of the proposed method are converted to the grayscale FSWT, so the computational complexity can be reduced to some extent. (2) The collaborative function of important information of multi-sensor are considered via obtaining the fused MTFR. (3) Compared with the independent FSWT, the presented approach containing the fused MTFR can achieve fault diagnosis tasks without increasing the number of samples, which means that the computational burden of MAFResCNN model training can also be alleviated via obtaining the fused MTFR.

To further show the effectiveness and necessity of MAFResCNN model used in the proposed method, take the noisy bearing multi-sensor IRF signal with −5 dB as an example, Fig. 24(a) and (b) show multichannel mode components obtained by AMFMD and their corresponding envelope spectrum, respectively. As seen from Fig. 24, bearing inner race fault frequency is basically invisible in the envelope spectrum, so MAFResCNN is used to further analyze multichannel mode components obtained by AMFMD. To better visualize the results, according to [40], the t-distributed stochastic neighbor embedding (t-SNE) technique is introduced for feature visualization. Specifically, for two scenarios with and without adding noises, three groups of sample data (i.e., the original input data, AMFMD-based envelope spectrum data and the output data of fully-connected layer 1 of MAFResCNN) are respectively mapped into two-dimension space from high-dimensional space by using t-SNE technique. The results are shown in Fig. 25. As seen from Fig. 25, for the original input data, bearing fault states can be identified by the AMFMD-based envelope spectrum, but the discrimination of features extracted by the AMFMD-based envelope spectrum is not as obvious as the features learned by MAFResCNN. More importantly, when the original input data are corrupted by strong noise, it is difficult to identify bearing fault types via the AMFMD-based envelope spectrum. However, compared with the AMFMD-based envelope spectrum, bearing fault patterns can be more clearly when being identified by observing the features learned by MAFResCNN even in strong noise environments. In other words, after AMFMD-based data processing, it is still necessary and beneficial to use MAFResCNN for intelligent fault identification.

5.2. Case 2: Wind turbine multi-sensor data from real engineering

5.2.1. Description of wind turbine and data collection

Machinery vibration measurement on wind turbine seated at a wind farm in North China is conducted to collect wind turbine multi-sensor data [64,65]. Fig. 26 shows the structural diagram of wind turbine, which is mainly composed of an impeller, gearbox and generator, where the gearbox includes a sun gear, intermediate stage gear and high-speed stage gear. During the operation of wind

Table 6
Recognition results of 10 trials of different approaches in case 1.

Methods with different sensor data	Average accuracy (%)	Standard deviation (SD)	CPU time (s)
Presented approach with MS_{123}	99.97% (2998/3000)	0.1044	39.62
Presented approach with S_1	94.60% (2838/3000)	01,434	38.45
Method 1 (PSO-VMD + SVM) with S_1	84.47% (2534/3000)	0.2352	89.41
Method 2 (CWT + CNN) with S_1	87.76% (2633/3000)	0.1594	46.86
Method 3 (HHT + CNN) with S_1	83.00% (2490/3000)	0.2200	39.57
Method 4 (SST + CNN) with S_1	90.53% (2716/3000)	0.2830	50.13
Method 5 (CSCoh + CNN) with S_1	91.10% (2733/3000)	0.1594	52.79
Method 6 (FSWT + CNN) with S_1	92.13% (2764/3000)	0.2333	51.37
Method 7 (CWT + ResCNN) with S_1	92.43% (2773/3000)	0.1642	57.52
Method 8 (MSCNN + BiLSTM) with MS_{123}	96.10% (2883/3000)	0.1594	114.65

Note: MS_{123} is multi-sensor data with three channels and S_1 are single-sensor data with one channel.

Table 7
Diagnosis results of different data processing way for bearing vibration signal with different noise levels in case 1.

Data processing way	Noisy signal with 5 dB		Noisy signal with 0 dB		Noisy signal with −5 dB	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
MCI-based mode selection	99.36 ± 0.1075	2.807	98.40 ± 0.1434	2.823	96.74 ± 0.1391	2.899
SESGI-based signal reconstruction	98.33 ± 0.1579	3.034	96.10 ± 0.1594	3.139	93.40 ± 0.1434	3.153

Table 8
Diagnosis results of the presented approach with and without information fusion in case 1.

Methods	Accuracy (%)	CPU time (s)
The presented approach with the fused MTFR	99.97 ± 0.1044	39.62
The presented approach with the independent FSWT	98.70 ± 0.1895	57.92

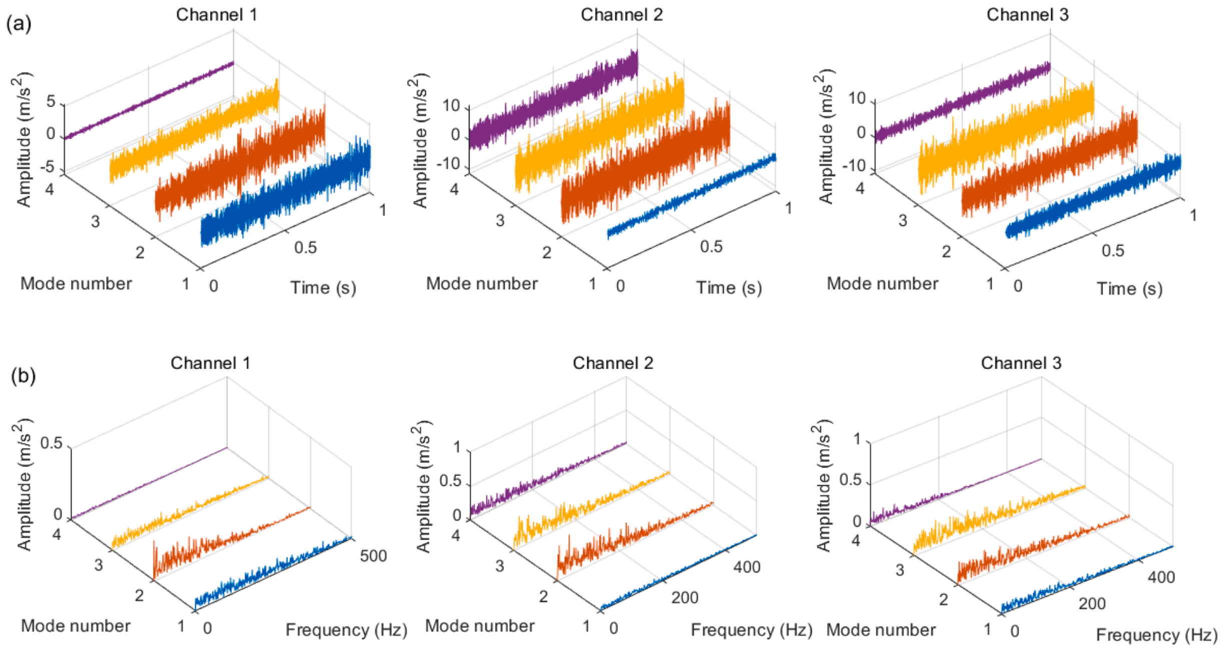
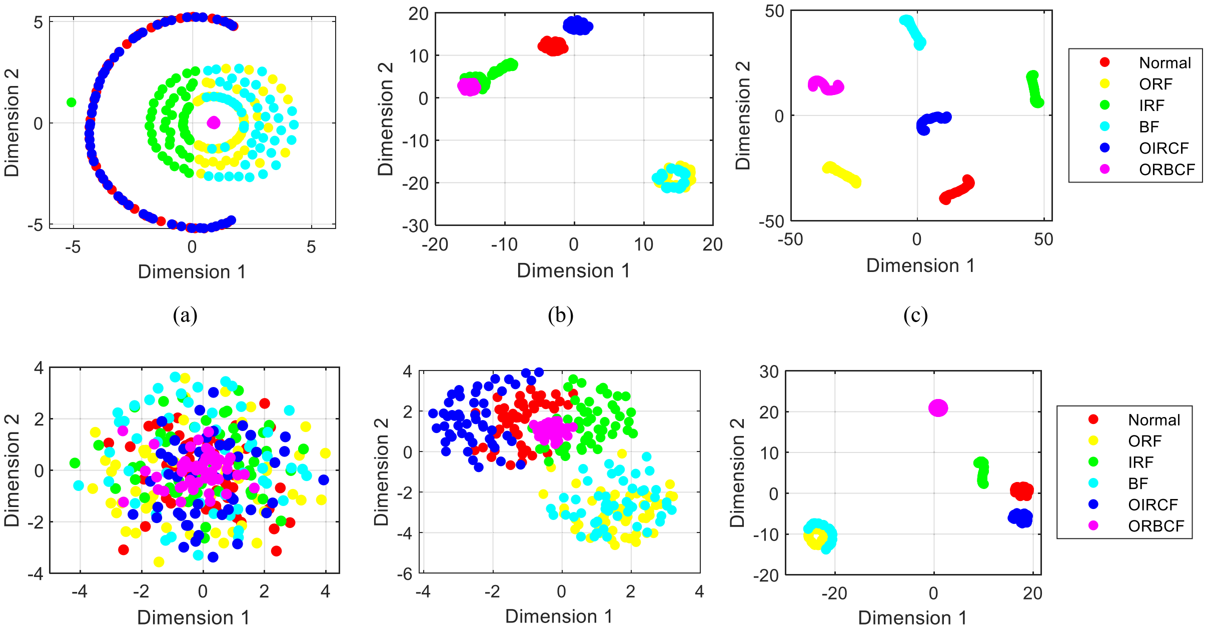


Fig. 24. Results of AMFMD in the noisy bearing multi-sensor IRF signal with −5 dB noise adding: (a) multichannel mode components and (b) their envelope spectrum.

turbine, six groups of acceleration sensors (see Fig. 27) with three directions (i.e., vertical, axial, horizontal) are respectively installed at the main shaft, low-speed stage gear, intermediate stage gear, high-speed stage gear and generator front-end and back-end bearing to collect wind turbine multi-sensor vibration data under four operating states (i.e. normal, high-speed gear fault, generator front-end



(a) (b) (c)

(d) (e) (f)

Fig. 25. Feature visualization by t-SNE under two scenarios (i.e., The original data and the noisy data with −5 dB noise adding) in case 1: (a) The original input data; (b) AMFMD-based envelope spectrum data of (a); (c) the output of MAFResCNN of (a); (d) the noisy input data; (e) AMFMD-based envelope spectrum data of (d); (f) the output of MAFResCNN of (d).

bearing fault and generator rotor imbalance fault). Fig. 28 shows the pictures of the acceleration sensors installation during vibration testing. During data acquisition, the rotating speed of the input shaft is about 21.8 rpm, the rotating speed of the generator shaft is about 1523 rpm (i.e., rotating frequency f_{r1} = 25.4 Hz), the sampling frequency and sampling time are set to be 16384 Hz and 1 s, respectively. A total of 120 data samples were collected for each type of wind turbine operating states in vibration testing. The type of generator bearing is SKF6324C3, and the generator bearing outer race fault frequency f_g is 79.21 Hz. Fig. 29 depicts the time domain waveform of wind turbine multi-sensor signal under different states. As seen from Fig. 29, due to the presence of a large amount of noise in multi-sensor signal, it is difficult to effectively identify the working state of wind turbine by observing the time domain waveform directly. Similarly, to facilitate intelligent fault identification, 200 samples of each bearing state are obtained in wind turbine vibration measurement, where 100 samples of each state are selected as training samples and the remainder is regarded as testing samples. Table 9 lists the detailed description of wind turbine data samples.

5.2.2. Results and comparison

To illustrate the diagnostic performance of the proposed approach, the real wind turbine multi-sensor data are analyzed via the proposed approach. Similarly, AMFMD is firstly adopted to decompose wind turbine multi-sensor data into a set of multichannel mode components. Note that the IWOA method is employed to automatically determine the combination parameters (i.e., mode number K and filter size L) of MFMD as (4, 30). Figs. 30 to 32 plot the decomposed multichannel mode components of MFMD under three states (i.e., state 2, state 3 and state 4) and their corresponding envelope spectrum, respectively. It is obvious that Fig. 30(b) that bearing outer

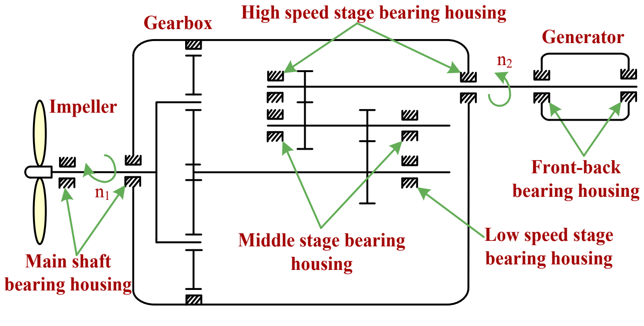


Fig 26. The structure diagram of the wind turbine.

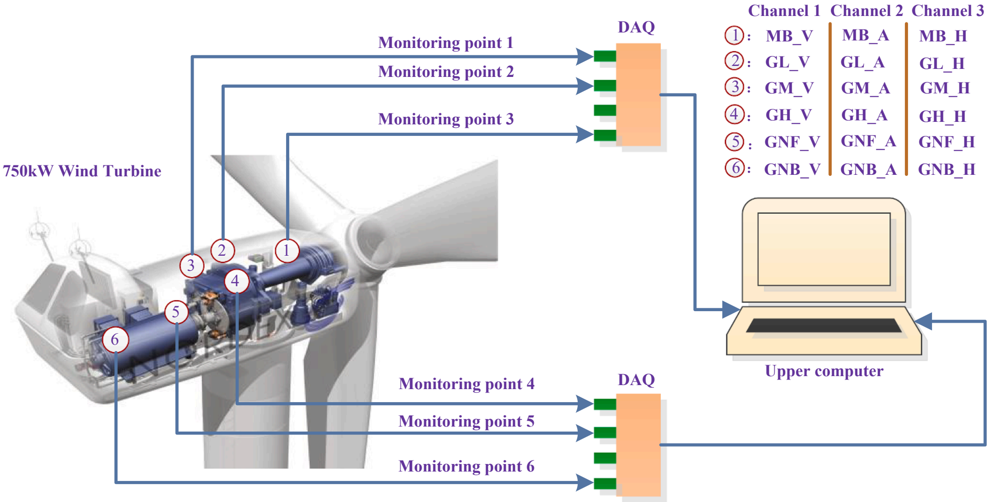


Fig 27. Schematic diagram of wind turbine vibration measuring points [64].

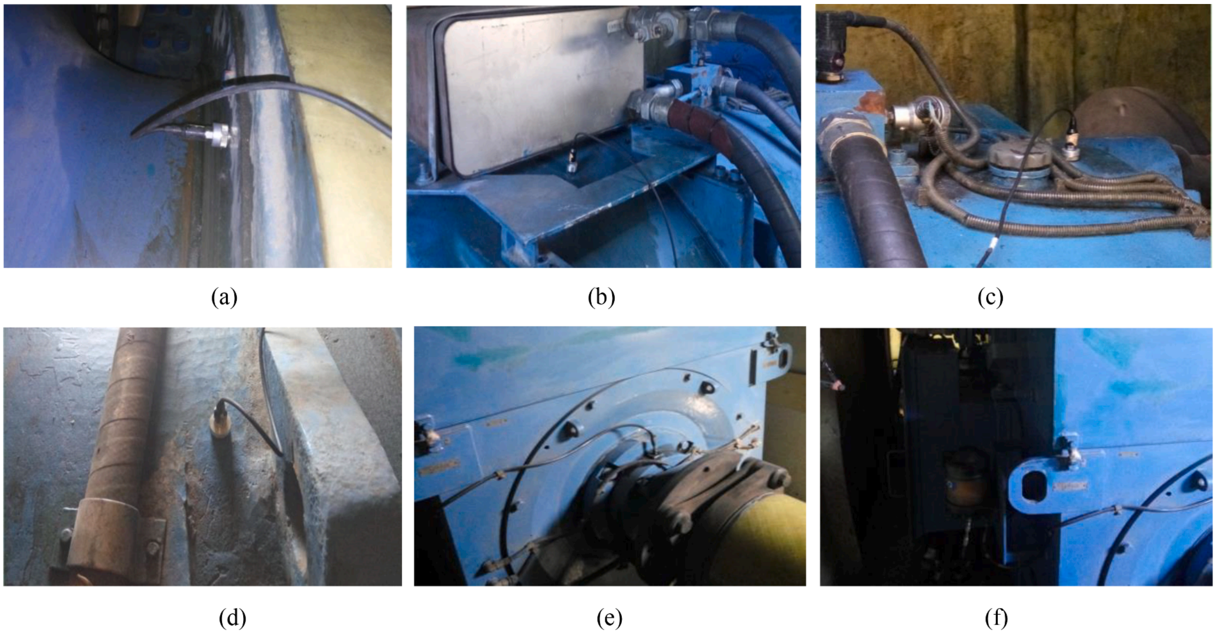


Fig 28. The pictures of acceleration sensors installation during vibration testing: (a) main shaft, (b) low-speed stage gear, (c) intermediate stage gear, (d) high-speed stage gear, (e) generator front-end bearing and (f) generator back-end bearing [65].

race fault frequencies $f_g = 79.21$ Hz and its frequency doubling are effectively extracted in the multichannel envelope spectrum. Additionally, as shown in Fig. 31(b) and Fig. 32(b), rotating frequency information (e.g., $f_{r1} = 25.4$ Hz and its harmonics) is relatively visible. The primary study reveals that AMFMD can reveal wind turbine fault features under three states (i.e., state 2, state 3 and state 4). Afterwards, to avoid the high requirement of expert experience in the direct observation of spectral frequency and achieve intelligent fault identification. MCI is adopted to select multichannel mode components. Due to limited length, take state 2 as an example, Fig. 33(a) and (b) show the MCI and the selected multichannel mode components, respectively. Additionally, FSWT of the selected multichannel mode components is further calculated and fused to obtain MTFR. Fig. 34 plots of the fused MTFR under different states. Conclusively, the fused MTFR is input into the MAFResCNN to execute network model training and identify automatically wind turbine fault states, where the model parameters of MAFResCNN are the same as in Case 1, except for the fully connected 2 is $1 \times 1 \times 4$.

As a contrast, five representative deep network models (i.e., MK-ResCNN, ResCNN, MSCNN, CNN and LSTM) are also adopted to analyze the same MTFR and identify wind turbine fault states. Fig. 35(a) to (f) show the confusion matrix of the proposed MAFResCNN

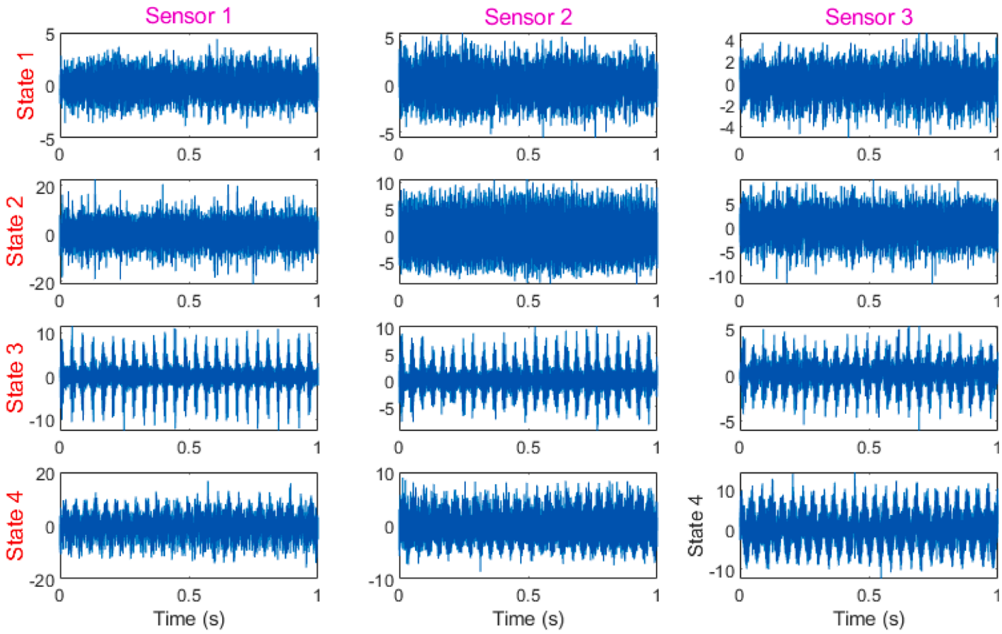


Fig. 29. Time domain waveform of wind turbine multi-sensor signal under different states.

Table 9

Description of wind turbine datasets.

State abbreviation	State category	Number of training samples	Number of testing samples	State labels
State 1	Normal	100	100	1
State 2	Generator front-end bearing fault	100	100	2
State 3	High-speed gear fault	100	100	3
State 4	Generator rotor imbalance fault	100	100	4

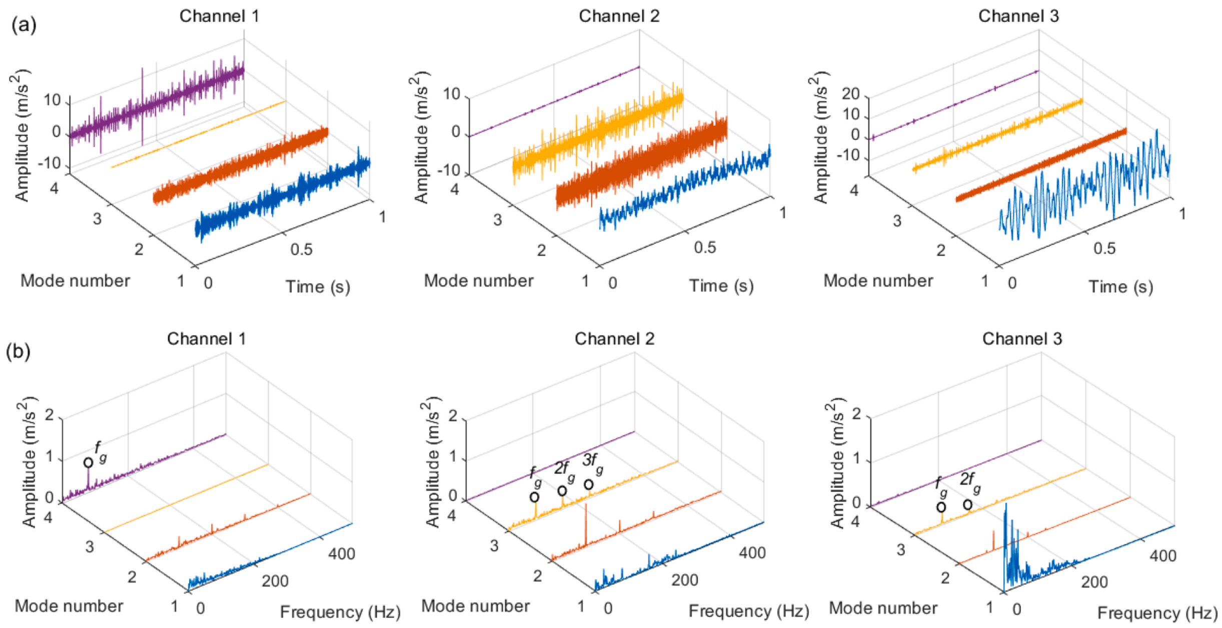


Fig. 30. Multi-sensor decomposition result of wind turbine vibration signal using MFMD under state 2.

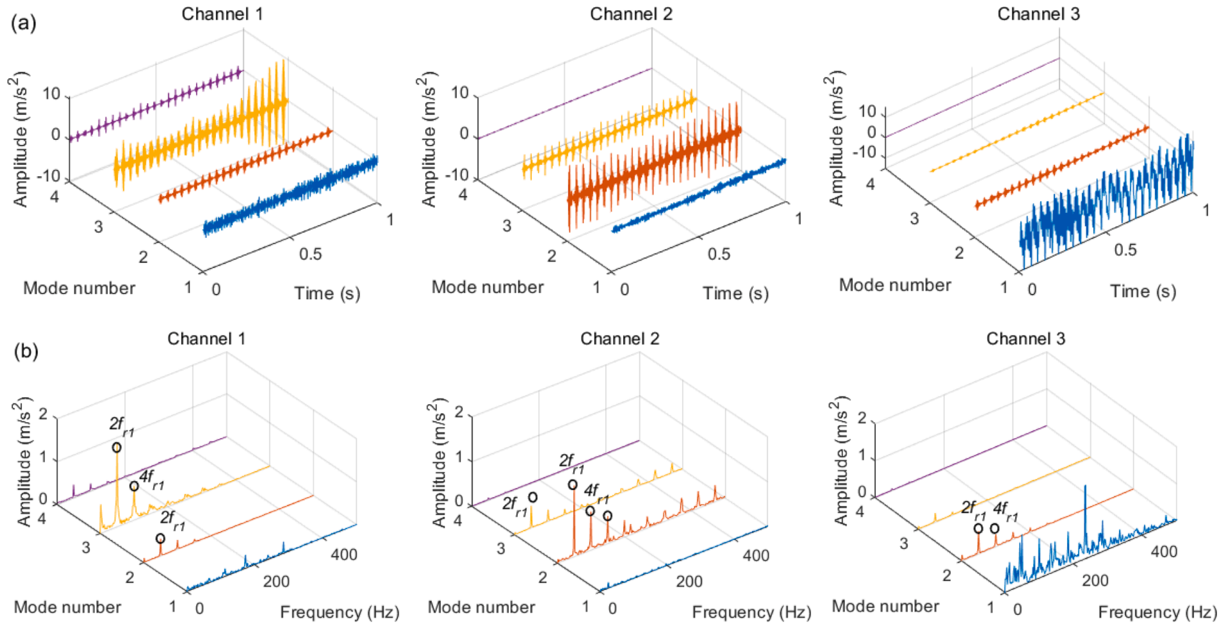


Fig. 31. Multi-sensor decomposition result of wind turbine vibration signal using MFMD under state 3.

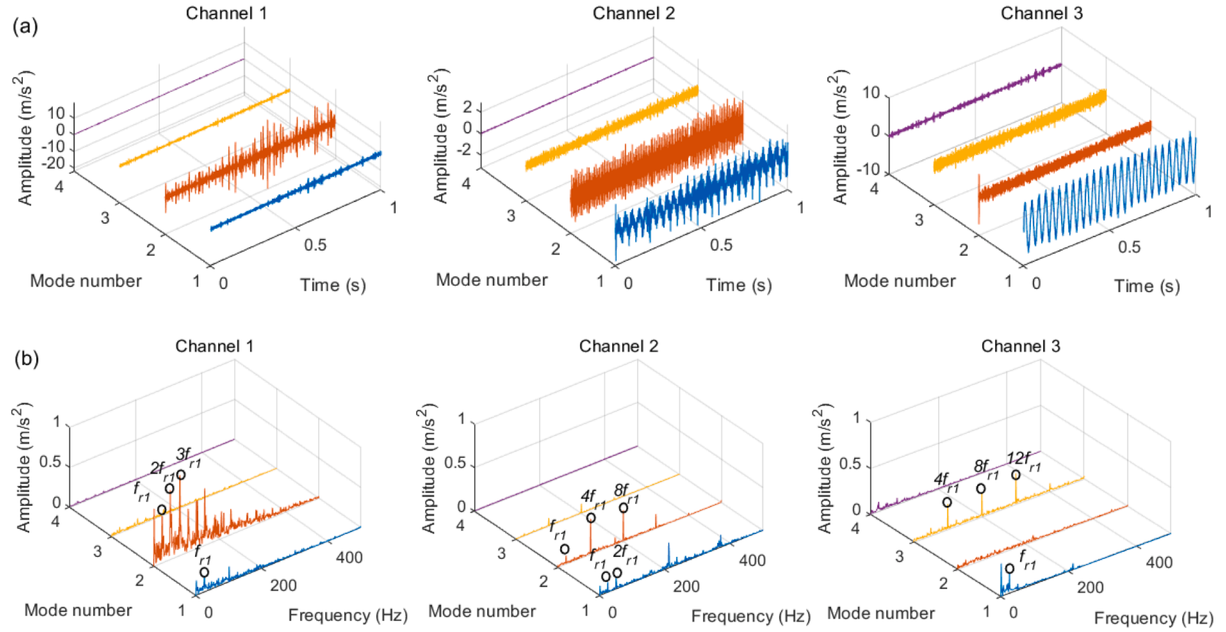


Fig. 32. Multi-sensor decomposition result of wind turbine vibration signal using MFMD under state 4.

and other representative deep network models, respectively. Seen from Fig. 35, the recognition accuracy of MAFResCNN is 100% (400/400), which is higher than that of other approaches (i.e., MK-ResCNN, ResCNN, MSCNN, CNN and LSTM) which are 99.50% (398/400), 98.75% (395/400), 98.00% (392/400), 97.25% (389/400) and 96.50% (386/400), respectively. This comparison results demonstrate the validity and superiority of the proposed MAFResCNN model in wind turbine intelligent fault identification.

To account for the efficacy of AMFMD in the presented approach, the same wind turbine multi-sensor data are processed by integrating two similar decomposition methods (i.e., MVMD and MEMD) with MAFResCNN. Furthermore, all methods are conducted 10 times to avoid the randomness of identifying results, enabling us to make an objective and fair comparison. Fig. 36 shows the recognition accuracy of 10 trials of different combination approaches, and Table 10 lists the detailed diagnostic results. As can be observed from Fig. 36 and Table 10, the proposed approach (i.e., AMFMD and MAFResCNN) can obtain average recognition accuracy

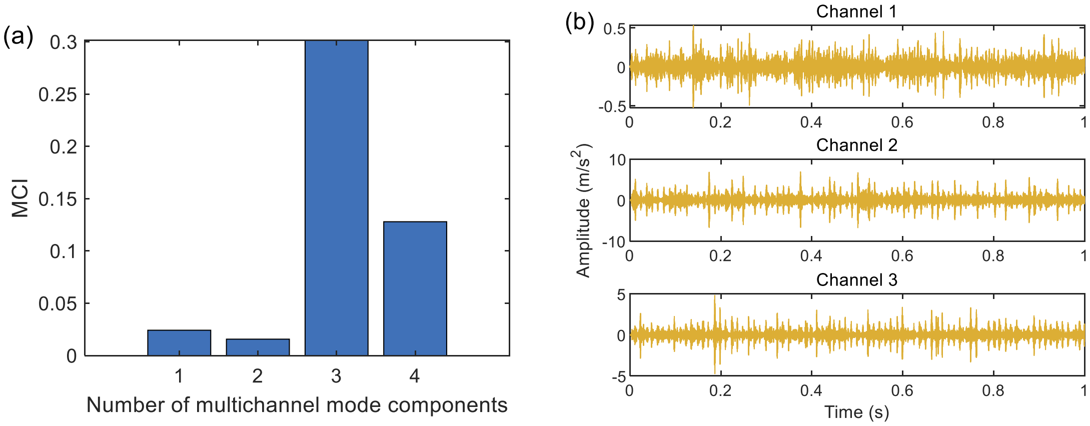


Fig. 33. Mode selection results of wind turbine multi-sensor signal under state 2: (a) MCI and (b) the selected multichannel mode components.

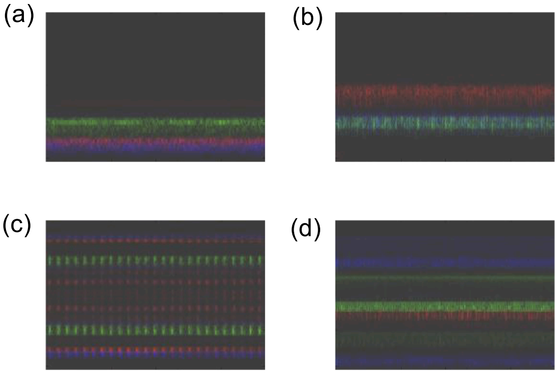
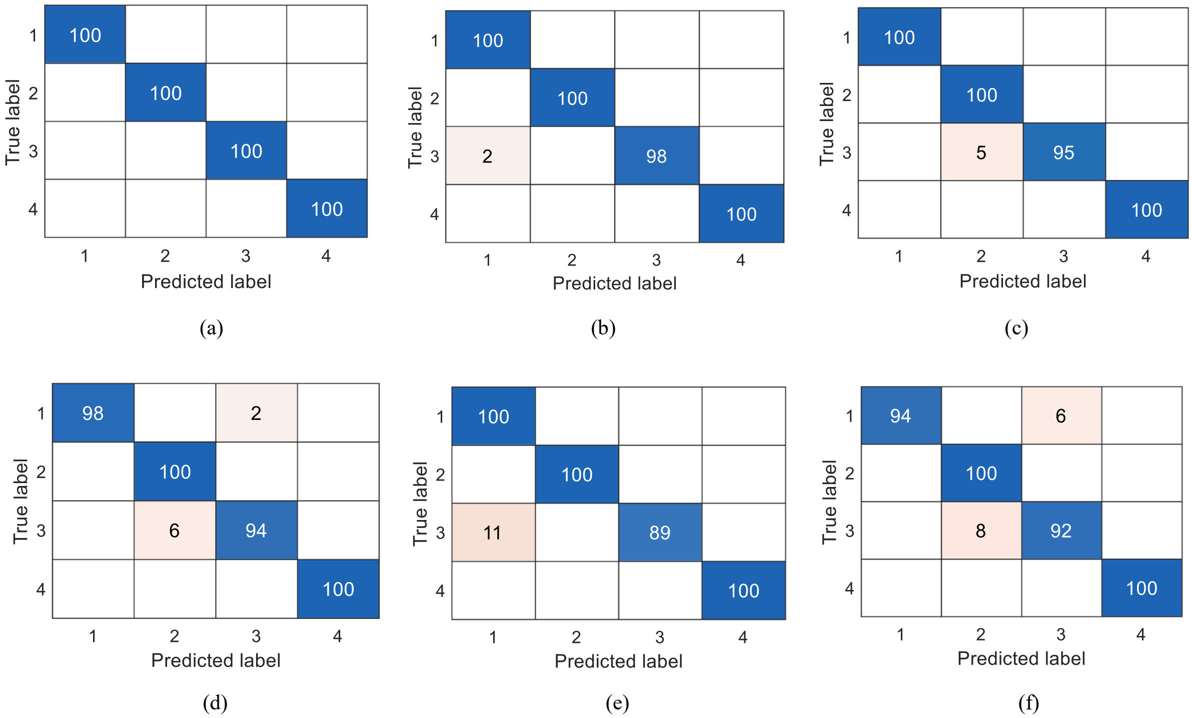


Fig. 34. MTFR of different wind turbine operation states: (a) state 1, (b) state 2, (c) state 3 and (d) state 4.

of 99.98%, which is larger than that of the other two combination approaches (i.e., MVMD and MAFResCNN, MEMD and MAFResCNN), which are 99% and 98.43%, respectively. Additionally, the proposed approach has the smallest standard deviation compared with the other two combination approaches. This not only shows that AMFMD in the proposed method is useful and necessary, but also shows that the proposed method has good recognition stability.

To further verify the usefulness of the presented approach in multi-sensor fault diagnosis, the presented approach is compared with some existing reported fault diagnosis methods (i.e., PSO-VMD + SVM, CWT + CNN, HHT + CNN, SST + CNN, CSCoh + CNN, FSWT + CNN, CWT + ResCNN, MSCNN + BiLSTM). Similarly, to ensure the reliability of diagnostic results of various approaches, 10 trials are repeated under the same configuration of each method. Fig. 37 draws the recognition accuracy of 10 trials of all comparison approaches and Table 11 displays synthetically the diagnosis results of different approaches, including average recognition accuracy, standard deviation and CPU time. According to Fig. 37 and Table 11, the average recognition accuracy of the proposed method with multi-sensor data is 99.98% (3999/4000), which is higher than that of all other comparison methods (i.e., method 1 ~ 7) with single sensor data. This proves the availability of our presented method with multi-sensor data in wind turbine fault diagnosis. Even though the proposed method and other comparison methods (i.e., method 1 ~ 7) are used for analyzing the same single sensor data, the average recognition accuracy of the proposed method is also larger, which indicates that the superiority of the proposed method is further explained. Additionally, compared to the previously reported multi-sensor method (MSCNN + BiLSTM), the average recognition accuracy of the proposed method can also be improved by 2.85%. In terms of standard deviation, the standard deviation (0.0791) of the proposed method is the smallest. In other words, the stability of the proposed method also wins over other comparison methods. Additionally, due to the way of achieving multi-sensor information fusion without increasing the number of samples, the CPU time of the proposed method in the model training process has competitiveness among these methods.

Similar to case 1, to illustrate the effectiveness and superiority of MCI-based mode selection used in the proposed method, for wind turbine multi-sensor data containing the different noise levels (i.e., 5 dB, 0 dB and −5 dB), the identification results of MCI-based mode selection and SESGI-based signal reconstruction are calculated. Table 12 lists their comparison results. As seen from Table 12, whether recognition accuracy or time cost, MCI-based mode selection has superiority compared with SESGI-based signal reconstruction. This further demonstrates that MCI-based mode selection used in the proposed method is effective and feasible.



(a) (b) (c) (d) (e) (f)

Fig. 35. Identification results of different approaches in wind turbine multi-sensor signal: (a) Proposed approach, (b) MK-ResCNN, (c) ResCNN, (d) MSCNN, (e) CNN and (f) LSTM.

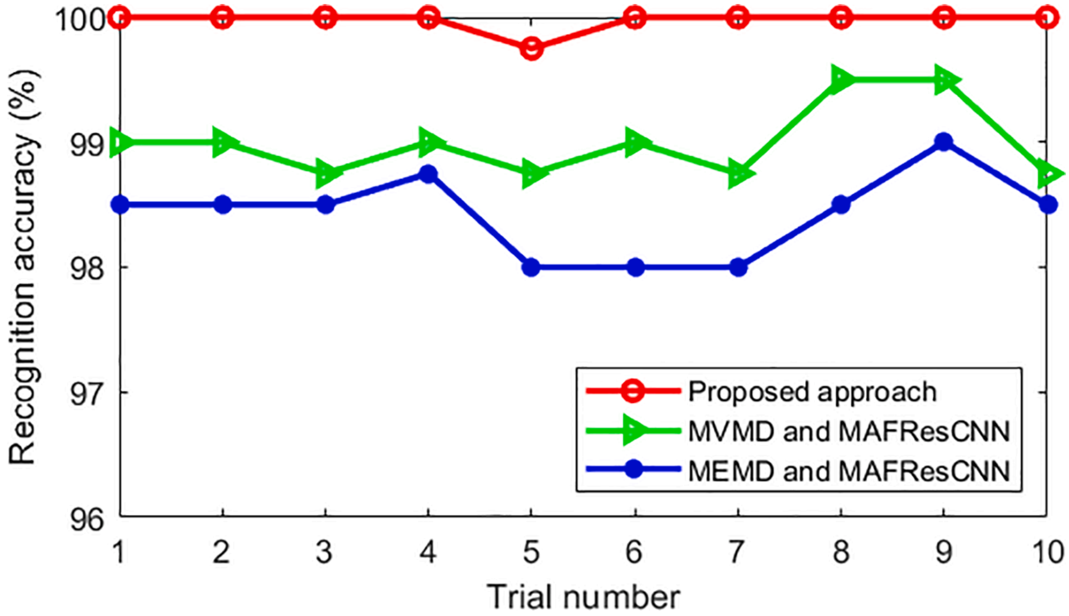


Fig. 36. Recognition accuracy of 10 trials of different combination approaches in case 2.

To show the validity and necessity of the fused MTFR used in the proposed method, the proposed method with and without multi-sensor information fusion are adopted to analyze the same wind turbine multi-sensor data. Table 13 lists the comparison results. It can be found from Table 13 that, compared with the proposed method containing the independent FSWT, the proposed method containing the fused MTFR has higher recognition accuracy and less model training time. This once again demonstrates the availability and superiority of multi-sensor information fusion used in the proposed method.

Table 10
Recognition results of 10 trials of different approaches in case 2.

Different approaches	Maximum	Minimum	Mean	Standard deviation
The proposed approach	100%	99.75%	99.98%	0.0791
MVMD and MAFResCNN	99.50%	98.75%	99.00%	0.2887
MEMD and MAFResCNN	99.00%	98.00%	98.43%	0.3344

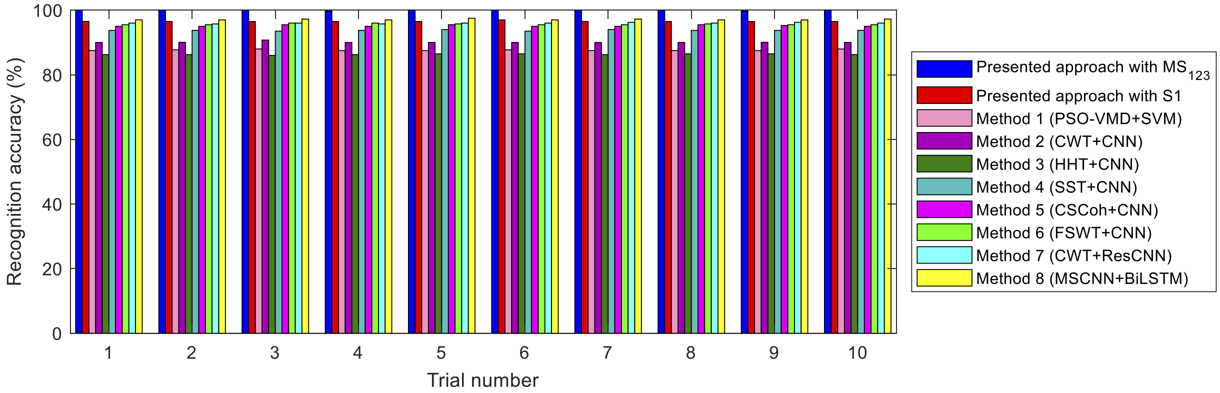


Fig. 37. Recognition accuracy of 10 trials of different comparison approaches in case 2.

Table 11
Recognition results of 10 trials of different approaches in case 2.

Methods with different sensor data	Average accuracy (%)	Standard deviation (SD)	CPU time (s)
Presented approach with MS_{123}	99.98% (3999/4000)	0.0791	20.35
Presented approach with S_1	96.73% (3869/4000)	0.1845	19.67
Method 1 (PSO-VMD + SVM) with S_1	87.65% (3506/4000)	0.2108	45.62
Method 2 (CWT + CNN) with S_1	90.08% (3603/4000)	0.2346	23.79
Method 3 (HHT + CNN) with S_1	86.33% (3453/4000)	0.1687	18.02
Method 4 (SST + CNN) with S_1	93.75% (3750/4000)	0.1667	25.79
Method 5 (CSCoh + CNN) with S_1	95.17% (3807/4000)	0.2372	26.21
Method 6 (FSWT + CNN) with S_1	95.65% (3826/4000)	0.2108	25.79
Method 7 (CWT + ResCNN) with S_1	96.00% (3840/4000)	0.1667	28.46
Method 8 (MSCNN + BiLSTM) with MS_{123}	97.13% (3885/4000)	0.1768	59.37

Note: MS_{123} is multi-sensor data with three channels and S_1 are single-sensor data with one channel.

Similarly, to display the effectiveness and necessity of MAFResCNN used in the proposed method, take the noisy wind turbine multi-sensor data with −5 dB under state 2 as an example, Fig. 38(a) and (b) plot multichannel mode components obtained by AMFMD and their corresponding envelope spectrum, respectively. As shown in Fig. 38, it is difficult to find generator bearing outer race fault frequency f_g by directly observing the envelope spectrum. Therefore, to enhance fault features and automatically achieve fault recognition, after the AMFMD-based signal decomposition, FSWT of multichannel mode components containing the most abundant fault information is further calculated and fused to obtain the fused MTFR, and the fused MTFR are fed into the proposed MAFResCNN for feature learning and intelligent fault identification. For the convenience of analysis, Fig. 39 shows feature visualization results by t-SNE under two scenarios (i.e., The original data and the noisy data with −5 dB noise adding). As seen from Fig. 39, whether the input data with or without adding noise, compared with the input data and the AMFMD-based envelope spectrum, the features learned by MAFResCNN are clearer for identifying wind turbine fault types. This once again indicates the effectiveness and necessity of MAFResCNN used in the proposed method.

Table 12
Diagnosis results of different data processing way for wind turbine vibration signal with different noise levels in case 2.

Data processing way	Noisy signal with 5 dB		Noisy signal with 0 dB		Noisy signal with −5 dB	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
MCI-based mode selection	99.53 ± 0.0791	4.263	98.55 ± 0.1054	4.415	96.83 ± 0.1208	4.542
SESGI-based signal reconstruction	98.50 ± 0.1179	5.282	96.13 ± 0.1318	5.374	93.55 ± 0.1581	5.638

Table 13
Diagnosis results of the presented approach with and without information fusion in case 2.

Methods	Accuracy (%)	CPU time (s)
The presented approach with the fused MTFR	99.98 ± 0.0791	20.35
The presented approach with the independent FSWT	98.78 ± 0.1419	38.59

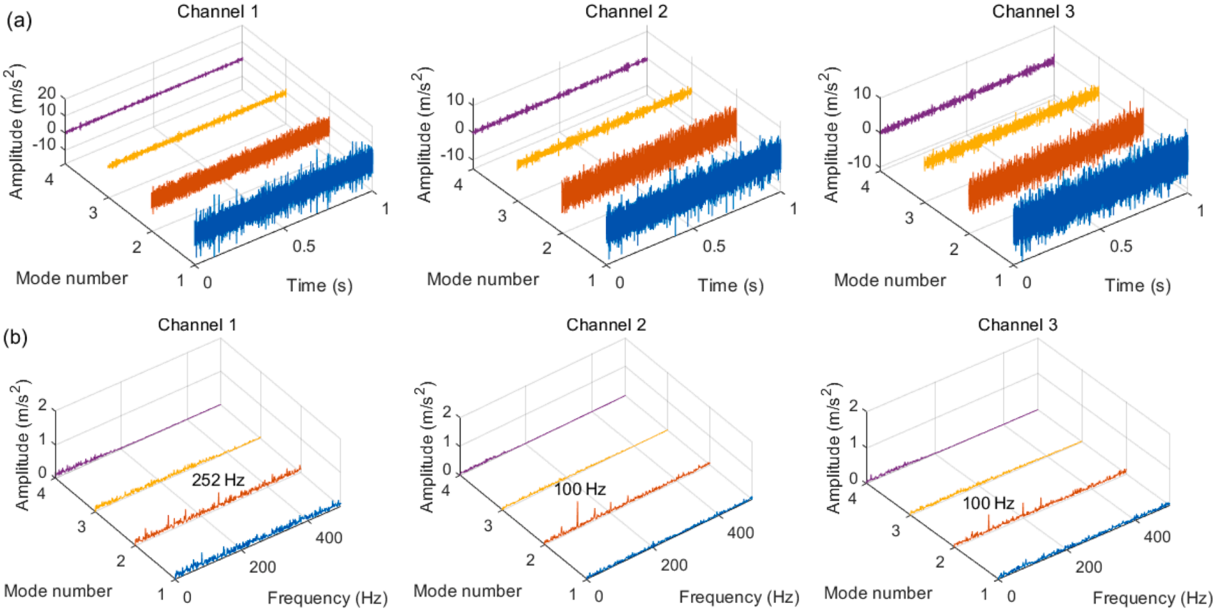
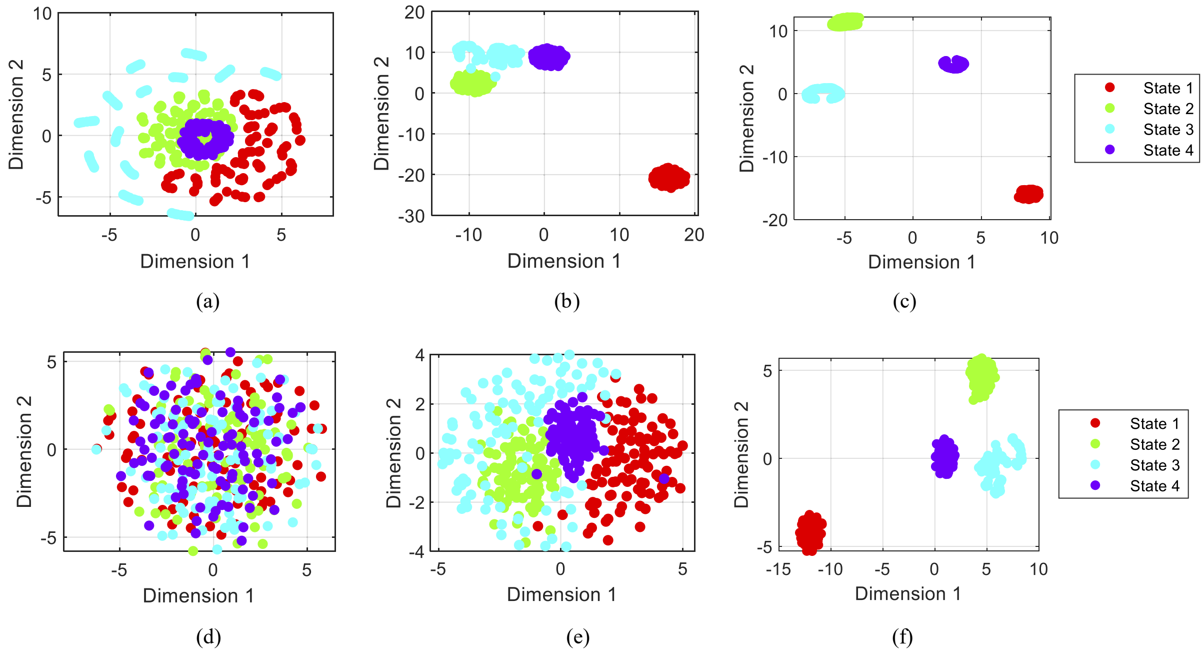


Fig. 38. Results of AMFMD in the noisy wind turbine multi-sensor data under state 2: (a) multichannel mode components and (b) their envelope spectrum.



(a) (b) (c)

(d) (e) (f)

Fig. 39. Feature visualization by t-SNE under two scenarios (i.e., The original data and the noisy data with −5 dB noise adding) in case 2: (a) The original input data; (b) AMFMD-based envelope spectrum data of (a); (c) the output of MAFResCNN of (a); (d) the noisy input data; (e) AMFMD-based envelope spectrum data of (d); (f) the output of MAFResCNN of (d).

5.3. Discussion and research prospects

Taken altogether, the main motivation of our study is to propose a new machinery multi-sensor intelligent fault diagnosis method without unique specialist experience. Specifically speaking, AMFMD and MAFResCNN are used in conjunction with each other to achieve automatic feature extraction and intelligent fault identification, which not only have good robustness against noise, but also promote fault identification accuracy compared with some similar techniques. Although two experimental examples and a series of comparative analysis have verified the effectiveness and superiority of the proposed method from multiple perspectives, it is necessary to objectively point out the problems left by our method and several aspects worth studying. The detailed research prospects are summarized as follows:

- Multi-sensor information fusion is recently regarded as a hot research direction, which can comprehensively process multichannel data or image information [66]. The familiar information fusion strategy mainly include data-level, feature-level and decision-level fusion, where the implementation way of data-level fusion includes the direct concatenation fusion, weighted average fusion and Laplacian pyramid transform fusion etc. The proposed method achieves multi-sensor information fusion in the input layer of network model, which can be understood as the simplest direct concatenation fusion way of data-level, but other fusion ways are not yet considered. Therefore, more multi-sensor information fusion technology will be investigated in our future work.
- Although the proposed approach with multi-sensor information fusion provides a new perspective for mechanical health status monitoring and fault diagnosis, it needs artificial sample markers in the fault identification process. In other words, manual labeling of samples will have a high cost for enterprises. Therefore, in our future work, some semi-supervised or unsupervised models (e.g., deep weighted adversarial network [67] and unsupervised transfer network [68,69]) can be considered as substitutes to save the cost caused by manual labeling.
- Considering that time cost of our proposed method with multivariate signal decomposition technology-based data prepossessing is relatively higher compared to the intelligent fault diagnosis method without data preprocessing. Consequently, how to effectively improve the computational efficiency and meanwhile guarantee the satisfactory recognition accuracy can also be regarded as a valuable subject in our future work.
- Although the proposed method can obtain a satisfactory recognition accuracy, it is difficult to explain the features learned by the proposed method in theory. Therefore, in our future research plans, some interpretable modules (e.g., wavelet basis, principal component analysis, singular value decomposition and morphological analysis) can be considered to embed into the proposed network architecture to enhance the interpretability of learning features in the circumstances of ensuring efficient diagnostic results.

6. Conclusions

In this study, a novel approach based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network is presented for machinery multi-sensor fault diagnosis. The proposed adaptive multivariate feature mode decomposition not only can synchronously process multi-sensor data to obtain multichannel mode components with mode alignment property, but also avoids the disadvantage of manual selection of two key parameters (i.e., mode number K and filter size L) in the original feature mode decomposition. Additionally, the multi-attention fusion residual convolutional neural network is designed by integrating the squeeze-excitation module (SEM) and convolutional block attention module (CBAM) into the traditional residual convolutional neural network to sufficiently learn the global and local information of the input data, which can promote the feature learning performance of the network model and relieve performance degradation problem existing in traditional residual convolutional neural network. The effectiveness and superiority of the proposed approach is verified by analyzing two case studies and conducting contrastive analysis of various aspects. It can be concluded from the comparative results that the proposed adaptive multivariate feature mode decomposition perform better for multichannel data analysis than the previously reported multivariate empirical mode decomposition and multivariate variational mode decomposition. Moreover, the proposed approach outperforms some existing state-of-the-art algorithms for machinery intelligent fault identification. Therefore, the identification accuracy of the proposed approach is higher than that of several existing representative intelligent fault diagnosis methods (e.g., variational mode decomposition-based support vector machine, time frequency graph-based convolution neural network and time–frequency graph-based residual convolutional neural network).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

The authors would like to thank the editors and the anonymous reviewers for their valuable comments.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the National Natural Science Foundation of China (No. 52005265), the Natural Science Fund for Colleges and Universities in Jiangsu Province (No. 20KJB460002), Macau Young Scholars Program (No. AM2021002), the Science and Technology Development Fund, Macau SAR (File no.: 0101/2021/A2, 0010/2021/AGJ, SKL-IOTSC(UM)-2021-2023), the Research Committee of University of Macau (No. MYRG2020-00073-IOTSC, MYRG2022-00096-IOTSC) and Guangdong-Hong Kong-Macau Joint Laboratory Program (No. 2020B1212030009).

References

[1] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—A tutorial, Mech. Syst. Signal Process. 25 (2) (2011) 485–520.
[2] K. Feng, J.C. Ji, Q. Ni, et al., A review of vibration-based gear wear monitoring and prediction techniques, Mech. Syst. Signal Process. 182 (2023), 109605.
[3] K. Feng, J.C. Ji, Y. Zhang, et al., Digital twin-driven intelligent assessment of gear surface degradation, Mech. Syst. Signal Process. 186 (2023), 109896.
[4] Z. Wang, J. Zhou, W. Du, et al., Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution, Mech. Syst. Signal Process. 162 (2022), 108018.
[5] S. Wei, Q. He, D. Wang, Zhike Peng, Two-level variational chirp component decomposition for capturing intrinsic frequency modulation modes of planetary gearboxes, Mech. Syst. Signal Process. 177 (2022), 109182.
[6] J. Zheng, H. Pan, S. Yang, J. Cheng, Adaptive parameterless empirical wavelet transform based time-frequency analysis method and its application to rotor rubbing fault diagnosis, Signal Process. 130 (2017) 305–314.
[7] H. Pan, H. Xu, J. Zheng, Q. Liu, J. Tong, Cyclic symplectic component decomposition with application in planetary gearbox fault diagnosis, Mech. Mach. Theory 172 (2022), 104792.
[8] C. Yin, Y.L. Wang, G. Ma, Y. Wang, Y. Sun, Y. He, Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed EMD and adaptive threshold denoising, Mech. Syst. Signal Process. 171 (2022), 108834.
[9] Z. Liu, Y. Jin, M.J. Zuo, Z. Feng, Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis, Mech. Syst. Signal Process. 95 (2017) 468–487.
[10] L. Wang, Z. Liu, An improved local characteristic-scale decomposition to restrict end effects, mode mixing and its application to extract incipient bearing fault signal, Mech. Syst. Signal Process. 156 (2021), 107657.
[11] H. Pan, Y.u. Yang, X. Li, J. Zheng, J. Cheng, Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis, Mech. Syst. Signal Process. 114 (2019) 189–211.
[12] M. Zhang, Z. Jiang, K. Feng, Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump, Mech. Syst. Signal Process. 93 (2017) 460–493.
[13] Z. Li, J. Chen, Y. Zi, J. Pan, Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive, Mech. Syst. Signal Process. 85 (2017) 512–529.
[14] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, Z. Zhu, A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines, Mech. Syst. Signal Process. 116 (2019) 668–692.
[15] G. Li, G. Tang, G. Luo, H. Wang, Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition, Mech. Syst. Signal Process. 120 (2019) 83–97.
[16] Q. Ni, J.C. Ji, K. Feng, et al., A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis, Mech. Syst. Signal Process. 164 (2022), 108216.
[17] Y. Miao, B. Zhang, C. Li, et al., Feature mode decomposition: New decomposition theory for rotating machinery fault diagnosis, IEEE Trans. Ind. Electron. 70 (2) (2022) 1949–1960.
[18] Y. Chen, M.J. Zuo, A sparse multivariate time series model-based fault detection method for gearboxes under variable speed condition, Mech. Syst. Signal Process. 167 (2022), 108539.
[19] N. Rehman, D. P. Mandic, Multivariate empirical mode decomposition. Proceedings of the Royal Society A: Mathematical, Phys. Eng. Sciences 466 (2117) (2010) 1291-1302.
[20] J. Zhou, Y. Yang, X. Li, et al., Multivariate local characteristic-scale decomposition and 1.5-dimensional empirical envelope spectrum based gear fault diagnosis, Mech. Mach. Theory 172 (2022), 104772.
[21] N.u. Rehman, H. Aftab, Multivariate variational mode decomposition, IEEE Trans. Signal Process. 67 (23) (2019) 6039–6052.
[22] L. Zuo, L. Zhang, Z.-H. Zhang, X.-L. Luo, Y.u. Liu, A spiking neural network-based approach to bearing fault diagnosis, J. Manuf. Sys. 61 (2021) 714–724.
[23] D. Yao, J. Yang, Z. Pang, C. Nie, F. Wen, Railway axle box bearing fault identification using LCD-MPE and ELM-Adaboost, J. Vibroeng. 20 (1) (2018) 165–174.
[24] Z. Jin, D. He, R. Ma, X. Zou, Y. Chen, S. Shan, Fault diagnosis of train rotating parts based on multi-objective VMD optimization and ensemble learning, Digit. Signal Process. 121 (2022), 103312.
[25] Q. Zhang, S. Chen, Z.P. Fan, Bearing fault diagnosis based on improved particle swarm optimized VMD and SVM models, Adv. Mech. Eng. 13 (2021) 16878140211028451.
[26] X. Chen, Y. Yang, Z. Cui, J. Shen, Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy, Energy 174 (2019) 1100–1109.
[27] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Signal Process. 138 (2020), 106587.
[28] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237.
[29] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, K. Gryllias, A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges, Mech. Syst. Signal Process. 167 (2022), 108487.
[30] Z. Wang, X. He, B. Yang, N. Li, Subdomain adaptation transfer learning network for fault diagnosis of roller bearings, IEEE Trans. Ind. Electron. 69 (8) (2022) 8430–8439.
[31] L.-H. Wang, X.-P. Zhao, J.-X. Wu, Y.-Y. Xie, Y.-H. Zhang, Motor fault diagnosis based on short-time Fourier transform and convolutional neural network, Chin. J. Mech. Eng. 30 (6) (2017) 1357–1368.
[32] Y. Cheng, M. Lin, J. Wu, H. Zhu, X. Shao, Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network, Knowl-Based Syst. 216 (2021), 106796.
[33] A. Kumar, C.P. Gandhi, Y. Zhou, R. Kumar, J. Xiang, Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images, Appl. Acoust. 167 (2020), 107399.

[34] M. Islam, J.-M. Kim, Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network, Comput. Ind. 106 (2019) 142–153.

[35] J. Gu, Y. Peng, H. Lu, X. Chang, G. Chen, A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN, Measurement 200 (2022), 111635.

[36] H. Liang, J. Cao, X. Zhao, Average descent rate singular value decomposition and two-dimensional residual neural network for fault diagnosis of rotating machinery, IEEE Trans. Instrum. Meas. 71 (2022) 3512616.

[37] M.-F. Guo, N.-C. Yang, W.-F. Chen, Deep-learning-based fault classification using Hilbert-Huang transform and convolutional neural network in power distribution systems, IEEE Sens. J. 19 (16) (2019) 6905–6913.

[38] Z. Wang, W. Zhao, W. Du, N. Li, J. Wang, Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network, Process Saf. Environ. Prot. 149 (2021) 591–601.

[39] L. Ke, Y. Zhang, B. Yang, Z. Luo, Z. Liu, Fault diagnosis with synchrosqueezing transform and optimized deep convolutional neural network: An application in modular multilevel converters, Neurocomputing 430 (2021) 24–33.

[40] Z. Chen, A. Mauricio, W. Li, K. Gryllias, A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks, Mech. Syst. Signal Process. 140 (2020), 106683.

[41] Y. Kim, K. Na, B.D. Youn, A health-adaptive time-scale representation (HTSR) embedded convolutional neural network for gearbox fault diagnostics, Mech. Syst. Signal Process. 167 (2022), 108575.

[42] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, IEEE Trans. Ind. Electron. 66 (4) (2019) 3208–3216.

[43] R. Liu, F. Wang, B. Yang, S.J. Qin, Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions, IEEE Trans. Ind. Inform. 16 (6) (2020) 3797–3806.

[44] W. Zhao, Z. Wang, W. Cai, Q. Zhang, J. Wang, W. Du, N. Yang, X. He, Multiscale inverted residual convolutional neural network for intelligent diagnosis of bearings under variable load condition, Measurement 188 (2022), 110511.

[45] K.N. Ravikumar, A. Yadav, H. Kumar, K.V. Gangadharan, A.V. Narasimhadhan, Gearbox fault diagnosis based on multi-scale deep residual learning and stacked LSTM model, Measurement 186 (2021), 110099.

[46] H. Wei, Q. Zhang, M. Shang, Y. Gu, Extreme learning machine-based classifier for fault diagnosis of rotating machinery using a residual network and continuous wavelet transform, Measurement 183 (2021), 109864.

[47] Y. Miao, M. Zhao, V. Makis, J. Lin, Optimal swarm decomposition with whale optimization algorithm for weak feature extraction from multicomponent modulation signal, Mech. Syst. Signal Process. 122 (2019) 673–691.

[48] D. Wang, Z. Peng, L. Xi, The sum of weighted normalized square envelope: A unified framework for kurtosis, negative entropy, Gini index and smoothness index for machine health monitoring, Mech. Syst. Signal Process. 140 (2020), 106725.

[49] J. Cheng, Y. Yang, N. Hu, Z. Cheng, J. Cheng, A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis, Mech. Syst. Signal Process. 149 (2021), 107351.

[50] B. Hou, D. Wang, T. Xia, L. Xi, Z. Peng, K. Tsui, Generalized Gini indices: Complementary sparsity measures to Box-Cox sparsity measures for machine condition monitoring, Mech. Syst. Signal Process. 169 (2022), 108751.

[51] I.S. Bozchalooi, M. Liang, A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection, J. Sound Vib. 308 (1-2) (2007) 246–267.

[52] Y. Miao, M. Zhao, J. Lin, X. Xu, Sparse maximum harmonics-to-noise-ratio deconvolution for weak fault signature detection in bearings, Meas. Sci. Technol. 27 (2016), 105004.

[53] Z. Huo, M. Martinez-Garcia, Y.u. Zhang, R. Yan, L. Shu, Entropy measures in machine fault diagnosis: insights and applications, IEEE Trans. Instrum. Meas. 69 (6) (2020) 2607–2620.

[54] Y. Cheng, S. Wang, B. Chen, G. Mei, W. Zhang, H. Peng, G. Tian, An improved envelope spectrum via candidate fault frequency optimization-gram for bearing fault diagnosis, J. Sound Vib. 523 (2022), 116746.

[55] Y. Miao, M. Zhao, J. Lin, Y. Lei, Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings, Mech. Syst. Signal Process. 92 (2017) 173–195.

[56] J. Hu, L. Shen, G. Sun, Ieee, Squeeze-and-excitation networks, 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), UT, Salt Lake City, 2018, pp. 7132–7141.

[57] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018, pp. 3–19.

[58] J. Liu, C. Zhang, X. Jiang, Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet, Mech. Syst. Signal Process. 168 (2022), 108664.

[59] X. Yan, Y. Liu, D. Huang, M. Jia, A new approach to health condition identification of rolling bearing using hierarchical dispersion entropy and improved Laplacian score, Struct. Health Monit. 20 (3) (2021) 1169–1195.

[60] X. Yan, W. Yan, K.-V. Yuen, Z. Yang, X. Wang, An adaptive variational mode extraction method based on multi-domain and multi-objective optimization for bearing fault diagnosis, Struct. Health Monit. 22 (4) (2023) 2708–2733.

[61] Y. Chen, M. Rao, K. Feng, M.J. Zuo, Physics-Informed LSTM hyperparameters selection for gearbox fault detection, Mech. Syst. Signal Process. 171 (2022), 108907.

[62] Z. Xu, X. Mei, X. Wang, M. Yue, J. Jin, Y. Yang, C. Li, Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors, Renew. Energy 182 (2022) 615–626.

[63] D. Wang, Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients, Mech. Syst. Signal Process. 108 (2018) 360–368.

[64] X. Yan, Y. Liu, Y. Xu, M. Jia, Multichannel fault diagnosis of wind turbine driving system using multivariate singular spectrum decomposition and improved Kolmogorov complexity, Renew. Energy 170 (2021) 724–748.

[65] X. Yan, M. Jia, Bearing fault diagnosis via a parameter-optimized feature mode decomposition, Measurement 203 (2022), 112016.

[66] Y. Ta, Y. Li, W. Cai, Q. Zhang, Z. Wang, L. Dong, W. Du, Adaptive staged remaining useful life prediction method based on multi-sensor and multi-feature fusion, Reliab. Eng. Syst. Safe. 231 (2023), 109033.

[67] Z. Wang, J. Cui, W. Cai, Y. Li, Partial transfer learning of multi-discriminator deep weighted adversarial network incross-machine fault diagnosis, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.

[68] G. Michau, O. Fink, Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer, Knowl-Based Syst. 216 (2021), 106816.

[69] Y. Xiao, H. Shao, SongYu Han, Z. Huo, J. Wan, Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain, IEEE/ASME Trans. Mechatron. 27 (6) (2022) 5254–5263.