Fault Diagnosis of Rolling Bearing of Wind Turbines Based on the Variational Mode Decomposition and Deep Convolutional Neural Networks

Zifei Xu[a], Chun Li [a], Yang Yang [a, b]

([a]. School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai 200093，P. R. China
[b]. Department of Maritime and Mechanical Engineering, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, UK)

Abstract: Machine learning techniques have been successfully applied in intelligent fault diagnosis of rolling bearings in recent years. However, in the real world industrial application, the dissimilarity of data due to changes in the working conditions and data acquisition environment often cause a poor performance of the existing fault diagnosis methods. Consequently, to address these inadequacies, this paper developed a novel method by integrating the Convolutional Neural Networks (CNNs) with the Variational Mode Decomposition (VMD) algorithms. Named as “Variational Mode Decomposition with Deep Convolutional Neural Networks (VMD-DCNNs)”, the method, in an end-to-end way, directly processes raw vibration signals without artificial experiences and manual intervention to realize the fault diagnosis of rolling bearings. In addition, the CNN technique is used to extract features from each Intrinsic Mode Function (IMF) in order to address the deficiency in extracting features from a single source and to achieve an effective and efficient fault diagnosis of rolling bearings under different environments and states. The value of parameter *K* of the VMD-DCNNs model is optimized by considering time complexity and generalization ability of the model. Lastly, bearing experiments are conducted to verify the superiority of the VMD-DCNNs in diagnosing fault under different conditions. The visualizations of the signals in the convolutional layer explain the reasonability in selecting the value of parameter *K* and they also indicate that the translational invariances in a raw IMF component have been learned by the VMD-DCNNs model.

Keywords: Deep learning; Convolutional neural networks; Variational mode decomposition; fault diagnosis; rolling bearing;

1. Introduction

As the fossil fuels continue to deplete, renewable and clean energy are increasingly attracting attention. Of the renewable energy sources, wind power has become a viable alternative to fossil fuels and it is attracting investments and global attention in research development [1]. In addition, wind power offers a faster means of adding electricity to the grid at a reduced cost. The world's wind power capacity was 566 GW in 2018, with the accumulated capacity expected to reach 2930 GW in 2040, accounting for 21.4% of the overall electricity generation [2].

The growth in the wind power industry has been very rapid over the past decade, resulting in wind turbines facing acute technological challenges due to their increasingly larger sizes and severe operation conditions [3]. One of these challenges is the reliability and maintenance cost of wind turbines, which has been a considerable factor in the uptake of renewable energy generation by investors and policymakers. The maintenance and operation costs about 75% of the total investments during the wind turbine service life [4]. The impact of the costs on the levelised cost of electricity (LCoE) for wind turbines is significant in determining its viability. Therefore, it is particularly important to reduce the failure rate of wind turbines in order to reduce its LCoE and ensure their service lives are free from major defects [5]. Most importantly, gearboxes and bearings, are among the most critical wind turbine components. Fig. 1 shows that bearing failure accounts for 64% of all failures in a typical wind turbine [6]. The efficiency and safety of wind turbines in working condition are significantly affected by the health of the bearings [7]. Consequently, it is crucial to find a reliable fault diagnosis methodology to accurately detect faults in the bearings in order to reduce the operation and maintenance costs, and extend, their service lives [8,9].

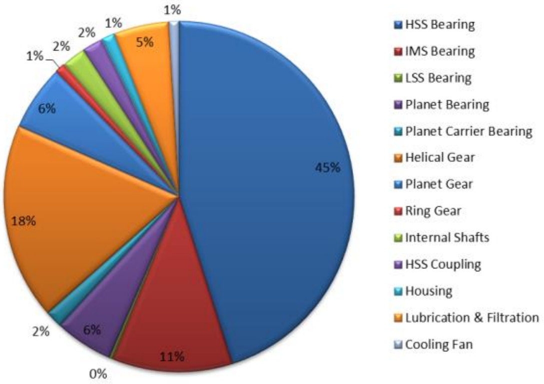


Fig. 1 Damage Distribution

The methodologies for analysis and diagnosis of bearing faults are commonly based on vibration signals which offer more useful information than other types of signals. Hitherto, various techniques of signal processing are used to analyze and extract vibration information from raw signals [10-13]. Many of these techniques rely on intelligent algorithms and they have been developed to analyze and diagnose bearing faults. Generally, a fault diagnosis framework is composed of three parts: data collection, feature extraction and condition classification. The most important step in the framework is the useful feature extraction, which determines the upper limit of the accuracy of fault classification [14]. Various advanced time-frequency analysis techniques are adapted for feature extraction and condition classification. Some of the notable techniques with good performance include Wavelet Transform (WT), Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Local Mean Decomposition (LMD), and Variational Mode Decomposition (VMD). Amongst these techniques, the EMD is a widely used option although it has some inherent shortcomings, such as the problems of the mode mixing and endpoint effects in data analysis when diagnosing a fault. On the other hand, the VMD algorithm offers a better capability in fault diagnosis because of its ability in solving the problems of the mode mixing and endpoint effects, produced by the EMD. The VMD method has successfully been used in the field of fault feature extraction. Chen *et al.* [15] used the VMD algorithm to decompose the raw bearing vibration signals. The energy entropies of each mode were introduced as the characteristic values to be fed into Support Vector Machine (SVM) for addressing the fault classification. The results show that the accuracy of VMD-Entropy-SVM in fault diagnosis was higher than that of WT-Entropy-SVM and EMD-Entropy-SVM. Zheng *et al*. [16] combined the VMD algorithm with the envelope demodulation method, and diagnosed the fault forms of bearing through envelope spectrum analysis of the decomposed modes. They also verified the reliability of the method based on standard rolling bearing experimental data. The results show that the rolling bearing fault types were accurately determined, and the decomposition results of the VMD method were better than those obtained using the EMD and other methods. Gu *et al*. [17] decomposed and reconstructed raw vibration signals using the VMD algorithm and Teager Energy Operator (TEO). The results indicate that the VMD-TEO method could improve the Signal to Noise Ratio (SNR) of the vibration signals and is capable of extracting the useful rolling bearing characteristics. The above studies confirm that the VMD algorithm has a higher effectiveness in feature extraction compared to other methods, such as EMD, EEMD and LMD.

Other tools such as, SVM, Multi-Layer Perceptron (MLP) and various deep learning techniques have been used to address the states classification. Li *et al*. [18] combined the VMD algorithm with the Back Propagation (BP) neural network in addressing the intelligent fault diagnosis of four working conditions of a rolling bearing. The capabilities of feature self-learning were considered. Shao *et al*. [19] integrated the deep wavelet auto-encoder within the extreme learning machine to finish the fault diagnosis of a rolling bearing. Wen *et al*. [20] proposed the use of Convolutional Neural Networks (CNNs) for fault diagnosis. They converted the raw signals into 2-D images, which were used as the input in the CNNs for extracting effective features. However, the diagnostic accuracy of the above method was not sufficiently high, implying that the model lacks generalization in coping with the changing environmental condition. In order to enhance the generalization ability of the model based on CNNs when the environmental condition is changing. Zhang *et al.* [21] proposed a new training interference CNN method, which directly utilized the raw vibration signal to realize the fault diagnosis of the rolling bearings subjected to noise environment and different workloads. The result show that the training interference CNN has better diagnosis performance than traditional CNNs.

In summary, there are two distinct groups in the field of fault diagnosis. The first group consists of a shallow learning algorithm and a feature extraction method based on artificial experiences used to address fault diagnosis. In the second group, the method uses the deep learning technique for the end-to-end feature extraction and state classification. The deep learning technology can deal with the shortcomings of shallow learning models that rely on human experiences and lack of generalization. However, the models based on the deep learning technique still have a poor generalization ability when the working environment changes. In addition, it will be difficult to maintain and update the model in subsequent fault diagnosis operations. The challenges in fault diagnosis of wind turbine bearings as revealed by the review of the above studies are summarized as follows:

Firstly, the working conditions of the bearings of a wind turbine are constantly changing. The measured vibration signals are complex due to environmental noises, data acquisition errors and variable bearing loads. Secondly, the diagnosis model based on the deep learning technique that uses images as the learning samples would have a shortage of capability of feature self-learning. Consequently, the extrapolation performance from the preceding method will not be of high quality. Thirdly, the diagnostic performance of the model would be deficient of the fault characteristics due to the singleness of the input signals.

In order to address these problems, this study has developed a method named Variational Mode

Decomposition with Deep Convolutional Neural Networks (VMD-DCNNs). The fault diagnosis accuracies and generalization abilities of the VMD-DCNNs model are examined using experimental data under different noise environments, variable loading conditions, sensor reset error scenarios and a progressively developing fault condition. Furthermore, the time series in the convolutional layers has been visualized to illustrate the learning mechanism in helping to realize the types of features that the model learns. In addition, the visualization results that shows the features learned from a residual IMF would support the selection of mode number *K* in the VMD-DCNNs model. The contributions of this paper are summarized as follows:

(1) A novel fault diagnosis system is developed by integrating the VMD with the CNNs. The advantages of the VMD in decoupling vibration and the superiorities of the CNNs in features extraction are utilized in the proposed VMD-DCNNs model.

(2) Some experimental signals containing environmental noise, variable load domains, sensor reset errors in data acquisition and the process of a failure occurrence of a bearing are used to test the extrapolation performance of the developed VMD-DCNNs model for its application to full-scale wind turbines.

(3) The parameter *K* of the VMD-DCNNs model is optimized by considering its associated time complexity and the diagnostic performance. The reason for choosing parameter *K* is tentatively explained by visualizing the time series in the convolutional layer. The features learned in the convolutional layer for each mode component have the same characteristics and they play a complementary role in, improving the performance of the VMD-DCNNs model.

The remaining parts of this paper are organized as follows. A brief description of VMD is given in Section 2. The intelligent fault diagnosis method based on VMD-DCNNs is introduced in Section

3. In addition, Section 4 analyzes the superiority and generalization ability of the proposed VMD-DCNNs model based on experimental data. The visualizations of the model are presented to show the features that have been learned by the developed VMD-DCNNs model are also presented in this section. Finally, the conclusions are presented in Section 5.

2. The Variational Mode Decomposition Algorithm

VMD is a new multi-component signal decomposition algorithm based on wiener filtering, Hilbert transform and heterodyne [22]. The algorithm aims to make decomposition of a target signal f into *K* modes u_k with sparsity properties achievable. In addition, the center frequency ω_k of each mode is assumed to have limited bandwidths. In the traditional VMD algorithm, the *k*th mode u_k is searched by constraining the variational model to minimize the estimated bandwidth. The constraint is satisfied that the sum of each mode is equal to the target signal. The 1-D input signal with a size of $N \times 1$ is assumed as the target signal f . The number of IMF components are evaluated by the preset parameter *K*.

The size of the analytic signal of each IMF component u_k is $N \times 1$. The analytic signal is obtained by using the Hilbert transform to convert every mode into a single-side spectrum. The estimated center frequency $e^{-j\omega_k t}$ is introduced to obtain the analysis signal of the mode as follows:

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t} \tag{1}$$

where t is the time, $\delta(t)$ represents the impulse function, and $\{u_k(t)\} = \{u_1(t), u_2(t), \ldots, u_k(t)\}$ represents the set of the decomposed *K* IMF modes. The set of the center frequencies of each IMF is given as $\{\omega_k(t)\} = \{\omega_1(t), \omega_2(t), \ldots, \omega_k(t)\}$

The constraint model is built by calculating the norm of the square L^2 of the given demodulation signal in Eq.(1) as follows

$$\begin{cases} \min\limits_{\{u_k\}\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum\limits_{k}^{K} u_k = f \end{cases} \tag{2}$$

where k stands for each mode, $k \in \{1, 2, \dots, K\}$, The partial derivative of time is represented by ∂_t, $j = \sqrt{-1}$.

The quadratic penalty factor α and Lagrange multiplication operator $\lambda(t)$ are introduced to change the constrained variational function to non-binding variational function. The extended Lagrange function is presented as follows:

$$\begin{aligned} & L(\{u_k\},\{\omega_k\},\{\lambda\}): \\ & = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k}^{K} u_k(t) \right\rangle \end{aligned} \tag{3}$$

where the reconstruction quantity is guaranteed by α when the signal contains a Gaussian noise. The Lagrange operator $\lambda(t)$ is introduced to enforce the constrains to be strict.

The Alternate Direction Method of Multiplier (ADMM) is used to solve the variational problem. The parameters $u_k^n(t)$, $\omega_k^n(t)$ and $\lambda_k(t)$ are alternately updated to find the saddle points of the Lagrange function. $u_k^{n+1}(t)$ is solved by Eq. (4).

$$\begin{aligned} & u_k^{n+1}(t) = \operatorname{argmin} \left\{ \alpha \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{i \neq k}^{K} u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\} \\ & i = \{1, 2, \dots, K\} i \neq k \end{aligned} \tag{4}$$

The Parseval/Plancherel Fourier isometry under the L^2 norm is introduced to solve the problem in the spectral domain and to change the variables $\omega \leftarrow \omega\text{-}\omega_k$ in the first term:

$$u_k^{n+1}(t) = \operatorname{argmin}\left\{ \alpha \left\| j(\omega-\omega_k)[(1+\operatorname{sgn}(\omega))\hat{u}_k(\omega)] \right\|_2^2 + \left\| \hat{f}(\omega) - \sum_{i \neq k}^{K} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right\|_2^2 \right\} \tag{5}$$

Eq. (5) is transformed to a nonnegative frequency interval-integral and its quadratic optimization solution is given in Eq. (6). The center frequency of each IMF is updated by Eq. (7) in the same process. The operator $\lambda(\omega)$ is updated by Eq. (8).

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i=1}^{k-1} \hat{u}_i^{n+1}(\omega) - \sum_{i=k+1}^{K} \hat{u}_i^{n}(\omega) + \frac{\hat{\lambda}_i(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{6}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega} \tag{7}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \tau \left(\hat{f}(\omega) - \sum_{k=1}^{K} \hat{u}_k^{n+1}(\omega) \right) \tag{8}$$

where, $\wedge$ denotes the Fourier transform; n is the number of iterations; and τ is the noise tolerance.

As revealed by [23], a notable influenced on the denoising performance could be produced by the variation of τ when it is not independent on the standard deviation of the added white Gaussian noise. Therefore, τ is set as a constant with a value of 0 [17,24,25] to obtain a better denoising effect for a signal with an arbitrary SNR.

The convergence tolerance in solving Eq. (9) is given as follows.

$$\sum_{k=1}^{K} \left(\left\| \hat{u}_k^{n+1}(\omega) - \hat{u}_k^{n}(\omega) \right\|_2^2 / \left\| \hat{u}_k^{n}(\omega) \right\|_2^2 \right) < \varepsilon \tag{9}$$

where ε is the convergence tolerance, and a typically value of 10^{-7} is adapted in this study.

Fig. 2 presents the flow chart of the VMD calculation process.

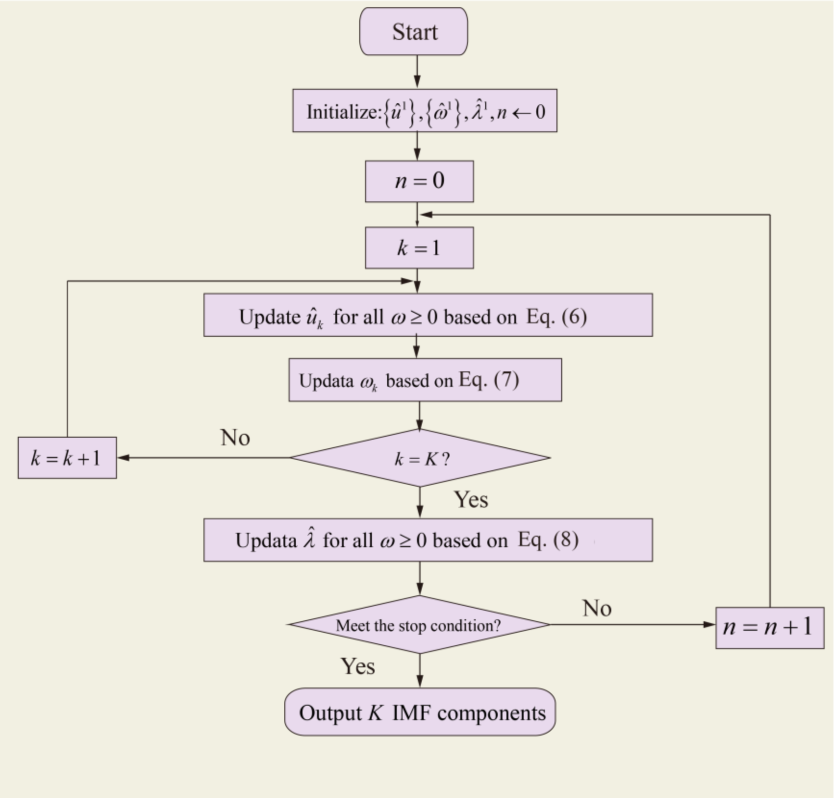


Fig. 2 The flow chart of the VMD

Step 1: Initialize $\{\hat{u}_k^1\}$, $\{\hat{\omega}_k^1\}$, $\{\hat{\lambda}^1\}$ and n in the first step, where $n\leftarrow 0$.

Step 2: Update $\{\hat{u}_k\}$ and $\{\hat{\omega}_k\}$ by using Eq. (6) and Eq. (7).

Step 3: Update the Lagrange multiplication operator by Eq. (8) when $\omega\geq 0$.

Step 4: Repeat steps 2-3 until the integration stopping criterion defined in Eq. (9) is satisfied.

3. Development of the VMD-DCNNs Fault Diagnosis System

3.1. VMD-DCNNs Architecture

This research's main goal is to realize an intelligent fault diagnosis of rolling bearings. However, the variable conditions of wind turbines produce some complexities and instability in collecting vibration signals. The poor extrapolation performance of the model based on traditional CNNs would fail in diagnosing bearing faults for such variable conditions. In order to improve the performance of the CNNs-based fault diagnosis model in a complex and changeable environment, the VMD algorithm is introduced to realize the functions of simultaneous signal filtering and data enhancement.

The VMD algorithm is used to decompose a set of 1-D raw vibration signals with a size of $N \times 1$ into K IMFs and a $N \times K$ singular matrix is obtained. More information is obtained by changing the value of K which significantly affects the performance of the VMD-DCNNs model. The $N \times K$ singular matrix is the input of the first convolution layer, $\boldsymbol{K}_i^l$ is the i^{th} filter in the layer l, and $\boldsymbol{X}^{l(\mathrm{R}^j)}$ denotes the j^{th} local area of the convolutional layer l. The advantage of the convolution kernel is that the characteristics of rotation invariance can be obtained by following the definition given in Eq. (10).

$$y^{l(i,j)} = \boldsymbol{K}_i^l \cdot \boldsymbol{X}^{l(\mathrm{R}^j)} = \sum_{j'=0}^{W} \boldsymbol{K}_i^l(j')\boldsymbol{X}^{l(j+j')} \tag{10}$$

where $y^{l(i,j)}$ is the dot product of the kernel and local area. W represents the width of the kernel. $\boldsymbol{K}_i^l(j')$ represents the j^{th} weight of the kernel l. The parameters in the model are normalized by the Batch Normalization (BN) operation to improve the generalization capability of a CNNs-based method in fault diagnosis.

The BN is usually placed following the convolutional layer or the full connection layer, or before the activation layer. The n-dimensional array $\mathbf{y}^l = (\mathbf{y}^{l(1)}, \mathbf{y}^{l(2)}, \ldots, \mathbf{y}^{l(n)})$ corresponding to the l^{th} BN layer is represented by $\mathbf{y}^{l(i)} = (y^{l(i,1)}, y^{l(i,2)}, \ldots, y^{l(i,n)})$ and $\mathbf{y}^{l(i)} = y^{l(i)} = y^{l(i,1)}$, when the BN layer is placed following convolutional layer and fully connected layer, respectively. The equation of the BN operation is described as follows:

$$\hat{y}^{l(i,j)} = \frac{y^{l(i,j)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}, z^{l(i,j)} = \gamma^{l(i)}\hat{y}^{l(i,j)} + \beta^{l(i)} \tag{11}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y^{l(i,j)} \tag{12}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (y^{l(i,j)} - \mu)^2 \tag{13}$$

where $z^{l(i,j)}$ represents the output of one neuron. μ and σ^2 are, respectively, the mean and variances of $y^{l(i,j)}$. ε is a small constant introduced to prevent the calculation from being invalid when the variance is 0. The $\gamma^{l(i)}$ and $\beta^{l(i)}$ are, respectively, the scale and shift parameters to be learned.

It is essential to add an activation function after the convolutional layer in order to enhance the non-linear expression ability of the input signal and to make the learned features more distinguishable. In recent years, Rectified Linear Unit (ReLU), which is the most widely used activation unit, has been applied to CNNs to accelerate its convergence. The ReLU is combined with a back propagation learning method to adjust the parameters when making the shallow weights become more trainable. The formula of the ReLU is presented in Eq. (14):

$$a^{l(i,j)} = f(z^{l(i,j)}) = \max\{0, z^{l(i,j)}\} \tag{14}$$

where $z^{l(i,j)}$ is the output array of the BN and $a^{l(i,j)}$ is the activation of $z^{l(i,j)}$.

A pooling layer is connected behind the convolutional layer to reduce the dimension of the feature map and to keep the invariance of the characteristic scale. The pooling technique typically includes maximum pooling, average pooling and stochastic pooling. The most widely used type is the maximum pooling which is presented as follows:

$$p^{l(i,j)} = \max_{(j-1)W+1\le q\le jW}{}^{\{a^{l(i,q)}\}} \tag{15}$$

where $a^{l(i,q)}$ is the value of the q^{th} neuron in the i^{th} framework of layer l, W is the width of pooling size, $p^{l(i,j)}$ is the corresponding value of the neuron in pooling layer l, and q falls within the range between $(j-1)W+1$ and jW.

The classification task of a fault diagnosis is a multiclass classification problem. The representations learned from the deep learning layers are fed into the last fully connected layer. In the output layer, a softmax function is adopted to output the conditional probability for every class. If there are η classes with the bearing conditions of the input sample χ, their output probability O_i for class i is calculated as follows.

$$O_i = e^{(\theta(i)\chi)} \Big/ \sum_{i=1}^{\eta} e^{(\theta(i)\chi)} \text{, } i = 1, 2, 3, \ldots, \mu \tag{16}$$

where θ represents the parameters of the model learned and $\sum_{i=1}^{\mu} O_i = 100\%$. This will be updated automatically along with the training samples.

In addition, the (Stochastic Gradient Descent with Momentum) SGDM optimizer is used in the model where its stabilization and reliability, and the momentum value of the SGDM is 0.95. The size of the mini-batch is 500.

Fig. 3 illustrates the concrete architecture of VMD-DCNNs model developed in this study. The input layer has the raw vibration signals composed of the mode components with a size of $N \times k \times 1$ obtained using the VMD method, where N denotes the length of the raw vibration signals, and k denotes the mode number.

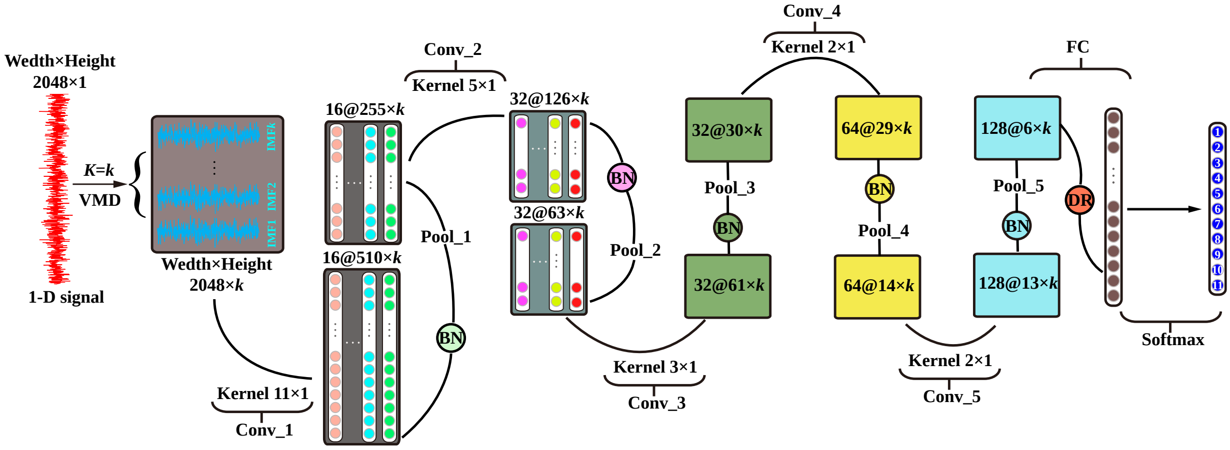


Fig. 3 The architecture VMD-DCNNs

3.2. Fault Diagnosis Based on VMD-DCNNs

The VMD-DNNs model is developed by combining the advantage of the VMD algorithm in processing time-frequency signals and the superiority of CNNs in machine learning technology. In this section, the superior ability of the fault diagnosis of wind turbine bearings is achieved by integrating

VMD with CNNs classification capability. The proposed VMD-DCNNs method consists of two parts. In the first part, VMD is used to decompose a 1-D raw vibration signals into K sub-signals. In the second part, the $N \times k \times 1$ vibration signal array is imported into the DCNNs model to address the fault diagnosis of the rolling bearings. In the proposed method, the fault diagnosis is performed directly based on the raw vibration signals in an end-to-end way without any manual feature extraction and artificial selection. The advantage of this method is reflected by considering the fact that raw vibration signals contain more information than images. The flowchart of a fault diagnosis system based on the VMD-DCNNs model is presented in Fig. 4 and the detailed process of fault diagnosis is given as follows:

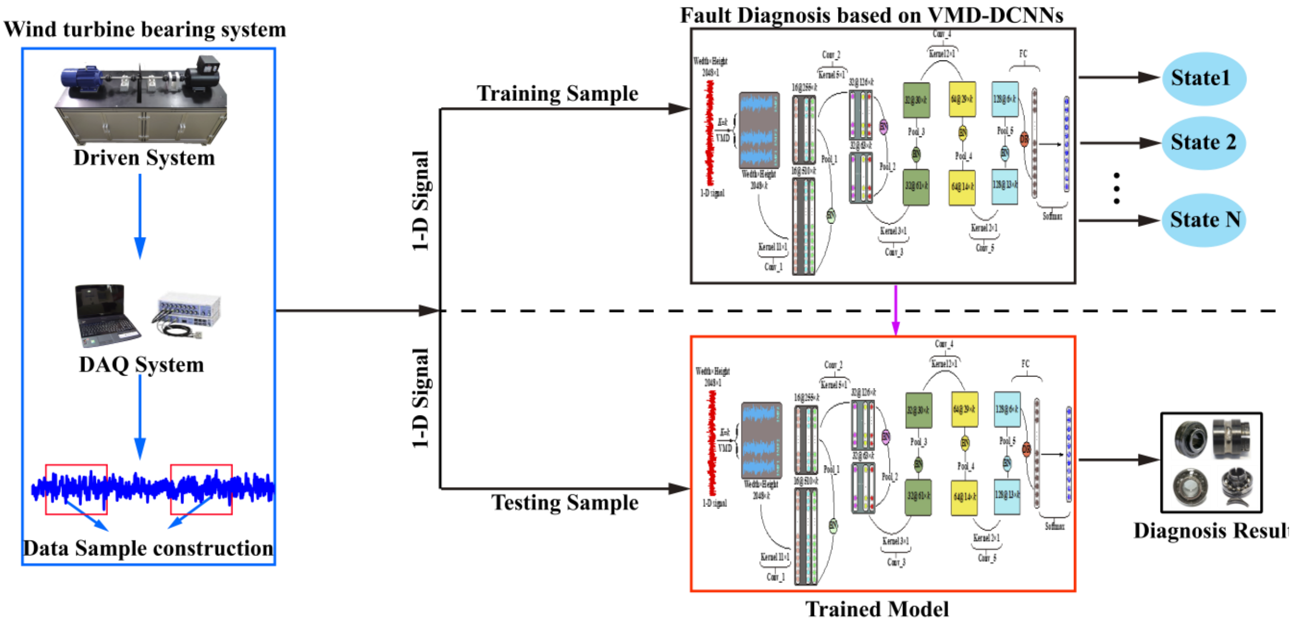


Fig. 4 The fault diagnosis system based on the VMD-DCNNs

Step 1: Vibration data of rolling bearings under different conditions are collected by the Data Acquisition (DAQ) systems.

Step 2: The collected 1-D signals data is converted into K modes using the VMD algorithm.

Step 3: The decomposed mode components are split into two parts with a proportion. One part is used to train the model of VMD-DCNNs, called the training samples; the other part is used to test model, as called the testing samples.

Step 4: The training samples are used to train DCNNs models

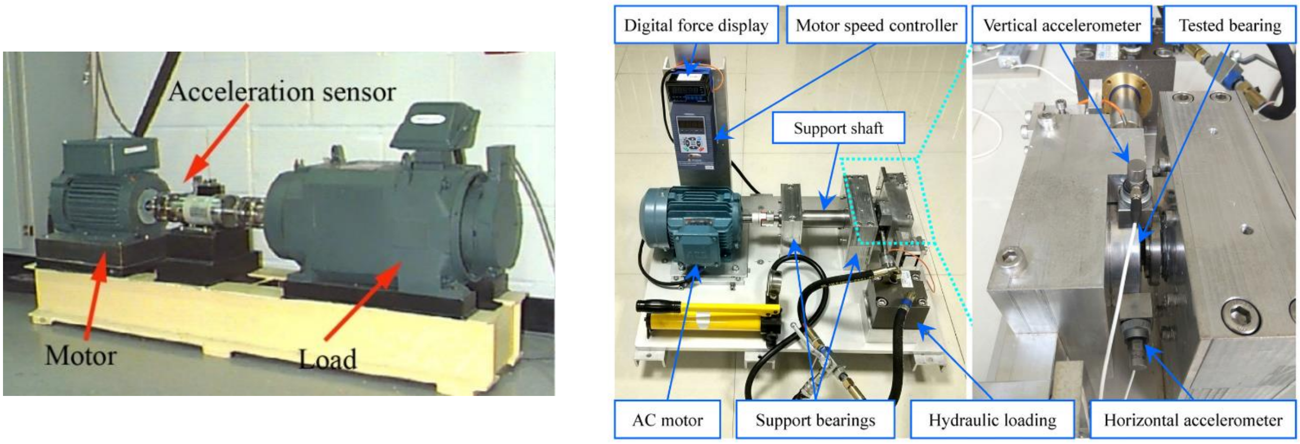
Step 5: The testing samples are imported into the model of the trained DCNNs to realize the fault diagnosis of bearings.

Moreover, the parameter α used in the VMD algorithm in this paper is chosen as the sampling frequency to ensure the accuracy of the VMD in the feature extraction of the rolling bearings.

4. Experimental Validation

4.1 Data Description

In this section, the rolling bearings data is used to from experiments that are available in the public domain to validate the proposed VMD-DCNNs. The datasets of the experiments conducted in Case Western Reserve University [26](CWRU) and Xi'an Jiao Tong University [27](XJTU) are used to demonstrate the superiority of the VMD-DCNNs model in detecting faults based on the data obtained under noisy and various loads conditions. This is in addition to highlighting the effectiveness of the VMD-DCNNs model in the extraction of initial multiple failures and associated features. The experimental platforms are shown in Fig. 5.



(a) The bearing dataset of CWRU (b) The bearing dataset of XJTU

Fig. 5 The experiment platform

The data from CWRU was produced by the experimental platform presented in Fig. 5(a). The faults of the rolling bearing were generated by an electro-discharge machining (EDM). In total, eleven sets of data covering the normal state, inner race fault, ball fault and outer race fault in different azimuths (3, 6 and 12o'clock directions) are used in this study. The motor loads range from 0 HP to 3 HP and the tested bearing model is SKF 6205. The details of the SKF 6205 bearing are shown in Table 1.

Table: 1 Bearing parameters of 6205 SKF (Size: inches)

Inside Diameter	Outside Diameter	Ball Diameter	Pitch Diameter	Thickness
0.9843	2.0472	0.3126	1.537	0.5906

Fig. 5(b) presents the experimental platform used in XJTU to produce vibration signals under the inner race fault, cage fault outer race fault, and hybrid faults conditions. Hybrid faults include the inner race, ball, cage and outer race failures. The experimental data is selected to validate the developed VMD-DCNNs model in diagnosing weak faults of rolling bearings. The type of tested bearings used in this study is the LDK UER 204, and the details of the experiments are given in Table 2.

Table: 2 Parameters of the XJTU tested bearings

Parameter	Value	Parameter	Value
Outer race diameter	39.90 mm	Inner race diameter	29.30 mm
Bearing mean diameter	34.55 mm	Ball diameter	7.92 mm
Number of balls	8	Contact angle	0°
Load rating (static)	6.65 kN	Load rating (dynamic)	12.82 kN

Fig. 6 shows the photographs of the bearings failure.

(a) Normal (b) Inner race (c) Cage (d) Ball (e) Outer race

Fig. 6 Photographs of the XJTU bearings

Table 3 and Table 4, respectively, present the details of the datasets of CWRU and XJTU used for the training and testing.

Table: 3 Details of CWRU rolling bearing datasets

Fault	Load	Normal	Inner Race	Ball	Outer Race@3	Outer Race@6	Outer Race@12
Diameter		-	0.007/**0.021**	0.007/**0.021**	0.007/**0.021**	0.007/**0.021**	0.007/**0.021**

Labels		0	1/**6**	2/**7**	3/**8**	4/**9**	5/**10**
Train	0	3200	3200	3200	3200	3200	3200
Test		400	400	400	400	400	400
Train	1	3200	3200	3200	3200	3200	3200
Test		400	400	400	400	400	400
Train	2	3200	3200	3200	3200	3200	3200
Test		400	400	400	400	400	400
Train	3	3200	3200	3200	3200	3200	3200
Test		400	400	400	400	400	400

Table: 4 Details of XJTU rolling bearing datasets

Type	Inner Race	Cage	Outer Race	Hybrids (Inner race and Outer Race)
Labels	0	1	2	3
Train	3200	3200	3200	3200
Test	400	400	400	400

4.2 Time complexity of the VMD-DCNNs

The time-consuming model training is significantly affected by the time complexity of the fault diagnosis model. This is important in a practical application because a higher time complexity requires a longer time for training the model. Therefore, the time complexities of the VMD-DCNNs model, depending on the parameter K and other CNNs-based methods are shown in Table 5.

Table: 5 The time complexity
of the VMD-DCNNs model and other CNNs-based methods

K	Time Complexity in Training (The length of input signal = 2048)
K=1	$O\ (6.7\times10^4)$
K=2	$O\ (13.5\times10^4)$
K=3	$O\ (20.3\times10^4)$
K=4	$O\ (27.1\times10^4)$
K=5	$O\ (33.8\times10^4)$
K=6	$O\ (40.6\times10^4)$
K=7	$O\ (47.4\times10^4)$
K=8	$O\ (54.1\times10^4)$
K=k	$O\ (k\times6.7\times10^4)$
WT-CNNs	$O\ (37.8\times10^4)$
MSCNNs	$O\ (178.8\times10^4)$

As shown in the Table 5, the time complexity O of the VMD-DCNNs increases with K. The VMD-DCNNs model with a larger K has a higher accuracy in some cases, but it will increase the time complexity of the model. For this reason, a rational K is selected by considering the performance of the model and its time complexity. In addition, the time complexities of the VMD-DCNNs model with $K = 3$ are less than those of the WT-CNNs [28] and MSCNNs [29].

The IMFs 1-8 obtained using the VMD method are shown in Fig. 7 for a better understanding of the signals imported into the input convolutional layer.

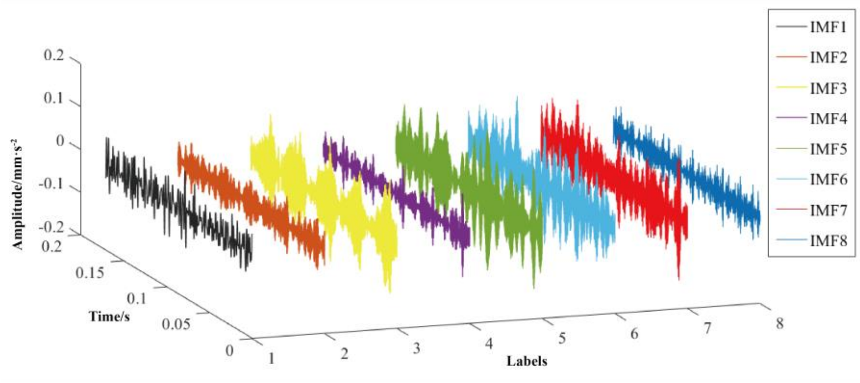


Fig. 7 The input of signals when K=8

As shown in Fig. 7, a 1-D signal is decomposed to 8 IMFs components when K is equal to 8. Therefore, data enhancement is successfully achieved by using the IMFs as the input of the VMD-DCNNs model. In addition, it is difficult to confirm the residual component without using human experiences. Hence, it should be easier to select the optimal parameter K by visualizing the features of the waveforms of each IMF in the convolutional layer.

4.3 Experimental Setup

Based on the architecture of the developed VMD-DCNNs model that is introduced in detail in Section 3.1, the parameters of the VMD-DCNNs model with “K=3”are presented in Table 6.

Table: 6 The details of VMD-DCNNs with the $K = 3$

No.	Layers	Kernel Size/Stride	Filter numbers	Outputs Size
1	Input Layer	-	-	[2048,3]
2	Conv_1	[11,1]/[4,1]	16	[510,3]
3	Pool_1	[2,1]/[2,1]	16	[255,3]
4	Conv_2	[5,1]/[2,1]	16	[126,3]
5	Pool_2	[2,1]/[2,1]	16	[63,3]
6	Conv_3	[3,1]/[1,1]	32	[61,3]
7	Pool_3	[2,1]/[2,1]	64	[30,3]

8	Conv_4	[2,1]/[1,1]	64	[29,3]
9	Pool_4	[2,1]/[2,1]	64	[14,3]
10	Conv_5	[2,1]/[1,1]	128	[13,3]
11	Pool_5	[2,1]/[2,1]	128	[6,3]
12	FC	-	-	[11,2304]
13	Softmax	-	-	-

4.4 Performance under different Noise

In this section, the diagnotic accuracy of the developed VMD-DCNNs is examined and discussed. As described in section 4.1, the test data produced in CWRU and XJTU is noiseless. Noises of different levels are added into the original experimental data by using the *awgn* function in MATLAB [30] to test the performance of the VMD-DCNNs model for applications in the real world condition.

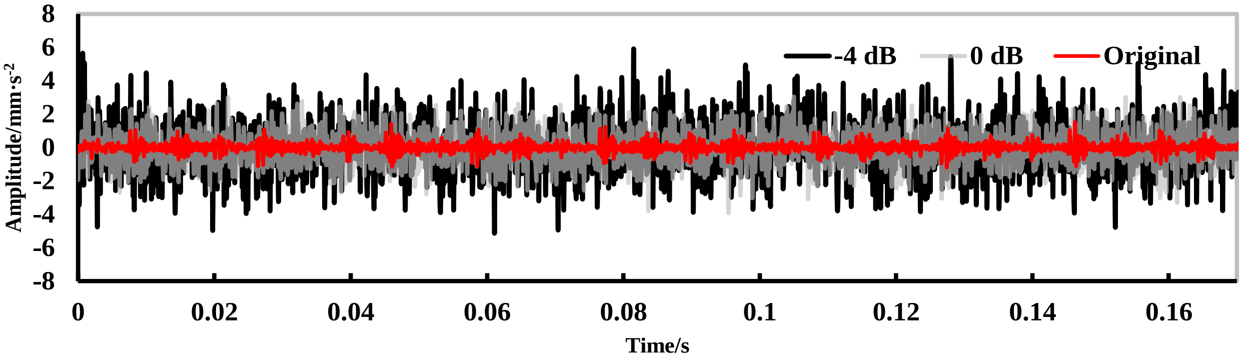
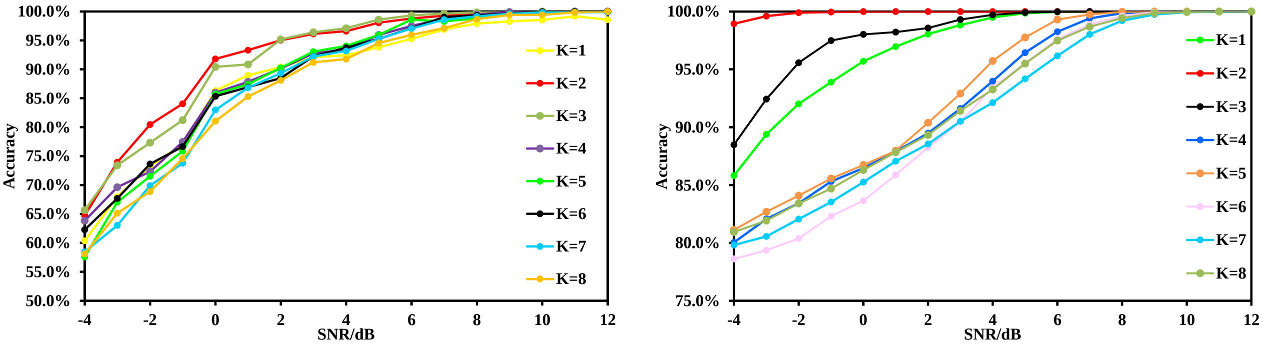


Fig. 8 The original signal of inner race fault with white Gaussian noise which SNR is 0dB

As shown in Fig. 8, there are significant differences among the original signal (red line), a 0dB signal with Gaussian noise (gray line) and a -4 dB signal with Gaussian noise (black line). Noise is a common problem in the actual operation of a wind turbine. This is because it changes the data distribution and conceals its fault features, which would affect the diagnostic accuracy of a model. The signals were examined with a noise level from -4dB to 12dB using the proposed VMD-DCNNs model. The performance of the VMD-DCNNs model for the *K* with different values in the VMD process is shown in Fig.9.



(a) CWRU (b) XJTU

Fig. 9 The accuracy of VMD-DCNNs under different noisy environments

The VMD-DCNNs model with *K* ranging from 1 to 8 is capable of identifying the fault for signals with a high SNR accurately (Fig. 9). Thus, the selection of the optimal value of K depends on the performance of the VMD-DCNNs model for low SNR signals. Table 7 and Table. 8 present the accuracy of the VMD-DCNNs model with different *K* values in identifying the faults based on the data from CWRU and XJTU, respectively, for different levels of noise.

Table: 7 The accuracy of VMD-DCNNs with different batch size and SNR in details

SNR(dB) Parameter k	-4	0	4	8	12
1	60.36±1.89	86.32±1.67	92.41±0.69	97.91±0.07	98.59 ± 0.05
2	64.73±2.06	91.79±1.01	96.59±0.98	99.73±0.03	99.89 ± 0.02
3	65.68±1.39	90.42±1.39	97.09±0.52	99.86±0.03	99.98 ± 0.01
4	63.86±1.25	85.97±0.92	93.45±0.72	99.45±0.74	99.52 ± 0.12
5	57.59±2.07	85.75±1.19	94.05±0.67	98.82±0.42	99.88 ± 0.18
6	62.27±1.09	87.35±0.91	93.73±0.72	99.32±0.37	99.44 ± 0.31
7	58.45±1.35	82.98±1.10	93.18±0.90	99.14±0.44	99.21 ± 0.36
8	58.09±1.77	81.05±1.24	91.77±0.65	98.68±0.67	99.72 ± 0.25

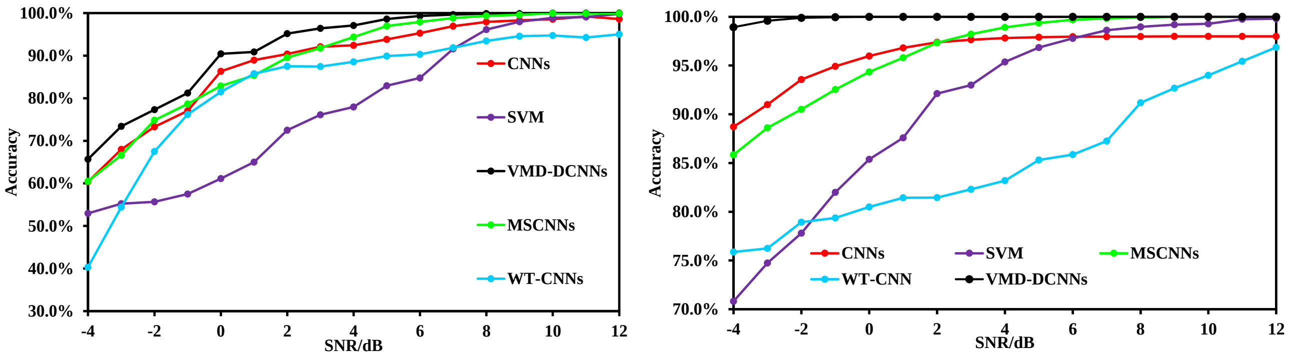
Table: 8 The accuracy of VMD-DCNNs with different batch size and SNR in details

SNR(dB) Parameter k	-4	0	4	8	12
1	85.82 ± 0.94	95.70 ± 0.63	99.48 ± 0.21	99.99 ± 0.01	99.99 ± 0.01
2	98.93 ± 0.19	99.99 ± 0.01	99.99 ± 0.01	99.99 ± 0.01	99.99 ± 0.01
3	88.48 ± 0.74	98.02 ± 0.40	99.71 ± 0.13	99.99 ± 0.01	99.99 ± 0.01
4	80.04 ± 0.87	86.47 ± 0.91	93.98 ± 0.48	99.85 ± 0.09	99.99 ± 0.01
5	81.13 ± 0.96	86.73 ± 0.75	95.73 ± 0.63	99.97 ± 0.02	99.99 ± 0.01
6	78.62 ± 1.09	83.65 ± 0.87	93.42 ± 0.42	99.62 ± 0.04	99.99 ± 0.01
7	79.82 ± 1.07	82.25 ± 0.68	92.12 ± 0.47	99.20 ± 0.21	99.99 ± 0.01

8	80.94 ± 0.81	86.30 ± 0.89	93.25 ± 0.56	99.43 ± 0.19	99.99 ± 0.01

As presented in Table 7 and Table 8, the VMD-DCNNs model has different accuracies in classifying the fault based on the signals from CWRU and XJTU for a same noise level and *K* value. This is because the XJTU's data only contains four categories of faults, while the data from CWRU covers 11 kinds of faults. Consequently, the classification accuracy of the VMD-DCNNs model is relatively higher when examining the data with fewer categories of faults. It also indicates that the optimal value of *K* is diffident for these two datasets from CWRU and XJTU. While the performance of the VMD-DCNNs model in examining a high SNR signal is reasonably good for all *K* values, the corresponding performance in examining a low SNR signal significantly depends on the value of *K*. For the CWRU data, the performance of the VMD-DCNNs model is less satisfactory for a large *K* value. The reason is that the input signals may be over decomposed when using a large *K*. In addition, the time complexity is increased with *K*. Therefore, an appropriate *K* value should be confirmed in evaluating the performance of the model under different noise signals. It is noted that the highest accuracy in dealing with the CWRU data having a SNR of -4dB is achieved when *K* is equal to 3. Similarly, the VMD-DCNNs model achieves the highest accuracy of 98.93% in dealing with the lowest SNR signal from XJTU when *K* is 2. In addition, the corresponding standard deviation of the accuracy is the lowest in the 20 trials. Therefore, the values of *K* are adopted as 2 and 3 when processing the XJTU's data and CWRU's data, respectively. In addition, the selections of this parameter is further proved rationally by the visualized results presented in Section 4.6.

Furthermore, in highlighting the superiority of the proposed VMD-DCNNs, Fig.10 compares the fault diagnosis of VMD-DCNNs, traditional CNNs, SVM [15], WT-CNNs [28], and MSCNNs [29] for signals with different SNRs.



(a) CWRU (b) XJTU

Fig. 10 Methods compared results under different noisy environment

As shown in Fig. 10, the proposed VMD-DCNNs method has the best performance in variable background noise level compared to other four methods. The shortcomings of the other methods are reflected under a strong noise environment although they perform well with a low noise level. When identifying different fault types and magnitudes of bearings, the VMD-DCNNs has an accuracy of 65.68% for the signal with a large noise (-4dB). This is higher than the results from SVM (52.95%), CNNs (60.36%), MSCNN (60.50%) and WT-CNNs (40.32%). When identifying fault types including hybrids failures, the VMD-DCNNs has an accuracy of 98.93% for the signal with a large noise (-4dB). Again, this is higher than for SVM (70.81%), CNNs (88.73%), MSCNN (85.86%) and WT-CNNs (75.88%).It is noted that the CNN-based methods have a higher accuracy compared to the SVM method; and this is because the generalization ability of the SVM method, which belongs to the shallow learning group, is insufficient. The accuracy of the WT-CNNs method is the lowest within the CNN-based methods because it uses the WT images as the input signal for training. Consequently, its capability in feature self-learning is relatively poor.

4.5 Performance across different conditions

The extrapolation performance of the model is validated using the XJTU data and CWRU data. The superiority of the model is further proved by different experiments.

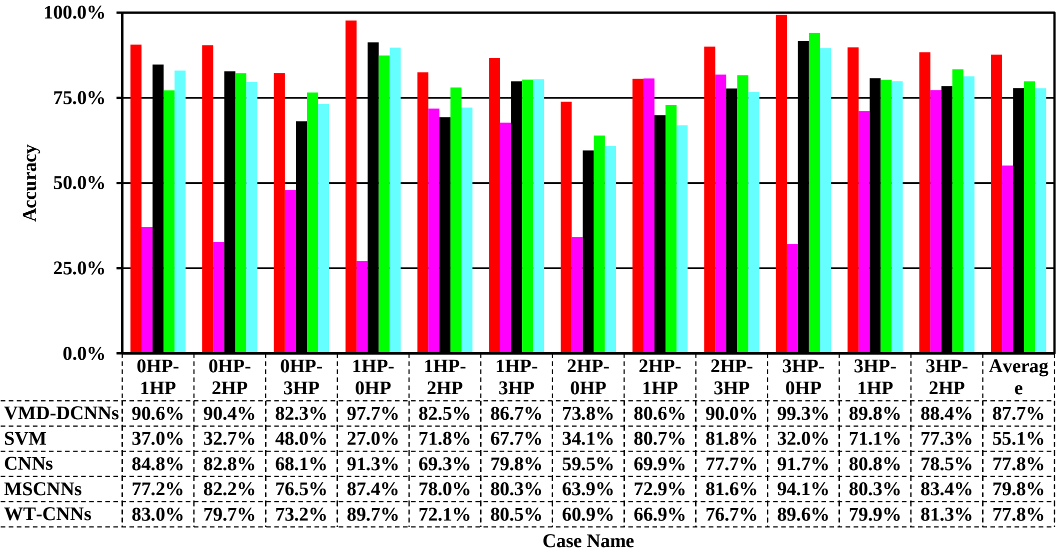
4.5.1 Different load domains

In this section, the adaptive capacity with different loads of VMD-DCNNs is tested by using the CWRU data. The description of the scene setup for the load domains adaptation is shown in Table 9.

Table: 9 The details of the load configuration

Case Name	Training samples	Testing samples
0 HP -1 HP	Load with 0 HP	Load with 1 HP
0 HP -2 HP	Load with 0 HP	Load with 2 HP
0 HP -3 HP	Load with 0 HP	Load with 3 HP
1 HP -0 HP	Load with 1 HP	Load with 0 HP
1 HP -2 HP	Load with 1 HP	Load with 2 HP
1 HP -3 HP	Load with 1 HP	Load with 3 HP
2 HP -0 HP	Load with 2 HP	Load with 0 HP
2 HP -1 HP	Load with 2 HP	Load with 1 HP
2 HP -3 HP	Load with 2 HP	Load with 3 HP
3 HP -0 HP	Load with 3 HP	Load with 0 HP
3 HP -1 HP	Load with 3 HP	Load with 1 HP
3 HP -2 HP	Load with 3 HP	Load with 2 HP

The result of VMD-DCNNs is compared to those of the SVM, CNNs, MSCNNs and WT-CNNs as shown in Fig. 11.



	0HP-1HP	0HP-2HP	0HP-3HP	1HP-0HP	1HP-2HP	1HP-3HP	2HP-0HP	2HP-1HP	2HP-3HP	3HP-0HP	3HP-1HP	3HP-2HP	Average
🟥 **VMD-DCNNs**	**90.6%**	**90.4%**	**82.3%**	**97.7%**	**82.5%**	**86.7%**	**73.8%**	**80.6%**	**90.0%**	**99.3%**	**89.8%**	**88.4%**	**87.7%**
🟪 **SVM**	**37.0%**	**32.7%**	**48.0%**	**27.0%**	**71.8%**	**67.7%**	**34.1%**	**80.7%**	**81.8%**	**32.0%**	**71.1%**	**77.3%**	**55.1%**
■ **CNNs**	**84.8%**	**82.8%**	**68.1%**	**91.3%**	**69.3%**	**79.8%**	**59.5%**	**69.9%**	**77.7%**	**91.7%**	**80.8%**	**78.5%**	**77.8%**
🟩 **MSCNNs**	**77.2%**	**82.2%**	**76.5%**	**87.4%**	**78.0%**	**80.3%**	**63.9%**	**72.9%**	**81.6%**	**94.1%**	**80.3%**	**83.4%**	**79.8%**
🟦 **WT-CNNs**	**83.0%**	**79.7%**	**73.2%**	**89.7%**	**72.1%**	**80.5%**	**60.9%**	**66.9%**	**76.7%**	**89.6%**	**79.9%**	**81.3%**	**77.8%**

Fig. 11 Methods compared results under different variable load environment

The VMD-DCNNs model performs better than the SVM, CNNs, MSCNNs and WT-CNNs models (Fig. 11), whose overall average accuracies in all scenes are 55.1%, 77.8%, 79.8% and 77.8%, respectively. Meanwhile the VMD-DCNNs model has an overall average accuracy of 87.7%. The better performance of the domain adaptation means that the features learned by the VMD-DCNNs model from the raw signals are more domain invariant compared to other methods. As observed from Fig.10 and Fig.11, it is worth noting that the SVM model has a better robustness than WT-CNNs in the noise environments, but it still performs poorly in variable load environments. Generally, it is uncommon for a model to have a good ability in the domain adaptation and a high noise robustness. However, the proposed VMD-DCNNs model has good performances in both noise environments and variable load domains.

4.5.2 Different errors of the sensor reset conditions

A supplementary experiment is set to further prove the extrapolation performance of the proposed VMD-DCNNs model. A sensor reset error experiment is examined using the CWRU experimental data. The extrapolation performance of the VMD-DCNNs model is tested by the simulated sensor reset error experiments. In the simulated experiments, the acquisition data has the same manifold but its mean value is different due to the sensor reset error [31]. Taking the Inner Race (IN007) fault of bearing as example, the signal of "Original_IN007" taken as $\boldsymbol{X}_{1,n} = (x_1, x_2, x_3 \ldots, x_n)$, the deviation with Zero-Position is $\boldsymbol{D}_{1,n} = \boldsymbol{Normal}_{1,n} \cdot ([\max(\boldsymbol{X}) - \min(\boldsymbol{X})])$, and the $n \times 1$ vector $\boldsymbol{Normal}_{1,n}$ is a normal distribution by $\mathrm{N}(\mu, \sigma^2)$, where $\mu \subset [0.1, 0.5]$. The error in the same experiment is stable, thus σ is equal to 0.001. The signal of a sensor reset error is $\boldsymbol{Y}_{1,n} = \boldsymbol{X}_{1,n} + \boldsymbol{D}_{1,n}$. For the sensor reset error scenario, the VMD-DCNNs model is trained using the data with a sensor reset error of 0%. Then, the data with sensor reset errors from 10% to 50% is tested. The time series of the signals captured under a sensor reset error condition is presented in Fig. 12.

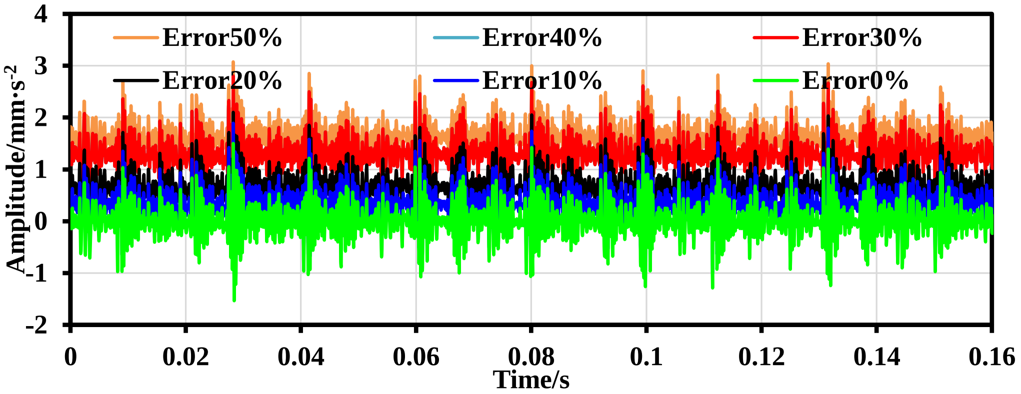


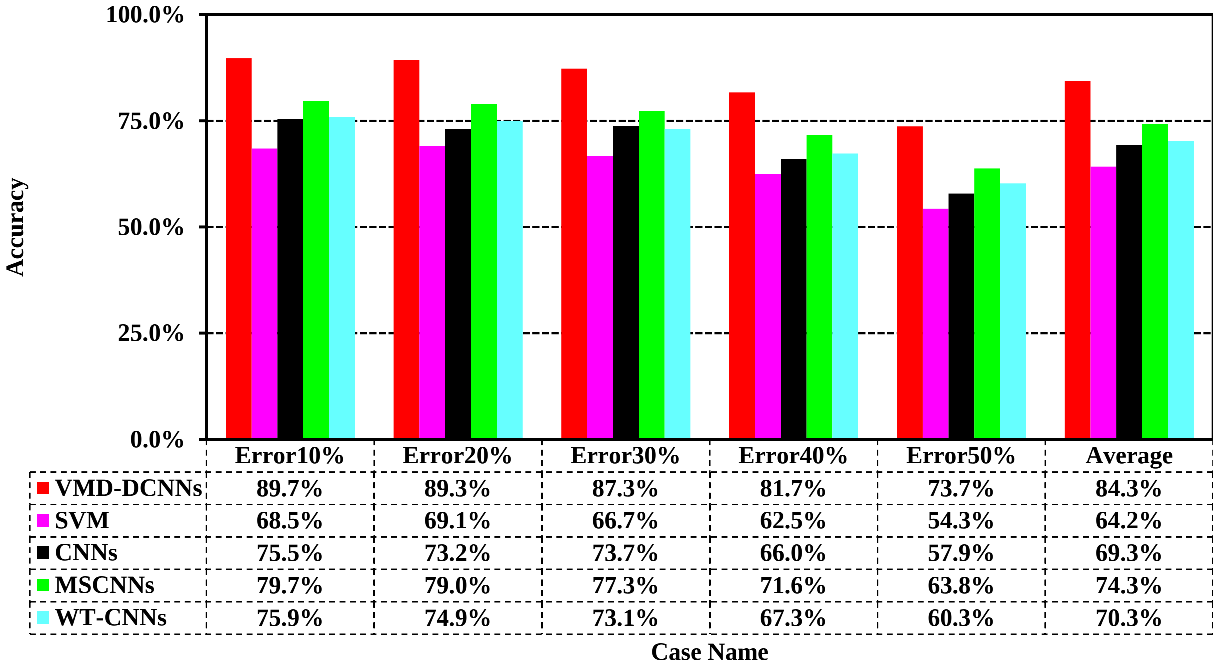
Fig. 12 The time series of the IN007 with different possition error

In order to validate the extrapolation performance of the VMD-DCNNs model, the above sensor reset error simulations are tested based on the data from CWRU that covers eleven fault states. Table 10 presents a detailed description of the scene setup for sensor position errors.

Table: 10 Details of the sensor position error configuration

Case Name	Training samples	Testing samples
Error 10%	Original Data	10% Reset error
Error 20%	Original Data	20% Reset error
Error 30%	Original Data	30% Reset error
Error 40%	Original Data	40% Reset error
Error 50%	Original Data	50% Reset error

The results of the experiments are shown in Fig. 13.



	Error10%	Error20%	Error30%	Error40%	Error50%	Average
🟥 VMD-DCNNs	89.7%	89.3%	87.3%	81.7%	73.7%	84.3%
🟪 SVM	68.5%	69.1%	66.7%	62.5%	54.3%	64.2%
⬛ CNNs	75.5%	73.2%	73.7%	66.0%	57.9%	69.3%
🟩 MSCNNs	79.7%	79.0%	77.3%	71.6%	63.8%	74.3%
🟦 WT-CNNs	75.9%	74.9%	73.1%	67.3%	60.3%	70.3%

Fig. 13 Comparison of accuracy for the data with different position errors

As shown in Fig. 13, the SVM, CNNs, MSCNNs and WT-CNNs models perform worse than the VMD-DCNNs, with overall average accuracies of 64.2%, 69.3%, 74.3% and 70.3%, respectively, while the proposed VMD-DCNNs model has an accuracy of over 81% for most scenarios with a sensor reset error. The VMD-DCNNs model achieves an average accuracy of 84.3%. This indicates that the VMD-DCNNs model can effectively diagnose the fault for signals with a sensor reset error, although the data used in the training process is not collected under a sensor reset error. This implies that the developed VMD-DCNNs model has a good extrapolation performance.

4.5.3 Different failure occurrence conditions

In this section, the extrapolation performance of the model is further tested using the data of XJTU. It is worth noting that the data distribution of the failure is different in the four phrases. The signals from different fault phases are presented in Fig. 14. Phase_1 denotes the weakest fault and Phase_4 represents a total failure.

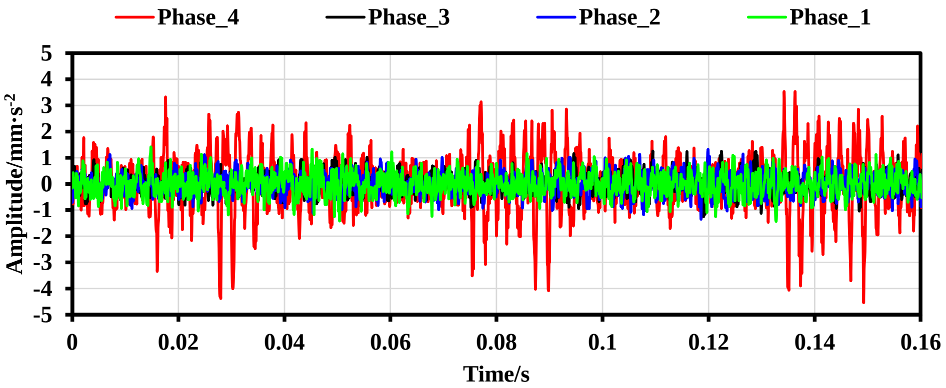


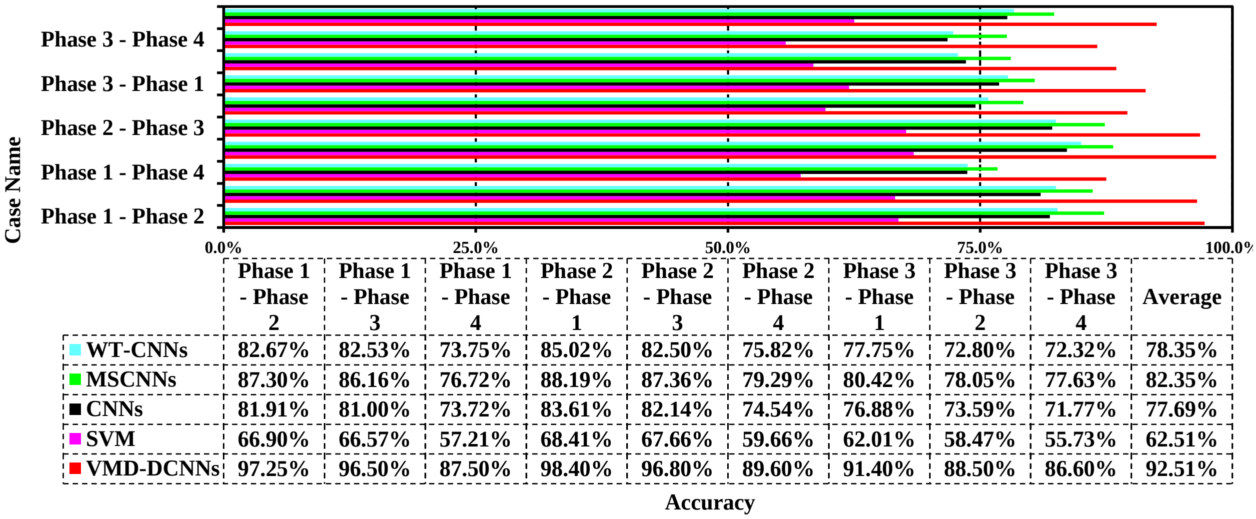
Fig. 14 The time series of different fault phases

The data of one phase is used for training while the data of the remaining three phases is used for tests. In total, nine cases were examined as presented in Table 11. The diagnosis results are shown in Fig. 15.

Table: 11 the details of the bearing fault development configuration

Case Name	Training samples	Testing samples
Phase 1 - Phase 2	Phase 1	Phase 2

Phase 1 - Phase 3	Phase 1	Phase 3
Phase 1 - Phase 4	Phase 1	Phase 4
Phase 2 - Phase 1	Phase 2	Phase 1
Phase 2 - Phase 3	Phase 2	Phase 3
Phase 2 - Phase 4	Phase 2	Phase 4
Phase 3 - Phase 1	Phase 3	Phase 1
Phase 3 - Phase 2	Phase 3	Phase 2
Phase 3 - Phase 4	Phase 3	Phase 4



	Phase 1 - Phase 2	Phase 1 - Phase 3	Phase 1 - Phase 4	Phase 2 - Phase 1	Phase 2 - Phase 3	Phase 2 - Phase 4	Phase 3 - Phase 1	Phase 3 - Phase 2	Phase 3 - Phase 4	Average
🟦 WT-CNNs	82.67%	82.53%	73.75%	85.02%	82.50%	75.82%	77.75%	72.80%	72.32%	78.35%
🟩 MSCNNs	87.30%	86.16%	76.72%	88.19%	87.36%	79.29%	80.42%	78.05%	77.63%	82.35%
■ CNNs	81.91%	81.00%	73.72%	83.61%	82.14%	74.54%	76.88%	73.59%	71.77%	77.69%
🟪 SVM	66.90%	66.57%	57.21%	68.41%	67.66%	59.66%	62.01%	58.47%	55.73%	62.51%
🟥 VMD-DCNNs	97.25%	96.50%	87.50%	98.40%	96.80%	89.60%	91.40%	88.50%	86.60%	92.51%

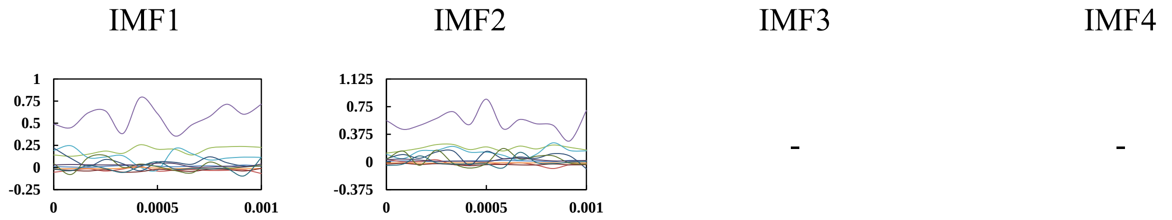
Fig. 15 Methods compared results tested by the data with fault development

Fig. 15 shows that the accuracy of the VMD-DCNNs model is higher than 86% for each examined case and the average diagnostic accuracy achieves 92.51% that is the highest among all the diagnostic models. The results indicate that the VMD-DCNNs model can effectively diagnose a fault in another developing phase by training the model based on the data from a distinct fault phase. This implies that the VMD-DCNNs model has a good extrapolation performance.

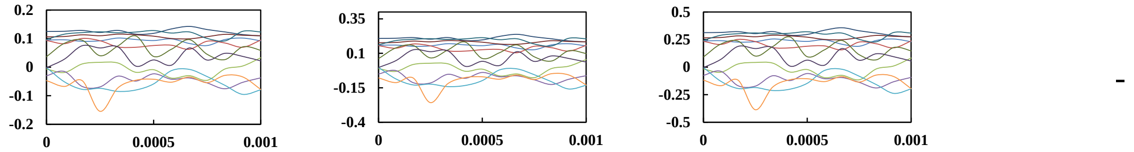
4.6 Visualization of VMD-DCNNs networks

In general, the CNNs are considered as a black box, thus it is difficult to understand their inner working mechanism. However, it is very much helpful to understand and then improve the fault diagnosis models by conducting research on the internal operating mechanism of CNNs. Hence, in this section, the internal operation of VMD-DCNNs is explored by visualizing the activations in the neural networks.

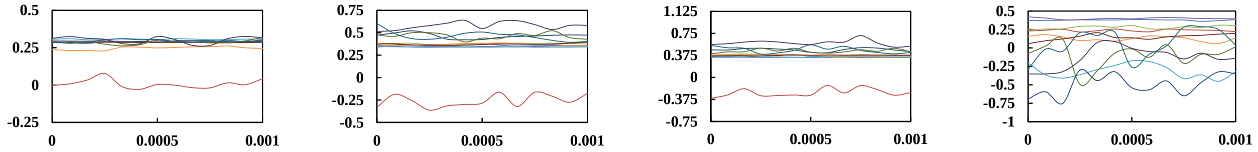
First, the visualization of each IMF in the last convolution layer (Conv5) is used to further prove the rationale in selecting the value of parameter *K* in the VMD-DCNNs model. The signals of CWRU and XJTU with different IMFs in the Conv5 are shown in Fig. 16 and Fig. 17, respectively.



（a） The Time domain representations of the Conv5 in the model of “*K*=2”



（b） The Time domain representations of the Conv5 in the model of “*K*=3”



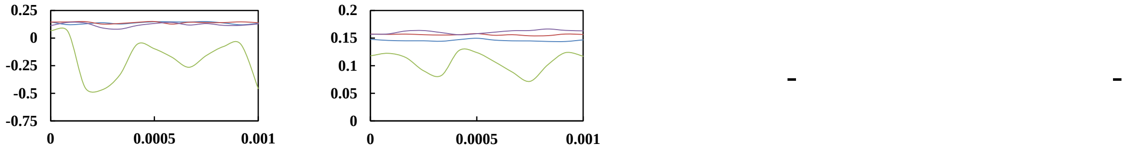
（c） The Time domain representations of the Conv5 in the model of “*K*=4”

NORMAL **IR007** **BAL007** **OR007@3**
OR007@6 **OR007@12** **IR021** **BAL021**
OR021@3 **OR021@6** **OR021@12**

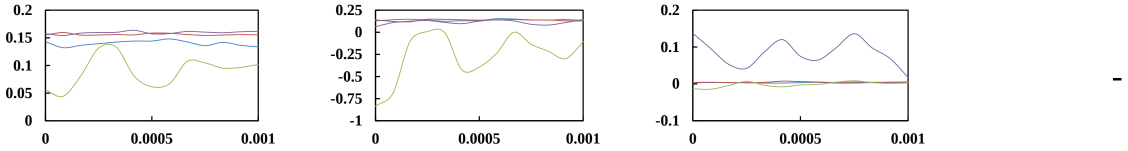
Fig. 16 The CWRU data visualization in the VMD-DCNNs model

As shown in Fig. 16, eleven states of bearings show the same waveform on different IMFs when *K* = 2 and *K* = 3. However, the IMF 4 shows a distinct waveform when *K* = 4. It is noted that the distributed feature representation in the fully connected layer is affected by the features learned in the Conv5. The probability of the labels corresponding to bearing states would be inaccurate when the features learned in the Conv5 are significantly different. Therefore, the IMF 4 is redundant and the *K* = 3 is the optimal parameter.

IMF 1 IMF 2 IMF 3 IMF 4



（a） The Time domain representations of the Conv5 in the model of “*K*=2”



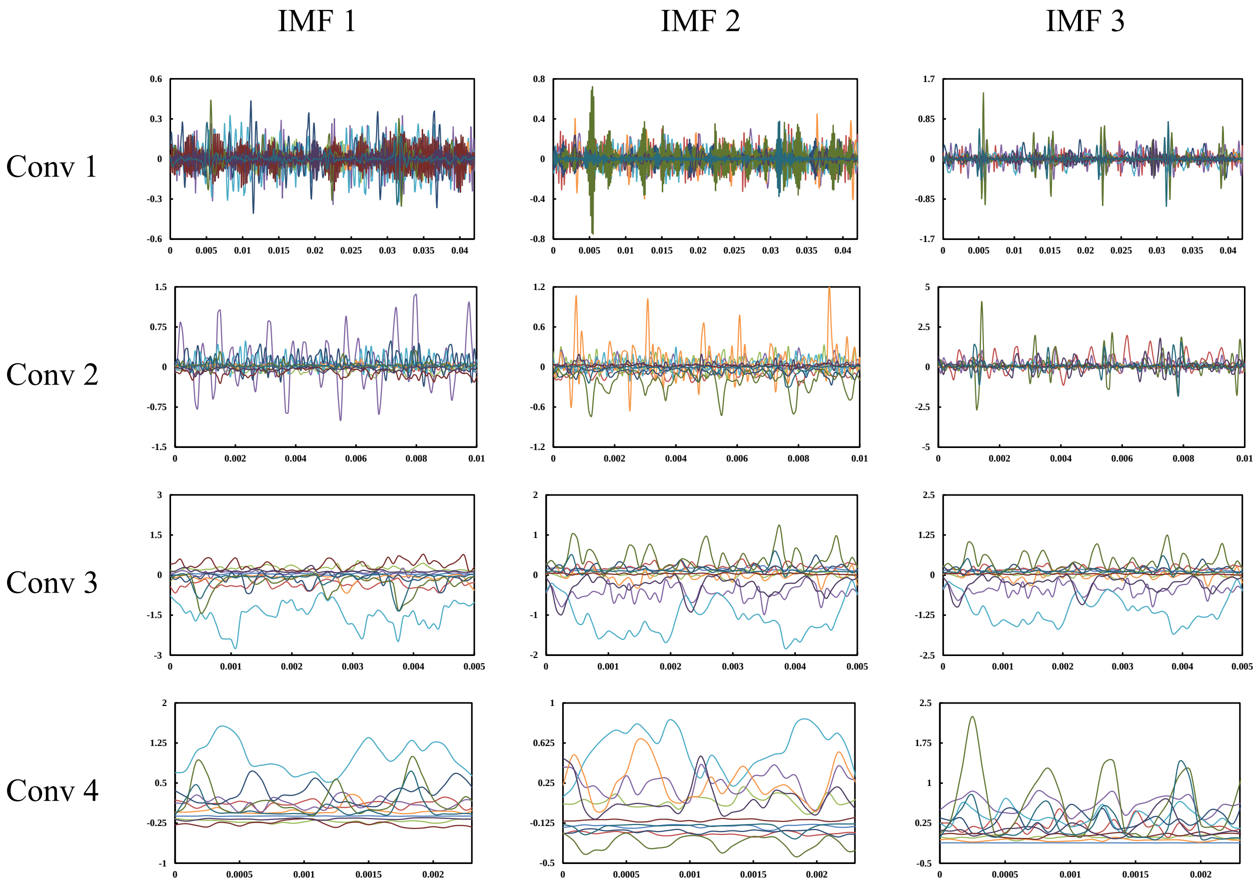
（b） The Time domain representations of the Conv5 in the model of “*K*=3”

—IR —Cage —OR —IR&OR

Fig. 17 The XJTU data visualization in the VMD-DCNNs model

As shown in Fig. 17, four states of the bearings show the same waveform on different IMFs when $K = 2$. However, the IMF 3 shows a distinct waveform when $K = 3$. This means that the IMF 3 is redundant and the $K = 2$ is the optimal parameter.

Secondly, the learned features of every IMF in the Conv1 to Conv5 are visualized in Fig. 18 in order to realize the kinds of features that every convolution layer has learned.



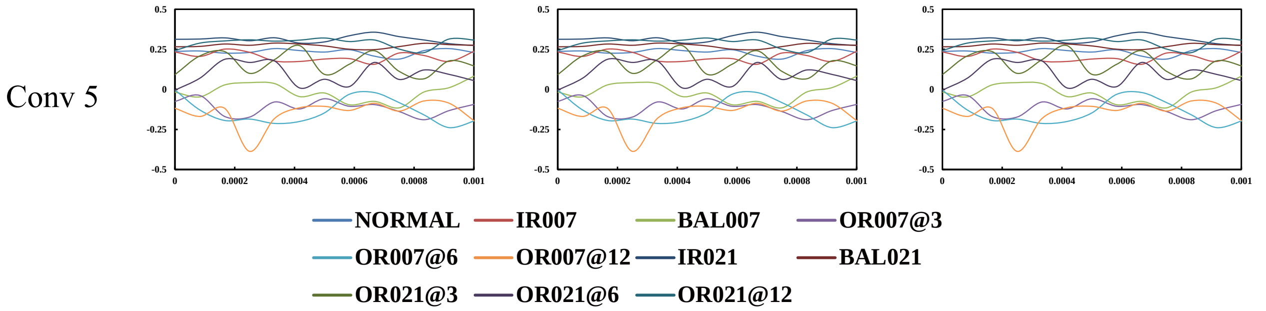


Fig. 18 Time domain representations of the each convolutional layers

The representations of time series indicate that the vibration characteristics in time domain become clearer than the original vibration signals in the Conv1 layer. As the convolutional layers go further, the vibration characteristics gradually disappear and the curve’s fluctuation gradually flattens out. However, as shown in Conv 5, the eleven states have the same waveform for different IMFs. This indicates that the VMD-DCNNs model has learnt the translation invariance characteristics of a one-dimensional signal from each IMF. As can be seen from Fig. 18, which is for Conv 1, the vibration waveforms in Conv 1 tends to be a real vibration signal. The plotted representation of Convolutional layer 1 (Conv1) in the frequency domain is shown in Fig. 19.

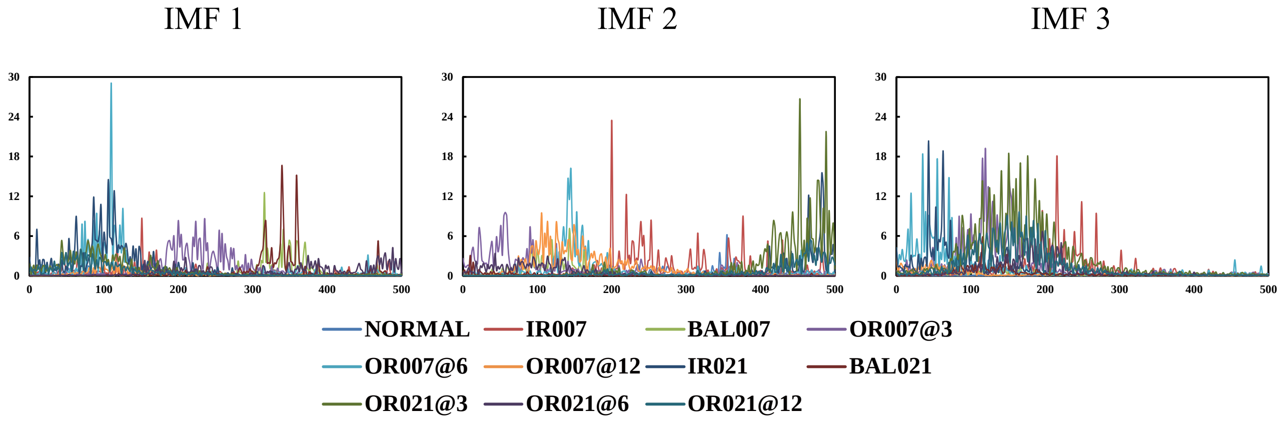


Fig. 19 Frequency domain representations of the first convolutional layer

Fig. 19 presents the information with high frequencies that have been filtered out by the Conv1. Some of the characteristic frequencies of the bearing vibrations have been retained, suggesting that the Conv1 layer acts as a high-frequency filter to make the network more robust in noise environments.

Lastly, the visualization of the feature distribution of test samples extracted from each layer by the t-SNE algorithm is shown in Fig. 20 in order to prove that the characteristics learned by different IMFs are complementary to each other intuitively.

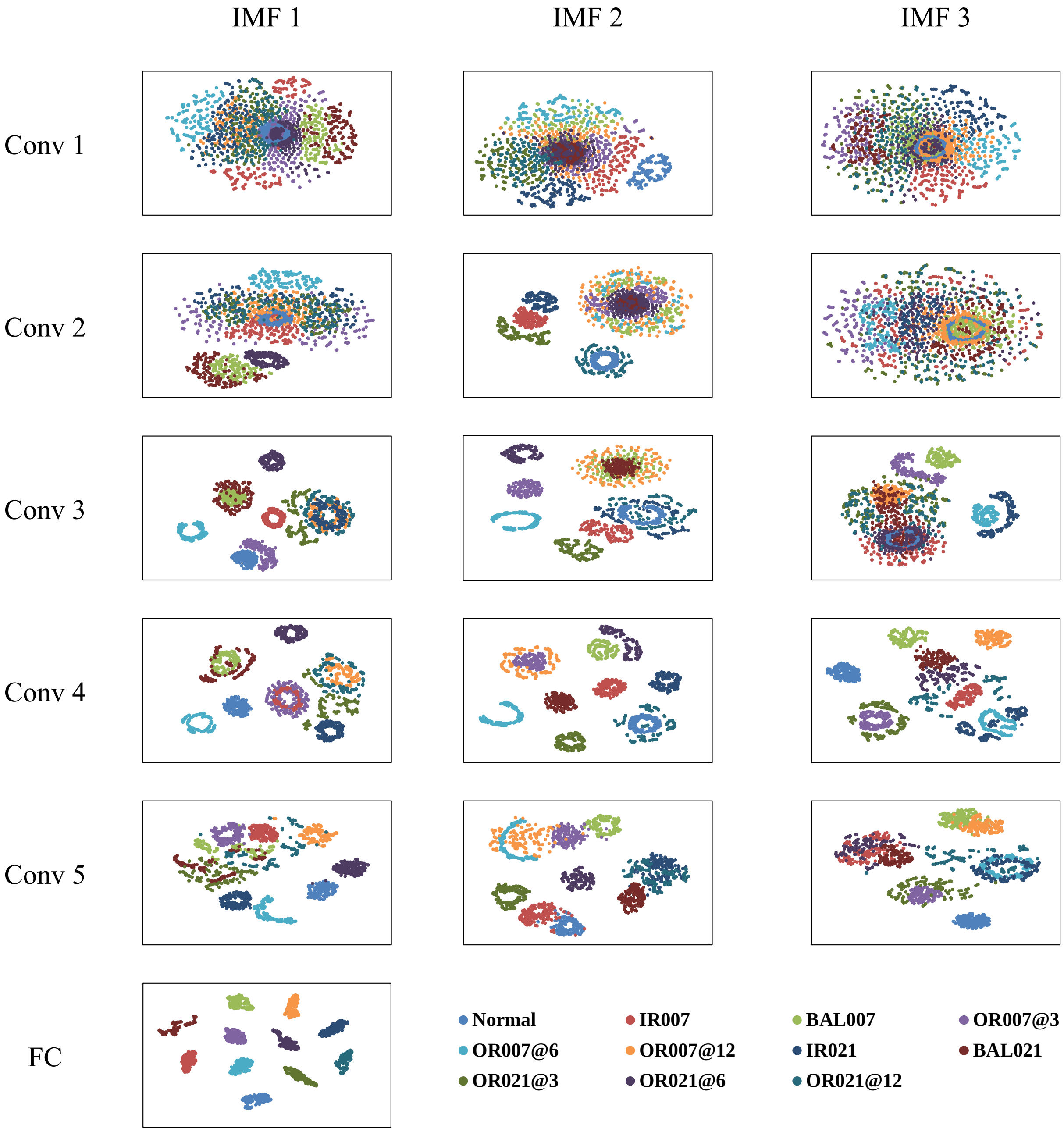


Fig. 20 Visualization of each convolutional layers

As shown in Fig. 20, the extracted features are becoming more and more separable as the layers go deeper. At the first convolutional layer, all IMF features are complicated. However, in the second convolutional layer, the IMF features have begun to be gradually separated. It is noted that the features of the IMF2 are more distinct than those of IMF1 and IMF3. As the layers go further, the features become markedly divisible, in the same convolutional layer, while the features in each IMF are different. For example, in the Conv 4 layer, OR007 cannot be obviously isolated in IMF1, but it is separated well in IMF3. Furthermore, IR007 is divisible in IMF2 and IMF3, but it fails to be isolated in IMF1. However, regardless of the degree of the separation of IMFs features in each Conv layer, the fault classifications become clearly divisible when the features are summarized in the fully connected layer. It can be concluded that, the characteristics learned from different IMFs are complementary to each other, which improves the diagnostic performance of the model.

5. Conclusions

In this research, a novel VMD-DCNNs model is developed to address fault diagnosis of rolling bearings. The VMD-DCNNs model works directly on raw vibration signals without any manual feature extraction process. By integrating CNNs with VMD, the method combines the flexibility and accuracy of the VMD algorithm in the decoupling vibration with the deep learning capability of CNNs. The diagnostic ability and extrapolation performance of the VMD-DCNNs model are validated using the experimental data from XJTU and CWRU. Compared to the traditional fault models including the CNNs, SVM, WT-CNNs, and MSCNNs, the VMD-DCNNs model developed in this study has the best performance in a variety of experimental tests including noise environments, variable load domains, sensor reset errors and bearing fault developments. In addition, the rationale behind selecting the value of parameter K for use in the VMD-DCNNs is explained by visualizing the signals in the last convolution layer. The features learned from the redundant IMFs are significantly different from those in the optimal IMFs. The features learned by VMD-DCNNs with the optimal K have similar waveforms indicating that the translation invariances of a 1-D vibration signal are learned by the VMD-DCNNs model.

Acknowledgements

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (grant numbers: 51676131, 51875361 and 51976131), Science and Technology Commission of Shanghai Municipality (grant number: 1906052200) and Royal Society (grant number: IEC\NSFC\170054).

References

[1] Global Wind Energy Council (GWEC). Global wind report-annual market update 2018. Brussels: Global Wind Energy Council, 2019.

[2] International Energy Agency (IEA). Offshore Wind Outlook 2019. Paris: International Energy Agency, 2019.

[3] Freiberg, Alice, et al. "Health effects of wind turbines in working environments – a scoping review." Scandinavian Journal of Work, Environment & Health 44.4 (2018): 351-369.

[4] Lin, Zhongwei, et al. "Coordinated pitch & torque control of large-scale wind turbine based on Pareto efficiency analysis." Energy (2018): 812-825.

[5] Li, Jimeng, Ming Li, and Jinfeng Zhang. "Rolling bearing fault diagnosis based on time-delayed feedback monostable stochastic resonance and adaptive minimum entropy deconvolution." Journal of Sound and Vibration (2017): 139-151.

[6] Sheng S. Wind Turbine Gearbox Reliability Database, Condition Monitoring, and O&M Research Update (Conference Paper) NREL/PR-5000e63868, National Renewable Energy Laboratory (NREL), Golden, CO (US), 2016.

[7] Glowacz, Adam, et al. "Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals." Measurement (2018): 1-9.

[8] Hu, Yongtao, et al. "A New Method of Wind Turbine Bearing Fault Diagnosis Based on Multi-Masking Empirical Mode Decomposition and Fuzzy C-Means Clustering." Chinese Journal of Mechanical Engineering 32.1 (2019): 1-12.

[9] Badihi, Hamed, Youmin Zhang, and Henry Hong. "Wind Turbine Fault Diagnosis and Fault-Tolerant Torque Load Control Against Actuator Faults." IEEE Transactions on Control Systems and Technology 23.4 (2015): 1351-1372.

[10] Lei Y, Li N, Lin J, et al. Fault Diagnosis of Rotating Machinery Based on an Adaptive Ensemble Empirical Mode Decomposition[J]. Sensors, 2013, 13(12): 16950-16964.

[11] Lei, Yaguo, et al. "Fault Diagnosis of Rotating Machinery Based on an Adaptive Ensemble Empirical Mode Decomposition." Sensors 13.12 (2013): 16950-16964.

[12] Yang, Boyuan, Ruonan Liu, and Xuefeng Chen. "Sparse Time-Frequency Representation for Incipient Fault Diagnosis of Wind Turbine Drive Train." IEEE Transactions on Instrumentation and Measurement 67.11 (2018): 2616-2627

[13] Qiu, Xueheng, et al. "Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting." Applied Soft Computing 54 (2017):246-255.

[14] Su, Zuqiang , et al. "Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier." Measurement 48(2014):136-148.

[15] Chen X J, Yang Y M, Cui Z X, et al. Vibration fault diagnosis of wind turbine based on variational mode decomposition and energy entropy[J]. Energy, 2019, 174: 1100-1109.

[16] Zheng, Xiaoxia, et al. "Variational mode decomposition applied to offshore wind turbine rolling bearing fault diagnosis." chinese control conference (2016): 6673-6677.

[17] Gu, Ran, et al. "Incipient fault diagnosis of rolling bearings based on adaptive variational mode

decomposition and Teager energy operator." Measurement (2020)

[18] Li Jiming, et al. “Multiscale local feature learning based on BP network for rolling bearing intelligent fault diagnosis”. Measurement 153 Mar.1 (2020): 107419.

[19] Haidong, Shao , et al. "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine." Knowledge Based Systems 140.jan.15(2018):1-14.

[20] Wen, Long , et al. "A New Convolutional Neural Network Based Data-Driven Fault Diagnosis Method." IEEE Transactions on Industrial Electronics PP.99(2017):1-1.

[21] Zhang, Wei , et al. "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load." Mechanical systems and signal processing 100.FEB.1(2018):439-453.

[22] Dragomiretskiy, Konstantin , and D. Zosso . "Variational Mode Decomposition." IEEE Transactions on Signal Processing 62.3(2014):531-544.

[23] Gilboa, Guy, Nir Sochen, and Yehoshua Y. Zeevi. "Variational denoising of partly textured images by spatially varying constraints." IEEE Transactions on Image Processing, 2006, 15(8): 2281-2289.

[24] Yong, Li , et al. "Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks." Measurement 130(2018):94-104.

[25] Li, Jimeng , et al. "Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis." Mechanical systems and signal processing 126.JUL.1(2019):568-589.

[26] http://csegroups.case.edu/bearingdatacenter/home/ .

[27] Biao Wang, Yaguo Lei, Naipeng Li, Ningbo Li, “A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings”, IEEE Transactions on Reliability, pp. 1-12, 2018.

DOI: 10.1109/TR.2018.2882682.

[28] Liang, Pengfei, et al. "Compound Fault Diagnosis of Gearboxes via Multi-label Convolutional Neural Network and Wavelet Transform." Computers in Industry (2019).

[29] Jiang, Guoqian, et al. "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox." IEEE Transactions on Industrial Electronics 66.4 (2019): 3196-3207.

[30] Singh, Inderpreet, S. Bahel, and S. B. Narang. "AWGN Channel Modeling using MATLAB." International Conference on Emerginging Technologies in Electronics and Communication (ICETEC 2013) 2013.

[31] Li, Jintao, and Jiancheng Fang. "Not Fully Overlapping Allan Variance and Total Variance for Inertial Sensor Stochastic Error Analysis." IEEE Transactions on Instrumentation and Measurement 62.10 (2013): 2659-2672.