Fault Diagnosis of Rolling Bearings Using an Improved Multi-Scale Convolutional Neural Network with Feature Attention mechanism

Abstract: Machine learning techniques have been successfully applied for the intelligent fault diagnosis of rolling bearings in recent years. This study has developed an Improved Multi-Scale Convolutional Neural Network integrated with a Feature Attention mechanism (IMS-FACNN) model to address the poor performance of traditional CNN-based models under unsteady and complex working environments. The proposed IMS-FACNN has a good extrapolation performance because of the novel IMS coarse grained procedure with training interference and the introduced the feature attention mechanism, which improves the model's generalization ability. The proposed IMS-FACNN model has a better performance than existing methods in all the examined scenarios including diagnosing the bearing fault of a real wind turbine. The results show that the reliability and superiority of the IMS-FACNN model in diagnosing faults of rolling bearings.

Keyword: multi-scale; convolutional neural network; fault diagnosis; deep learning; rolling bearings

Nomenclature

IMS-FACNN	Improved multi-scale convolutional neural networks with features attention mechanism
TICNN	Convolution neural networks with training interference
WT-CNN	Wavelet transform convolutional neural network
MS-CNN	Multi-scale convolutional neural network
MC-CNN	multi-scale cascade convolutional neural network
CNN	Convolutional neural network
SVM	Support vector machine
EMD	Empirical mode decomposition
VMD	Variational mode decomposition
ELMD	Ensemble local mean decomposition
SNR	Signal-Noise-Ratio
LR	Logistic regression
Conv	Convolutional layer
BN	Batch Normalization

ReLU	Rectified Linear Unit
$y^{l(i,j)}$	the dot product of kernel
W	represents the width of the kernel
$\mathbf{K}_i^l(j')$	the j^{th} weight of kernel l.
$z^{l(i,j)}$	the output of one neuron
μ	the mean of $y^{l(i,j)}$
σ^2	the variance of $y^{l(i,j)}$,
ε	a small constant
$\gamma^{l(i)}$	the scale to be learned
$\beta^{l(i)}$	the shift parameters to be learned
$a^{l(i,j)}$	the activation of $z^{l(i,j)}$
$y_j^s(j)$	the output of the $x(i)$ processed by IMS procedure with the interference
$O_i(k)$	the k^{th} output feature
φ_i	the output of fully connected layer
α_i	The feature weight of each scale
NREL	National Renewable Energy Laboratory
DAQ	Data Acquisition system

1. Introduction

The operation conditions and working environments have become more and more complicated as the rapid development of the wind turbine industries[1-2]. Generally, the operation conditions of a wind turbine have various uncertain factors including unsteady environment loads, constantly changing temperature and humidity [3], which leads to a high risk of damage on mechanical transmission components of a wind turbine and brings huge potential maintenance costs [4]. It is noted that rolling bearing is one of the damage hotspots taking a fault share of 30% in the mechanical transmission system of wind turbines [5]. An efficient fault diagnosis system is imperatively to be

developed for reducing maintenance cost, supporting the development of prognosis system to avoid more hazardous events. An intelligent fault diagnosis system consists of three basic steps: data acquisition, feature extraction and state classification. The classification accuracy was affected by the correctness of features extraction [6]. Its main purpose is to extract representative features which can characterize operation states and promote the accuracy in the recognition assignment of downstream conditions. To date, several methods of feature extraction have been developed and used in the fault diagnosis field. For instance, Guo *et al*. [7] used the improved Empirical Mode Decomposition (EMD) method combined with envelope spectrum, which could extract more fault information compared to the traditional EMD for the rolling bearing fault diagnosis. Liu *et al*. [8] utilized the multi-fractal detrended cross-correlation analysis and the EMD method to extract nonlinear information from different fault states. Zhang *et al*. [9] adopted the Variational Mode Decomposition (VMD) method to decompose vibration signals of rolling bearings, and combined with the Fast Fourier Transformation and envelope analysis to conduct the fault identification. Wang *et al*. [10] employed the Ensemble Local Mean Decomposition (ELMD) method to eliminate the residual noise for improving the Signal-Noise-Ratio (SNR) of a signal. The results indicated that the proposed ELMD algorithm had more advantages in feature extractions, which gathered more fault characteristic information from bearing vibration signals.

The aforementioned literatures indicate that various signal processing methods including all kinds of analysis methods in time-frequency domain, manifold learning and sparse representation have been used to obtain effective fault-related features on different levels from raw vibration signals. Following that, those extracted features are imported to a variety of shallow machine learning algorithms such as support vector machine (SVM) and Logistic Regression (LR). However, the upper boundary of the

performance of a machine learning algorithm heavily depends on the exactitude of the extracted features or representations [11]. Moreover, the limited diagnosis capabilities of shallow learning models lead to that their generalizations are not sufficient to address the complex state changes [12]. The limitations of a fault diagnosis model composed of a shallow learning approach and a feature extraction method are concluded as follows: i) the design of the feature extraction and classification processes which both affect diagnosis performance are examined independently. ii) the feature extraction process requires lots of diagnostic expertise and techniques of signal processing which are time-consuming and human experience dependent. iii) the existing methods of fault diagnosis are developed too specifically to be applied to other engineering areas.

As an alternative, deep learning approaches provide an effective way to overcome the shortcomings reviewed above for intelligent fault diagnosis. Deep learning methods have a good performance in classification and prediction tasks. The nonlinear processing units in a hierarchical architecture make a deep leaning method capable of modelling high-level representations of data [13]. Meanwhile, deep learning network connects to data directly rather than putting the extracted features into shallow machine learning algorithms to address diagnostic tasks. Due to the strong capabilities of learning and adapting, deep learning techniques have been applied to various engineering areas, e.g. computer vision, language processing and fault diagnosis. It has been well accepted that deep network models including deep brief network, sparse filtering, recurrent neural network and convolutional neural network (CNN) perform better than traditional approaches in fault diagnosis.

CNN is a typical algorithm of deep learning based on feed forward neural network involving convolutional computations and deep structures that are built by simulating visual perceptions [14]. They have representation learning abilities to perform translation-invariant classification of input

information according to its hierarchical structure [15], and thus CNN has been widely used in the computer vision, natural language processing and fault diagnosis fields [16]. Some studies show that CNN is capable of processing 1D to 3D signals. The main difference between CNN and traditional neural network is that the CNN approach is capable of stably learning the grid-like topology features while consuming less computational resources due to its features with weight sharing and sparse connectivity [17-18].

Due to the good performances on the classification and identification of structural damage states, CNN has been recently used for fault diagnosis of rotating machineries especially of rolling bearings. It is noted that the most common vibration signals acquired by sensors are naturally 1-D time series rather than 2-D images in the field of fault diagnosis. A preprocess is required for the sampled vibration signals before the use of a 2-D CNN models as stated in some recent studies. Chen *et al*. [19] used a 2-D CNN to address fault diagnosis of a gearbox, in which 256 statistical parameters including standard deviation, skewness, kurtosis and rotating frequency were used to build the input matrix with a size of 16×16. Janssens *et al*. [20] adopted the 2-D CNN to classify four conditions of a rotating machinery, where the input of the CNN was composed of frequency signals collected by two sensors. Similar to the research by Janssens, Wang *et al*. [21] used the wavelet spectrogram with a size of 32×32 as the input; the 2-D CNN model was utilized to recognize different working states of the rotor systems. Chen *et al*. [22] employed the continuous wavelet transform to obtain preprocessed representation images as the input of a 2-D CNN. The difference was that Chen's study used extreme learning machine as a classifier to complete the fault diagnosis task of rolling bearings.

However, a model trained by using 2-D images as the inputs of CNN cannot learn inherent vibration information from the images. In order to address this problem, a 1-D CNN integrating feature

extraction and classification together was proposed by Ince *et al.* for processing raw vibration signals which contained more information and were applicable for the capability of feature self-learning [23]. The 1-D CNN was applied to localize structural damage based on raw vibration signals [24]. Wei *et al.* adopted 1-D raw vibration signals of rolling bearing as the input of a deep CNN to achieve feature extraction and classification simultaneously. Their study showed that using raw signals as the inputs of a CNN model had better generalization ability and robustness in complex operation environments [25]. Nevertheless, except for variable operation conditions, the measured vibration signals are complicated due to the involvements of various mechanical rotating and reciprocating frequencies [26], indicating that vibration signals have usually contain characteristics in multiple time scales [27]. Consequently, the deep CNN could not well capture such inherent multiple characteristics. Therefore, Huang *et al*. used a multi-scale cascade CNN to enhance the classification information of the inputs, and tried to capture fault characteristic frequencies at different scales [28]. They found that signals convoluted by kernels of different sizes could have diverse resolutions in frequency domain. But the measured vibration signals of rolling bearing are very complicated due to the complex operation conditions which contains multiple modes in multiple time scale as discussed and revealed in the study [29], leading to inaccurate fault diagnosis results. Although convolutional kernels with different sizes were adopted to capture different information that contained distinct fault characteristic frequencies in Huang's study, the inherent multiple time characteristics of a vibration signal were ignored. In addition, most of the above studies adopted a small batch when training a model. However, it is noted that a small batch is more suitable for processing images rather than processing 1-D time series with a large mini-batch size.

As indicated by the existing literatures related to CNN applications for fault diagnosis, 1-D raw

vibration signals are more effective being the input of a CNN-based model compared to 2-D images [30]. The prediction accuracy of a CNN-based model is higher and more stable when more sensitive information can be extracted. Therefore, in order to satisfy the engineering needs, this study aims to develop an end-to-end intelligent fault diagnosis system based on a novel Improved Multi-Scale coarse grained procedure Convolutional Neural Networks with Features Attention mechanism (IMS-FACNN) framework. The developed IMS-FACNN considering inherent multiple time characteristics of raw signals, is capable of capturing different fault characteristic frequencies from sub-signals obtained from the IMS coarse-grained procedure extraction layer. The feature attention mechanism layer is introduced to adaptability weight to learned features. The weighted features are fused after the feature learning layer, which is no need to search the optimal time scale for the changing diagnostic objects. The key contributions of our study are summarized as follows:

Firstly, a novel Improved Multi-Scale coarse-grained procedure with training interference has been proposed to obtain vibration characteristics information with different scales from the input raw signal. Secondly, a feature attention mechanism has been introduced after the feature learning layer to address learned multiscale features fusion with adaptive weights. Thirdly, the optimal mini-batch of the vibration signals for the model training is obtained different from that used in the image processing. The features maps learned by the IMS-FACNN model are visualized to reveal the inner learning mechanism of the model, which has the complementary mechanisms at different scales.

The remaining parts of the paper are organized as follows. The CNN and related computational theories are presented in Section 2. The proposed IMS-FACNN for rolling bearing fault diagnosis is described in detail in Section 3. The experimental data of rolling bearing, superiorities of model structure, and investigation of influence by mini-batch are presented in Section4. Performances of the

model are examined in Section 5 throughout multiple scenarios test. Conclusions are presented in Section 6.

2. Convolutional neural network

CNN is a variant of multiple layers fully connected neural network that consists of various filter processes and one classification process. Compared with a fully connected network, CNN has more superior performance in a lot of engineering applications because of its local connects, shared weights and pooling operators [31]. A traditional CNN framework consists of convolutional layer, pooling layer, activation layer, classification layer and some techniques for improving model's generalization abilities e.g. batch normalization and dropout.

2.1. Convolutional Layer

The convolutional operation adopts sparse interactions, parameter sharing and equivariant representations to improve a machine learning system. In the convolutional layer, each convolutional filter has the same kernel. Through these kernels, local features are extracted from the local area. Each filter corresponds to the next layer one by one. The total number of the filters is called the layer depth. The $\mathbf{K}_i^l$ is the i^{th} filter in the layer l, and the $\mathbf{X}^{l(\mathrm{R}^j)}$ is the j^{th} local area in the convolutional layer l. The advantage of convolution kernel is that it can obtain the characteristics of rotation invariance, and the process is as follows:

$$y^{l(i,j)} = \mathbf{K}_i^l \cdot \mathbf{X}^{l(\mathrm{R}^j)} = \sum_{j'=0}^{W} \mathbf{K}_i^l(j')\mathbf{X}^{l(j+j')} \tag{1}$$

where $y^{l(i,j)}$ the dot product of kernel and the local area. W represents the width of the kernel. $\mathbf{K}_i^l(j')$ represents the j^{th} weight of kernel l.

2.2. Batch Normalization Layer

Due to the reduction of internal covariate shift, BN layer can accelerate the network training, and

solve the gradient dispersion problem. The BN operation solves the gradient vanishing problem caused by activation function. The parameters in a CNN model are normalized by the BN operation for improving network generalization capabilities. The BN is usually placed following the convolutional layer and the full connection layer, or before the activation layer. Inputting the *n*-dimensional array $\mathbf{y}^{l} = (\mathbf{y}^{l(1)}, \mathbf{y}^{l(2)}, \cdots, \mathbf{y}^{l(n)})$ to the l^{th} BN layer is represents as $\mathbf{y}^{l(i)} = (y^{l(i,1)}, y^{l(i,2)}, \cdots, y^{l(i,n)})$ 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 formula of the BN operation is described as follows:

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

$$\mu = \frac{1}{n} \sum_{i=1}^{n} y^{l(i,j)} \tag{3}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (y^{l(i,j)} - \mu)^2 \tag{4}$$

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

2.3. Activation Layer

It is essential to add an activation function after the convolutional layer for enhancing the non-linear expression ability of the input signal and making the learned features more distinguishable. In recent years, Rectified Linear Unit (ReLU) that is the most widely used activation unit, has been applied to CNNs for accelerating the convergence. Combined with back propagation learning method to adjust parameters, the ReLU makes shallow weights more trainable. The formula of the ReLU is shown in Eq. (5):

$$a^{l(i,j)} = f(z^{l(i,j)}) = \max\{0, z^{l(i,j)}\} \tag{5}$$

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

2.4. Pooling Layer

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

$$p^{l(i,j)} = \max_{(j-1)W+1 \le t \le jW}{}^{\{a^{l(i,t)}\}} \tag{6}$$

where $a^{l(i,t)}$ is the value of the t^{th} neuron in 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 layer l of the pooling, and $t \in [(j-1)W+1, jW]$.

3. The proposed IMS-FACNN framework

Considering the inherent multi-scale characteristic of measured signals contained with complex patterns at multiple time scales, a novel intelligent fault diagnosis approach named IMS-FACNN is developed to perform a more accurate fault diagnosis.

3.1. The Improved Multi-Scales coarse-grained procedure with Interference

The extrapolation capacity of a CNN model is improved due to the capability of feature self-learning ability by learning the inherent vibration characteristics from a raw signal. However, measured signals contain vibration inherent vibration characteristics in multiple time scales, which leads to the model have a bad robustness without considering such complex patterns.

Therefore, in this section, a novel Improved Multi-Scales coarse-grained procedure based on the traditional multiscale analysis approaches has been proposed to obtain more information in the

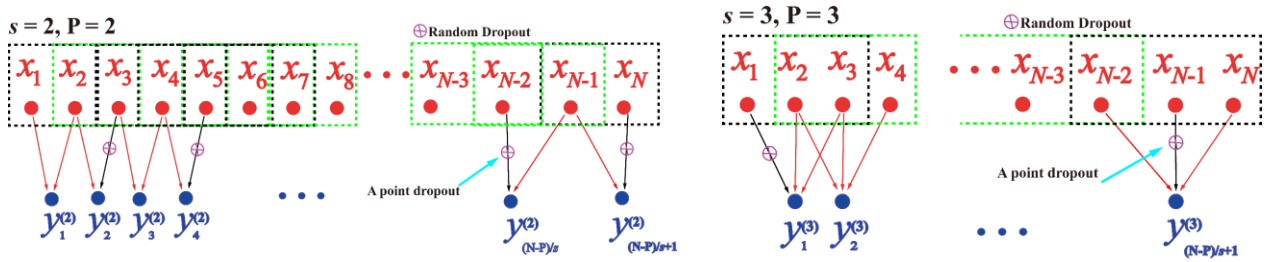
multiple time scales. [27]. In traditional multiscale studies for fault diagnosis, however, some inherent feature information would be ignored because of using non-overlapping window to divide time scales. The length of the sub-signals decrease exponentially leading to that the model must be modified when the input vibration signal changes. To address the above problems, a continuous shift operation by overlapping window is used to obtain average data in the proposed IMS coarse-grained procedure for obtaining more useful information from raw signals in the multiple time scales meanwhile addressing the shortcoming of the model needed to be modified frequently. In addition, a training interference is introduced in the IMS coarse-grained procedure for improving the extrapolation performance of the proposed model.

Figure 1 presents the improved multiscale coarse-grained procedure with interference for the scale factor s = 2 and s= 3. The IMS coarse-grained time series at a scale factor of s can be obtained by Eq. (7).

$$\mathrm{P}=\textbf{floor}(s/2)$$

$$y_j^s=\begin{cases}\dfrac{1}{s}\displaystyle\sum_{i=(j-1)\mathrm{P}+1}^{(j-1)\mathrm{P}+1+s} x(i),\ j\in[1,\dfrac{N-s}{\mathrm{P}}+1],s\geq 2\end{cases} \tag{7}$$

where **floor ()** is a round toward negative infinity function. P means the steps of sliding window.



(a) Scale = 2 (b) Scale = 3

Figure 1: The demonstration of the IMS procedure with interferences for s = 2 and s = 3

More importantly, as shown in Figure 1, some tricks are used to serve as interference in the IMS

procedure when training a model. In every coarse-grained procedure, as shown in Figure 1, some points of a raw signal are abandoned by dropout technology during training process. That not only prevents the model from overfitting, but also enhances the model's adaptability when working conditions changed. Thus, the outputs y is change by Eq. (8).

$$\begin{aligned} & p \sim Uniform(0.1 \sim 0.25) \\ & \mathrm{r}_i^1(k) \sim Bernoulli(p) \\ & \tilde{K}_i^1 = \mathrm{r}_i^1 \cdot K_i^1 \Big/ p \\ & y_j^s(j) = \tilde{K}_i^1 \cdot x(i) \end{aligned} \tag{8}$$

where $\cdot$ denotes the element-wise product, the value of dropout rate follows uniform distribution U(0.1,0.25), and $\mathrm{r}_i^1(k)$ follows Bernoulli distribution, which is used to decide whether the k^{th} element in the i^{th} frame of the first-layer convolutional. $y_j^s(j)$ is the output of the $x(i)$ processed by IMS procedure with the interference.

3.2 Feature Attention mechanism layer

A feature attention mechanism is proposed to adaptively score the features learned at different scales and assign them weights, thus the probability of each mode is better calculated in the final fully connected layer. The proposed Features Attention mechanism layer is shown in Figure 2.

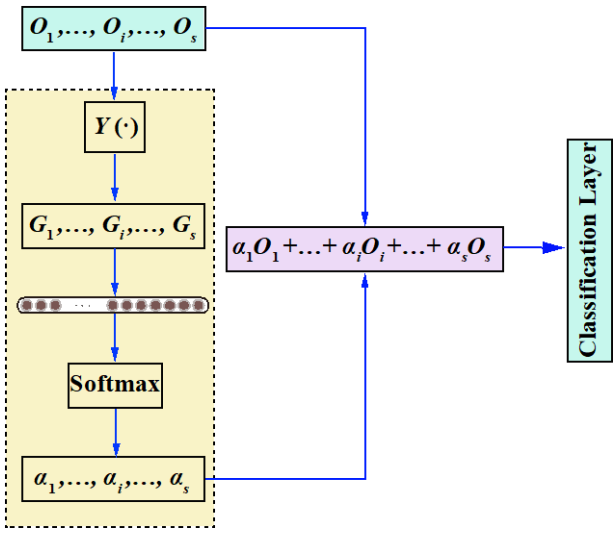


Figure 2: The feature fusion process with attention mechanism

In the feature fusion process with attention mechanism layer, the features learned from sub-signals that are extracted by IMS layer $O_1,\cdots,O_i,\cdots,O_s$ are summed by function $Y(\cdot)$ to get features $G_1,\cdots,G_i,\cdots,G_s$.

$$G_i = Y(O_i) = \sum_{k=1}^{M} O_i(k) \tag{9}$$

where $O_i(k)$ is the k^{th} output feature O_i, and M is the number of IMS coarse-grained output data, which is determined by the number of convolution kernels in the last convolution operation.

The feature weights of each scale $\alpha_1,\cdots,\alpha_i,\cdots,\alpha_s$ are obtained through a fully connected layer and a Softmax function, which are calculated by Eq.(10).

$$\begin{cases} \alpha_i = \text{Softmax}(\varphi_i) = e^{\varphi_i} \Big/ \sum_{k=1}^{s} e^{\varphi_k} \\ \sum_{i=1}^{s} \alpha_i = 1 \end{cases} \tag{10}$$

where φ_i is the output of fully connected layer. The feature weight of each scale α_i is calculated by Softmax function and mapped it to probability Space (0,1).

The feature fusion Z related to α_i and O_i is calculated by Eq.(11).

$$Z = \sum_{i=1}^{s} \alpha_i O_i \tag{11}$$

3.3. The IMS-FACNN architecture

Different Convolutional kernels are used to extract vibration characteristics from each sub-signal in the IMS extraction layer. A framework based on IMS-FACNN is developed for fault diagnosis of rolling bearings as presented in Figure 3. The IMS-FACNN framework consists of an input layer, an

IMS coarse-grained procedure extraction layer, feature learning layer and a classification layer. In addition, a feature attention mechanism layer is originally proposed to be integrated with the IMS-FACNN framework for a better feature extraction.

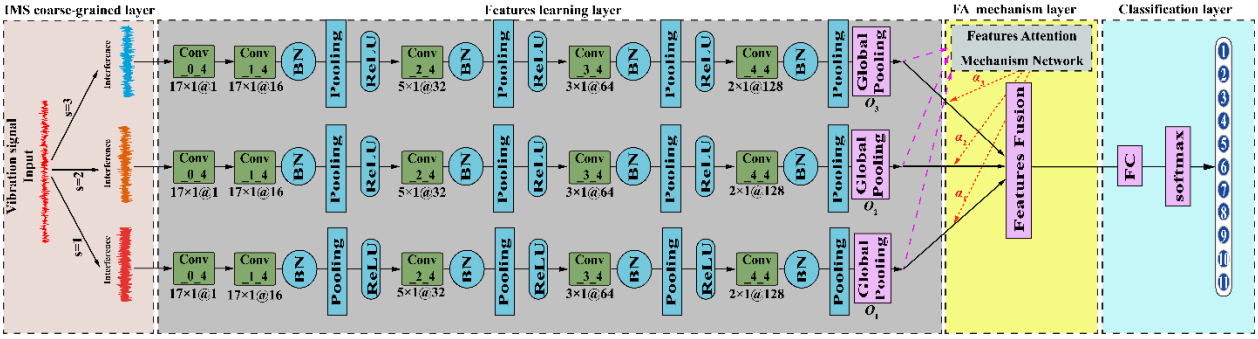


Figure 3: The structure of proposed IMS-FACNN model

Figure 3 shows that *n* sub-signals in different time scales are obtained from a raw signal by the IMS coarse-grained procedure. Features learned from each sub-signal by feature learning layer are scored by the feature attention mechanism layer, which are weighted and fused into classification layer to address fault diagnosis. In contrast to traditional CNN and TICNN model, the advantage of structure of the proposed IMS-FACNN model is that the IMS-FACNN model is capable of examining time multi-scale features, which can capture more information for improving the performance of the model. In contrast to the MS-CNN and MC-CNN models, the advantage of structure of the proposed IMS-FACNN model is that a feature attention mechanism layer is involved to improve the model performance by weighing features according to their scale information. In addition, there is no need to search the optimal time scale for the changing diagnostic objects. Moreover, the problem of high time complexity caused by a single large convolution kernel has been improved by adopting the form of continuous small convolution to reduce model complexity.

3.4. Fault diagnosis based on the IMS-FACNN

In Section 3.4, an end-to-end fault diagnosis system based on the proposed IMS-FACNN architecture is presented. The flowchart is shown in Figure 4, the proposed involved in the IMS-

FACNN architecture are summarized as follows.

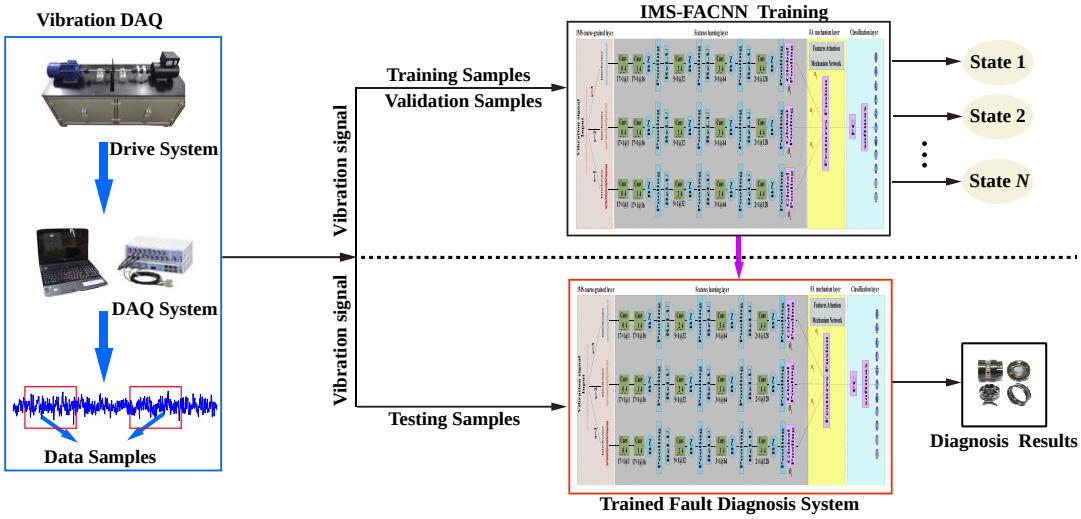


Figure 4: The flowchart of the proposed IMS-FACNN based on fault diagnosis system for rolling bearing

Step 1: Vibration data from different conditions of rolling bearings are collected by a Data Acquisition (DAQ) system. The whole signals would be segmented into multiple small segments for the model training, validation and testing.

Step 2: An end-to-end fault diagnosis system is established with training samples based on the IMS-FACNN that uses the vibration data as the inputs and takes the corresponding condition labels as the output. Through the offline training of the model, feature extraction and classification are realized.

Step 3: The testing samples will be imported to the trained fault diagnosis system for directly diagnosing the rolling bearings under different conditions.

4. Experiment Description

4.1 The description of experiment datasets

In this section, two public experiments of rolling bearings to validate the proposed the IMS-FACNN. The experimental data is the bearing dataset of Case Western Reserve University (CWRU) [32] and Xi'an Jiao Tong University (XJTU) [33], which are used to test the generalization performance of the IMS-FACNN model in different scenarios.

The data of CWRU covering normal state, inner race fault, ball fault and outer race fault in different azimuths (3, 6 and 12o'clock directions) are selected to validate the developed IMS-FACNN model, made by an electro-discharge machining (EDM) are examined for each fault category above. In total, 11 sets of data 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

As shown in Figure 5(b), the data of XJTU covering inner race fault, cage fault outer race fault, and hybrid faults that consist inner race, ball, cage and outer race failure are selected to validate the developed IMS-FACNN model when the failures are weak. The type of tested bearings is LDK UER 204, and the detailed 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

Figure 5 shows the photographs of the failure bearings.

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

Figure 5: Photographs of the XJTU bearings

4.2 The description of real wind turbines datasets

Two datasets of rolling bearings of the real wind turbines are used to examine the proposed IMS-FACNN model.

One of the datasets is the wind turbine vibration condition monitoring benchmarking datasets provided by National Renewable Energy Laboratory (NREL), which is used to examine the diagnostic

performance of the proposed IMS-FACNN in the real world condition [34]. In the NERL datasets, the test facilities are shown in Figure 6.

Figure 6: The real wind turbine drivetrain dynamometer installation

The other bearing dataset collected from a 3 MW wind turbine in Qingdao, China is used to examine the diagnostic performance of the proposed IMS-FACNN model in real world condition [35]. The dataset contains four conditions of bearings including normal, inner race fault, ball fault and outer race fault. The sets of training, validation and test are independence.

4.3 Data Acquisition

Two acceleration sensor PCB 352C33 are fixed to two direction of the test bearing by magnetic seat. The DT9837 portable dynamic signal collector was used to collect vibration signals during the test. The sampling parameters are set as shown in Figure 7. In the experiment, the sampling frequency is set to 25.6 kHz, the sampling interval is 1 minute, and the length of each sampling is 1.28 s. In each sampling, the obtained vibration signal is stored in a csv file.

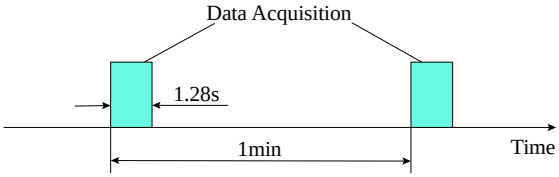


Figure 7:Sampling settings for vibration signals

4.4 Sample Division

Due to the short sampling time of CWRU, the total number of samples is insufficient. Therefore, when the scene tests based on CWRU data are constructed, the overlapping sampling technique is adopted to expand the samples. It is worth noting that the data of training, testing and validation sets are independent of each other. Figure. 8 presents the overlapping sampling technique.

In order to be more consistent with the realistic situations, the XJTU data is also used to produce sufficient samples of the test set in each scenario without the use of the overlapping sampling technique, which ensures the independence of the test set and the training set.

In the NREL dataset, four conditions includes normal, overheating, damages and dents are classified by the proposed IMS-FACNN model for examining the model's diagnostic performance. The sets of training, validation and test were selected on different days, which guarantees the independence between the data sets.

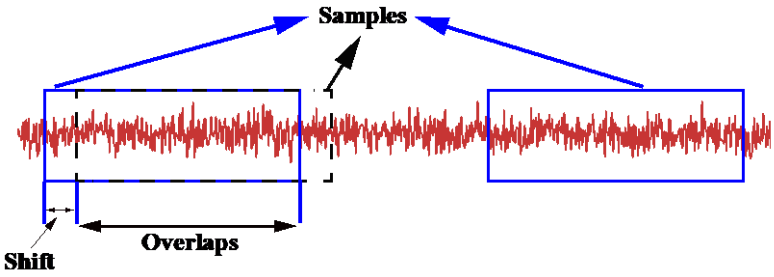


Figure 8: Data augment with overlap

Table 3 to Table 6 present the details of the samples of training, validation and testing, respectively. 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		3200	3200	3200	3200	3200	3200
Validation	0	400	400	400	400	400	400
Test		400	400	400	400	400	400
Train		3200	3200	3200	3200	3200	3200
Validation	1	400	400	400	400	400	400
Test		400	400	400	400	400	400
Train	2	3200	3200	3200	3200	3200	3200
Validation		400	400	400	400	400	400

Test		400	400	400	400	400	400
Train		3200	3200	3200	3200	3200	3200
Validation	3	400	400	400	400	400	400
Test		400	400	400	400	400	400

Table 4: Details of XJTU rolling bearing datasets

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

Table 5: Details of 750kW wind turbine bearing datasets

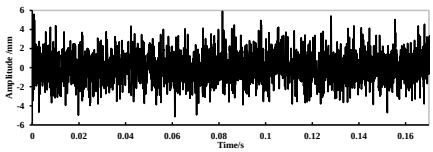
Type	Normal	Overheating	Damage	Dents
Labels	*0*	*1*	*2*	*3*
Train	*3200*	*3200*	*3200*	*3200*
Validation	*400*	*400*	*400*	*400*
Test	*400*	*400*	*400*	*400*

Table 6: Details of 3 MW wind turbine bearing datasets

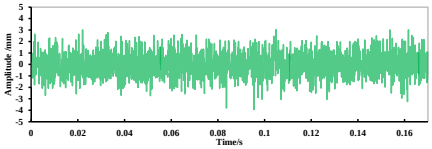
Type	*Normal*	*Inner Race*	*Ball*	*Outer Race*
Labels	*0*	*1*	*2*	*3*
Train	*3200*	*3200*	*3200*	*3200*
Validation	*400*	*400*	*400*	*400*
Test	*400*	*400*	*400*	*400*

4.5 Influence of mini-batch size

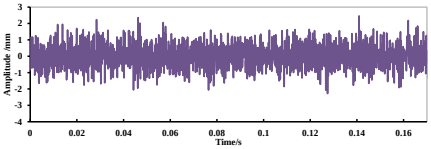
The mini-batch size determines the number interactions in every epoch, which also affects the anti-noise performance and calculation time of the proposed IMS-FACNN model. White Gaussian noises with different intensities are added into the raw vibration signals measured in CWRU. Figure 9 presents the raw and noised signals corresponding to the inner race fault condition.



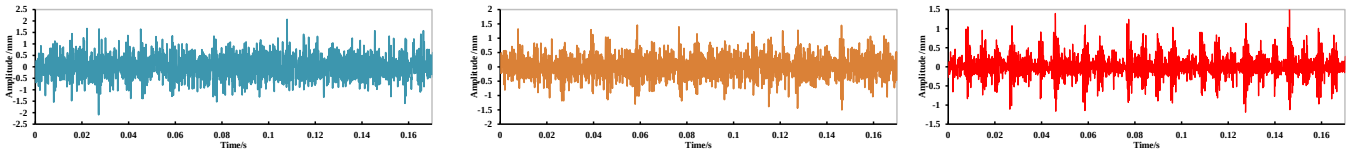
SNR = -4 dB



SNR = 0 dB



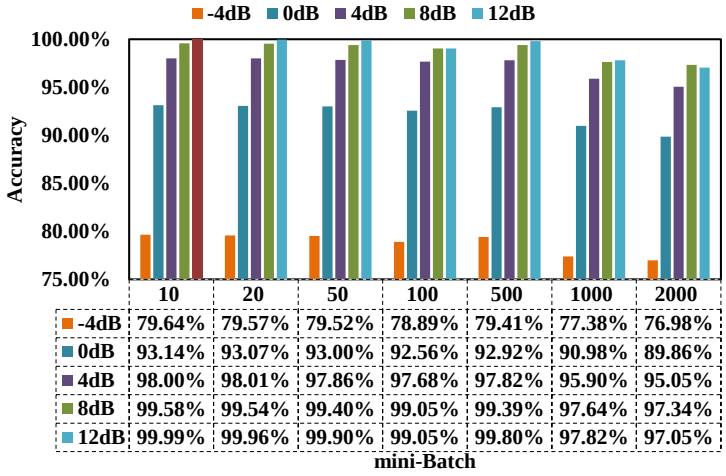
SNR = 4 dB



SNR = 8 dB SNR = 12 dB Original

Figure 9: The original signal of inner race fault with white Gaussian noise

The 0dB case means the noise has an equivalent power spectral density to the original signal. The -4dB means the ratio of the noise in a signal is much higher than that of the original signal. On the contrary, the 12 dB case means the proportion of noise in a signal is much smaller than that of the original signal. In order to cover the possible circumstances in the realistic engineering environments, the noisy signals with SNRs from -4dB to 12 dB are examined. As the batch size affects the learning efficiency and accuracy of the IMS-FACNN model, a sensitivity analysis of the mini-batch size to the accuracy is performed to find the optimal value for the subsequent researches. The mean diagnosis accuracy of the IMS-FACNN model in 20 trials for the noisy signals from -4dB to 12 dB are presented in Figure 10.



	10	20	50	100	500	1000	2000
-4dB	79.64%	79.57%	79.52%	78.89%	79.41%	77.38%	76.98%
0dB	93.14%	93.07%	93.00%	92.56%	92.92%	90.98%	89.86%
4dB	98.00%	98.01%	97.86%	97.68%	97.82%	95.90%	95.05%
8dB	99.58%	99.54%	99.40%	99.05%	99.39%	97.64%	97.34%
12dB	99.99%	99.96%	99.90%	99.05%	99.80%	97.82%	97.05%

Figure 10: The accuracy of the IMS-FACNN model using batch sizes from 10 to 2000

As shown in Figure 10, the diagnosis accuracy of the IMS-FACNN model is oscillatory along with the mini-batch size from 10 to 2000. The diagnosis accuracy of the IMS-FACNN model is relatively higher than using a bigger mini-batch for the signals with an arbitrary noise level. Generally,

the anti-noise ability of a CNN model is good when using a small batch (mini-batch size =10, 20 and 50) in the training process. When the SNR value equals to -4 dB, the accuracy of the model using a small batch (10, 20 and 50) is nearly 80%. It is noted that the diagnosis accuracy corresponding to the mini-batch of 500 is 79.41%, which is pretty high close to the accuracies corresponding to the small mini-batch (from 10 to 50) in the training process. However, the training process consumes more calculation resources when using a small mini-batch. Moreover, a small mini-batch is more suitable for dealing with images processing rather than time series, e.g. a raw vibration signal. Therefore, the combinations of the total epochs, batch size, total iterations and the epoch of model achieving 99% accuracy trained on GPU of 1050TI are presented in Table 7.

Table 7: The relationships about total epochs, size of batch, total iterations and the epoch of model achieved 99% accuracy

mini-batch size	2000	1000	**500**	100	50	20	10
Total Epochs	100	100	**100**	100	100	100	100
Time of 100 epochs (mins)	14	18	**12**	100	247	1270	5580
Total Iterations	1900	3900	**7900**	39500	79000	190000	395000
Achieve 99% accuracy at Epoch	53	38	**17**	8	3	2	2
Time of Achieve 99% accuracy (mins)	8	7	**3**	12	11	37	89

Table 7 indicates that a small mini-batch requires a long processing time. It is difficult to achieve this in the actual engineering application using limited computational resources, the time complexity of the diagnosis model is improved when using a small batch size. It is noted that an equivalent accuracy to the small mini-batch can be achieved for a larger mini-batch if examining more epochs. Moreover, the time of achieving the 99% accuracy decreases significantly. The IMS-FACNN model achieve the 99% accuracy at 17 epochs when the batch size is 500, which requires the shortest time among the examined configurations that can achieve a 99% accuracy.

4.6 Time complexity of the IMS-FACNN

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 IMS-FACNN model.

Table 8: The time complexity of the IMS-FACNN model and other CNNs-based methods

Method	Time Complexity in Training (The size of input signal = 2048×1)
CNN(s = 1)	O (45.1×10^4)
TICNN [25]	O (87.4×10^4)
MS-CNN [27]	O (178.8×10^4)
IMS-FACNN	O (98.5×10^4)
WT-CNN [35]	O (37.8×10^4)
MC-CNN [28]	O (218.2×10^4)

It is found in Table 8 that the complexity of the proposed IMS-FACNN method is smaller than other MS-CNN and MC-CNN, that is because the form of the first convolution sampling has been changed by adopting the form of continuous convolution instead of a large convolution kernel.

5. Experiment Validation

Different scenarios created based on the four aforementioned datasets are used to examine the proposed IMS-FACNN model. The data mining and setup of the deep learning model are conducted using the MATLAB® Deep Network Designer, MATLAB version 9.70 (R2019b, The MathWorks, Inc., Natick, MA, USA).

5.1 Noise scenario test

In order to confirmed the superiority of the IMS-FACNN model for identifying the failure types and the failure magnitudes of the bearings in the noise environments, the performances of other methods including a traditional CNN, the MS-CNN [27], the TICNN [25], the MC-CNN [28], the SVM [36] and the WT-CNN [37] are examined by using CWRU data, which are presented in Figure 11 and Table 9.

Table 9: The accuracy of six methods examined by using CWRU data under different noisy environments

Algorithms	SNR(dB)				
	-4	0	4	8	12

CNN	60.36±**0.49**	86.32±0.45	92.41±**0.20**	97.91±0.**04**	98.59±0.02
SVM	52.95±2.08	61.14±0.63	77.95±0.29	96.14±0.14	99.77±0.12
MC-CNN	61.09±1.30	83.00±0.59	95.05±0.26	98.77±0.14	99.66±0.12
WT-CNN	40.32±1.16	81.50±0.51	88.55±0.21	93.45±0.09	95.00±0.09
TICNN	65.68±1.15	88.09±0.44	92.75±0.23	94.89±0.09	93.34±0.06
MS-CNN	61.03±1.22	87.19±0.37	93.38±0.29	98.90±0.14	99.68±0.06
IMS-FACNN	**79.41**±0.59	**92.92**±0.**34**	**97.82**±0.26	**99.39**±0.14	**99.80**±**0.02**

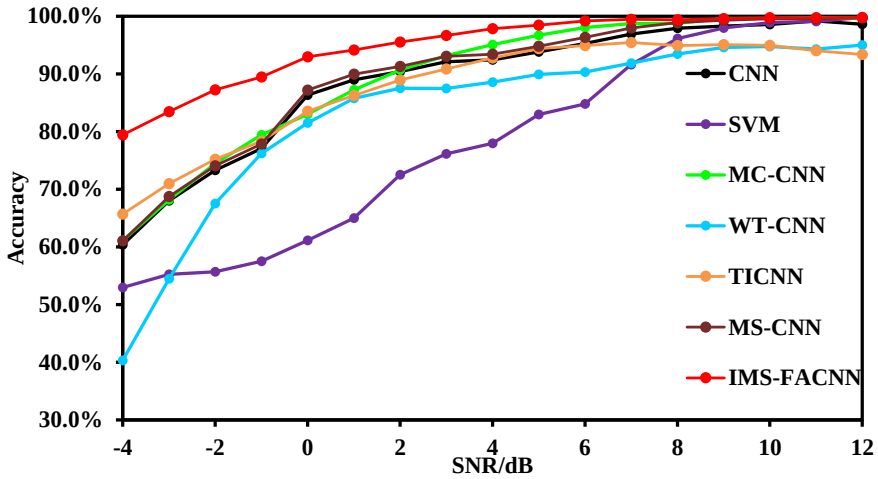


Figure 11: Noise test result

All of the six methods have a good diagnosis performance on high SNR signals. The responses time of the IMS-FACNN model is average 1.57s in 20 trails. However, the accuracies of the SVM and the WT-CNN decrease obviously the reduction of SNR. It can be explained by considering that these two model are built, respectively, by a shallow learning approach and a CNN developed specifically for processing images, resulting in a poor generalization ability. On the contrary, the TICNN, MC-CNN, CNN, MS-CNN and the IMS-FACNN that are capable of handling raw signals have good robustness under variety of noisy environments. Compared to the TICNN, the MC-CNN and MS-CNN, the proposed IMS-FACNN model has the best robustness and accuracy for each of examined SNRs. When the SNR equals to -4 dB, the accuracy of our IMS-FACNN model is nearly 80%, and the standard deviation of accuracy in 20 trials is lowest that is a unique performance that the other models cannot achieve.

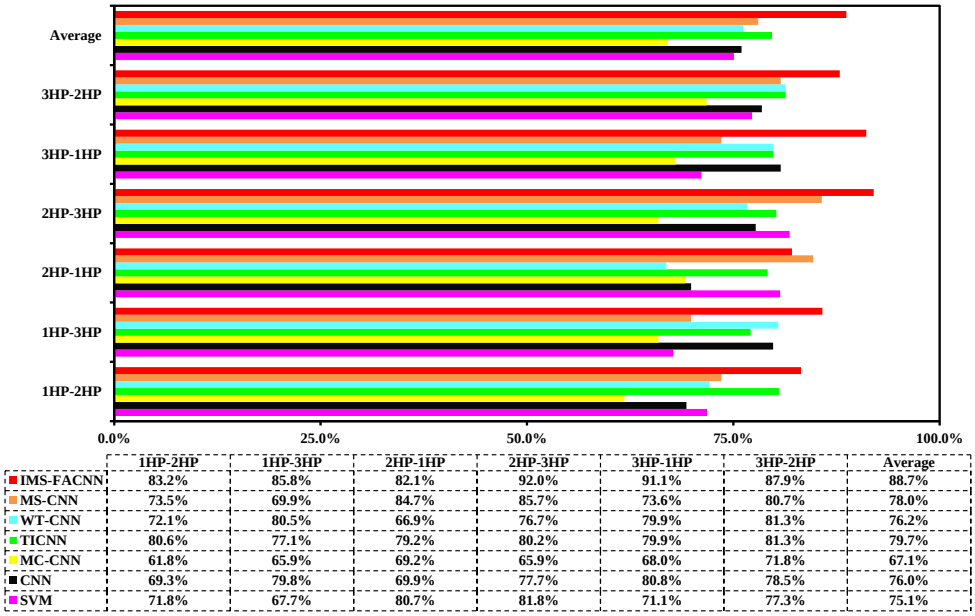
5.2 Load scenario test

In this section, the generalization ability of the IMS-FACNN under different load environments is investigated by using CWRU data. Table 10 illustrates the details of the load configuration of the training and testing processes. The mini-batch of 500 are considered in the IMS-FACNN.

Table 10: the details of the load configuration

Case number	Training samples	Testing samples
1	Load with 1 HP (*I*)	Load with 2 HP (*II*)
2	Load with 1 HP (*I*)	Load with 3 HP (*III*)
3	Load with 2 HP (*II*)	Load with 3 HP (*III*)
4	Load with 2 HP (*II*)	Load with 1 HP (*I*)
5	Load with 3 HP (*III*)	Load with 1 HP (*I*)
6	Load with 3 HP (*III*)	Load with 2 HP (*II*)

Figure 12 compares the result of the IMS-FACNN with the results of the traditional CNN, the SVM, the MC-CNN, the WT-CNN, the MS-CNN and the TICNN for load adaptation test.



	1HP-2HP	1HP-3HP	2HP-1HP	2HP-3HP	3HP-1HP	3HP-2HP	Average
IMS-FACNN	83.2%	85.8%	82.1%	92.0%	91.1%	87.9%	88.7%
MS-CNN	73.5%	69.9%	84.7%	85.7%	73.6%	80.7%	78.0%
WT-CNN	72.1%	80.5%	66.9%	76.7%	79.9%	81.3%	76.2%
TICNN	80.6%	77.1%	79.2%	80.2%	79.9%	81.3%	79.7%
MC-CNN	61.8%	65.9%	69.2%	65.9%	68.0%	71.8%	67.1%
CNN	69.3%	79.8%	69.9%	77.7%	80.8%	78.5%	76.0%
SVM	71.8%	67.7%	80.7%	81.8%	71.1%	77.3%	75.1%

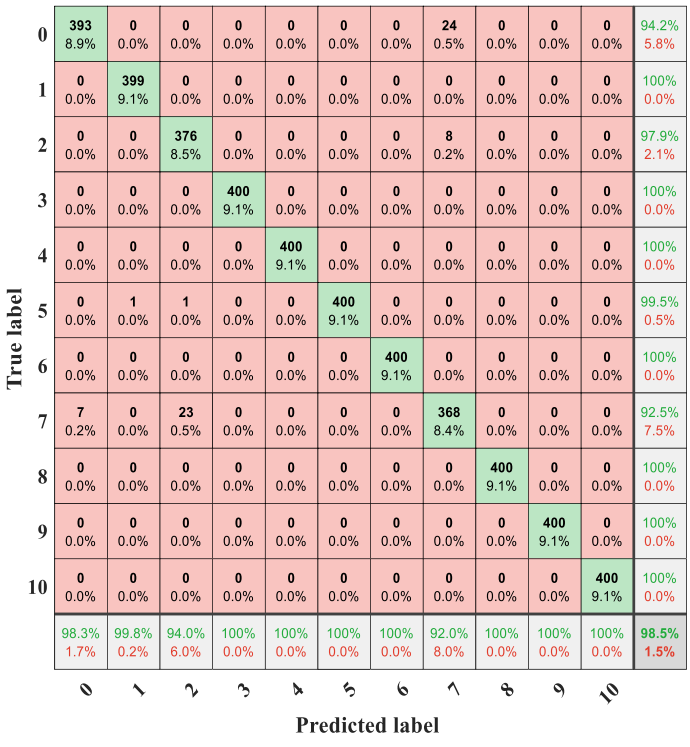
Figure 12: The results of IMS-FACNN
compared with SVM, WT-CNN,TICNN, MC-CNN, MS-CNN and CNN on load adaptation test

As shown in Figure 12, the mean accuracy of the IMS-FACNN model is higher than the others for the changing loads. The responses time of the IMS-FACNN model is average 1.42s in 20 trails. The SVM and the WT-CNN both have a mean accuracy around 75%, which are higher than the MC-CNN. It is because that the features used in the SVM model and WT-CNN model that are energy entropy and the WT images respectively, have a good robustness under variable load environments.

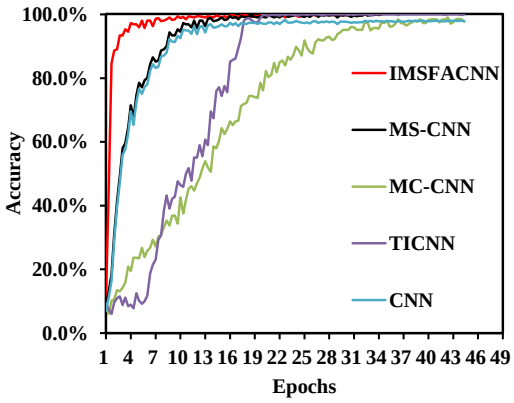
The mean accuracy rate of the IMS-FACNN reaches nearly 89% when the load environments changed, which is highest in the accuracies of all methods. The reason is that the IMS-FACNN model is capable of extracting more information than the other methods from a raw signal. Compared to the TICNN that is capable of load domain adaptation, the IMS-FACNN method helps a diagnosis system improve accuracy by nearly 9% on the load adaptation test.

5.3 Noise and load mixed adaptation scenario test

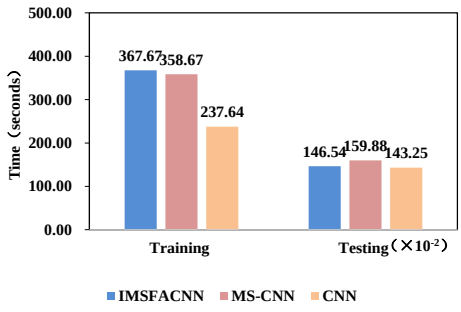
A scenario with the mixed load data (0HP, 1HP, 2HP and 3HP) is examined in this section to verify the reliability of the proposed IMS-FACNN model. The independent test set ensures the reliability of test results. The model diagnosis accuracy rate, false alarm rate and computational time when training and testing are shown in Figure 13.



(a) Confusion matrix of diagnosis results by IMS-FACNN



(b) Training accuracy of different methods based on raw vibration signals



(c) Computational time of IMS-FACNN, MS-CNN and CNN

Figure 13: Diagnosis results in the noise and load mixed adaptation scenario

Figure 13(a) shows the confusion matrix of the diagnosis result by IMS-FACNN model in the mixed load adaptation test. It gives correctly classified samples and misclassified samples for 11 bearing working conditions. The result shows that most of the working conditions can be accurately distinguished, only a small false alarm rate exists between the ball state and the normal state. Figure 13(b) shows the training accuracies of the five diagnostic methods based on raw vibration signals. The results show that although the five methods have achieved higher accuracy after several epochs, but the time required of five methods for reaching 99% accuracy is different. The accuracy of TICNN model at the beginning of training increases slowly because that it only introduced dropout technology for enhancing generalization ability of the model. The MC-CNN model does not consider the time multi-scale characteristics of a raw vibration signal, so the rising speed of training accuracy is slower than the MS-CNN model which considers time multiscale. Importantly, the features learned from different time scales have different influences, which is considered by the proposed IMSFACNN model, thus it can learn effective fusion features faster (at the fully connected layer) than the MC-CNN model. Figure 13(c) shows the computational time of the different models when training and testing. The computational time of training (50 epochs) the proposed IMS-FACNN model is longer than the MC-CNN model and the CNN model, since the IMS-FACNN model has been introduced the attention mechanism that leads to consuming more time. But the testing time of the proposed IMS-FACNN model is basically similar with the existing model. The computational time of training and testing the proposed model will be shorter in the case of better computing equipment, which represents the proposed model is suitable for actual projects.

In order to reflect the effectiveness and superiority of the proposed IMS-FACNN model, a scenario of mixed data with different noise is set, where training set includes 0HP,1HP, 2HP and 3HP

clean bearing vibration signals, the testing set includes 0HP,1HP, 2HP and 3HP bearing vibration signals with different SNR. The diagnostic results of the IMS-FACNN model compared to other existing models are shown in Figure 14.

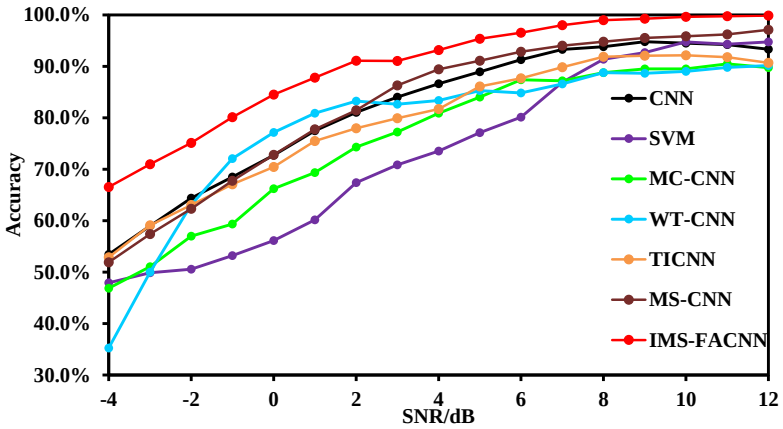


Figure: 14: The diagnostic results of the IMS-FACNN model
compared to other existing models on the noise and load mixed adaptation scenario test

Figure 14 shows the diagnosis results tested in the noise and load mixed adaptation scenario. The responses time of the IMS-FACNN model is average 1.49s in 20 trails. The diagnosis accuracies of all the methods in this scenario are lower than those in only noise scenario. In the case of such load changes and the presence of noise (-4dB), the proposed IMS-FACNN still has an accuracy close to 70%, which is nearly 20% higher than the existing methods. In this scenario (-4dB), MS-CNN and TICNN still have slightly better diagnostic performance of MC-CNN, that is because the MS-CNN consider multiple time scale and the TICNN use training interference. The proposed IMS-FACNN model combines the advantages of the MS-CNN and the TICNN models, considering more continuous multi-scale information and introducing the feature attention mechanism, leading to a better performance in the noise and load mixed adaptation scenarios.

5.4 Damage degree adaptation scenario test

In this section, the extrapolation performance of the model is further tested using the data of XJTU. It is worth noting that the data distributions in the degradation processes of bearings are different. Thus,

three phrases of the degradation processes are presented in Figure 15. Phase_1 to Phase_3 denotes the fault evolution processes of the bearings.

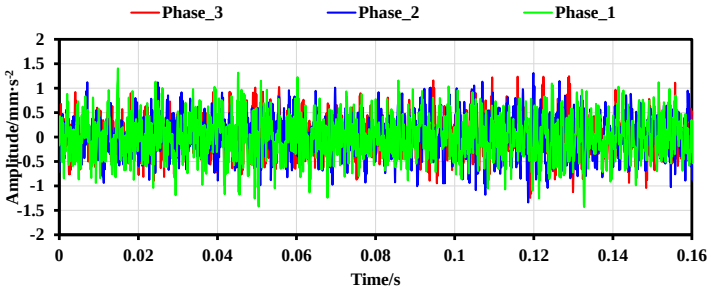
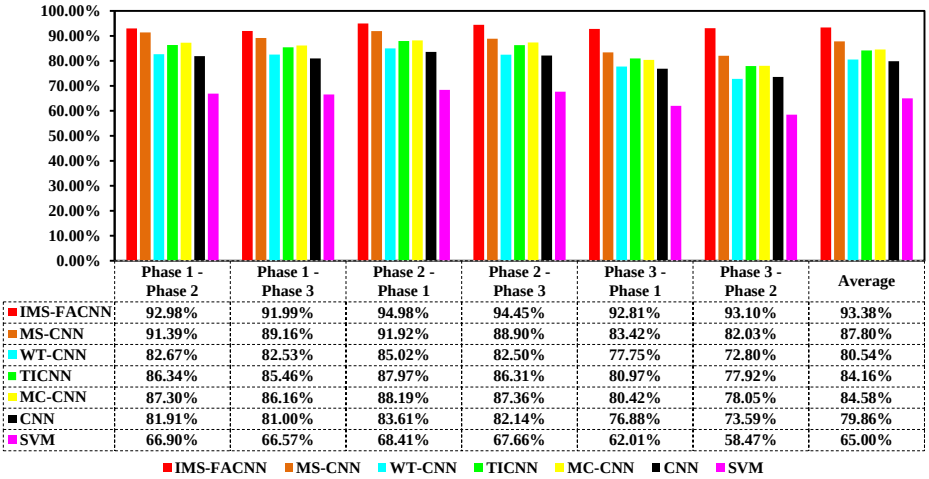


Figure 15: The signals in different damage phases

As shown in Figure 15, three phases represent the development process of bearing failures. Due to the evolution of the bearing fault, the data is regarded as unsteady, which is used to set the damage degree adaptation scenario to test the extrapolating performance of the proposed model. Eight cases are examined as presented in Table 11. The diagnosis results are shown in Figure 16.

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



	Phase 1 - Phase 2	Phase 1 - Phase 3	Phase 2 - Phase 1	Phase 2 - Phase 3	Phase 3 - Phase 1	Phase 3 - Phase 2	Average
IMS-FACNN	92.98%	91.99%	94.98%	94.45%	92.81%	93.10%	93.38%
MS-CNN	91.39%	89.16%	91.92%	88.90%	83.42%	82.03%	87.80%
WT-CNN	82.67%	82.53%	85.02%	82.50%	77.75%	72.80%	80.54%
TICNN	86.34%	85.46%	87.97%	86.31%	80.97%	77.92%	84.16%
MC-CNN	87.30%	86.16%	88.19%	87.36%	80.42%	78.05%	84.58%
CNN	81.91%	81.00%	83.61%	82.14%	76.88%	73.59%	79.86%
SVM	66.90%	66.57%	68.41%	67.66%	62.01%	58.47%	65.00%

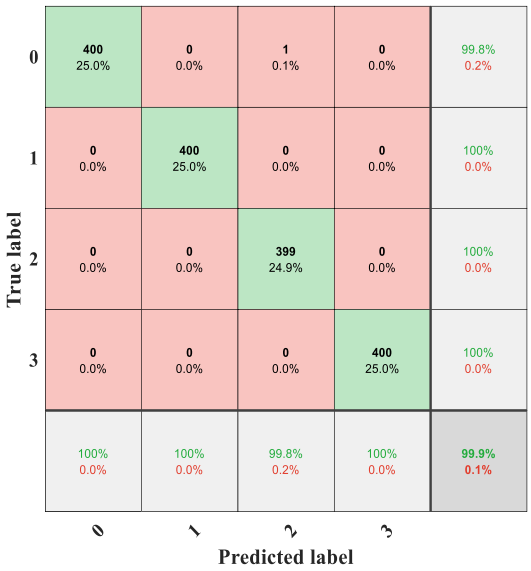
Figure 16: Methods compared results tested by the data with fault development

The responses time of the IMS-FACNN model is average 1.45s in 20 trails. The accuracy of the

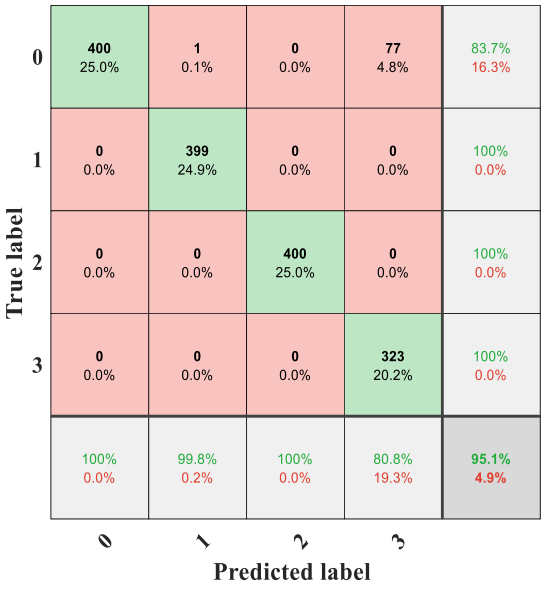
IMS-FACNN model is higher than 84% for each examined case and the average diagnosis accuracy achieves 93.38%, which is the highest among all the diagnostic model. The diagnosis accuracy of the IMS-FACNN model is higher than the MS-CNN model by nearly 6%, which shows that the introduced features attention mechanism and the improved multiscale layer is effective. The IMS-FACNN 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 IMS-FACNN model has a good extrapolation performance.

5.5 Actual Wind turbine bearing damage scenario test

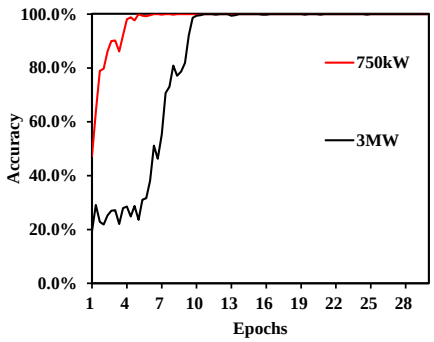
The vibration datum of the bearings respectively in a 750 kW wind turbine a 3 MW wind turbine are used to examine the fault diagnostic performance of the proposed IMS-FACNN model in this section. The diagnostic results of the IMS-FACNN model when dealing with the vibration datum of the bearings respectively in the 750 kW and 3 MW wind turbines are shown in Figure 16.

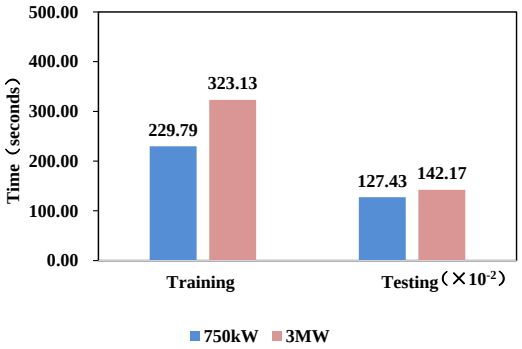


(a) Confusion matrix
of diagnosis results for the 750 kW model



(b) Confusion matrix
of diagnosis results for the 3 MW model





(c) The accuracy of the IMS-FACNN model (d) Computational time of the IMS-FACNN model

Figure 17: The diagnostic results of the IMS-FACNN model for diagnosing bearings in real wind turbines

Figure 17(a) and Figure 17(b) give the confusion matrixes of diagnosis results corresponding to the 750 kW and 3 MW wind turbine models. In Figure 17(a), it is observed that the proposed IMS-*FACNN model has a good performance when diagnosing the working conditions of the 750 kW wind* turbine bearing, although a minor false alarm rate exists between the normal state and the bearing damage due to the complexity of the real world working condition. Figure 17 (b) indicates that the proposed IMS-FACNN model is able to distinguish the working conditions of the 3 MW wind turbine bearing. Similarly, false alarm rates can be observed between the normal state, ball creak and outer race failures. Compared to the diagnosis performance for the 750 kW model, the false alarm rates of the 3 MW wind turbine are slightly larger. As shown in Figure 17(c) and (d), the computational time of the proposed IMS-FACNN model when diagnosing the bearing fault of the 750 kW wind turbine is less than diagnosing the bearing fault of the 3 MW wind turbine. The above results proves the proposed IMS-FACNN model has a good performance when diagnosing bearing fault in a real wind turbine.

5.6 Networks Visualization

The black-box property of CNN makes what have been CNN learned become difficult to understand. Thus, for understanding the inner operation of our IMS-FACNN visually, we visualize the activations in our neural network. The features distribution of testing samples of the signals with a SNR of 0 dB are visualized in Figure 18 by t-SNE [38].

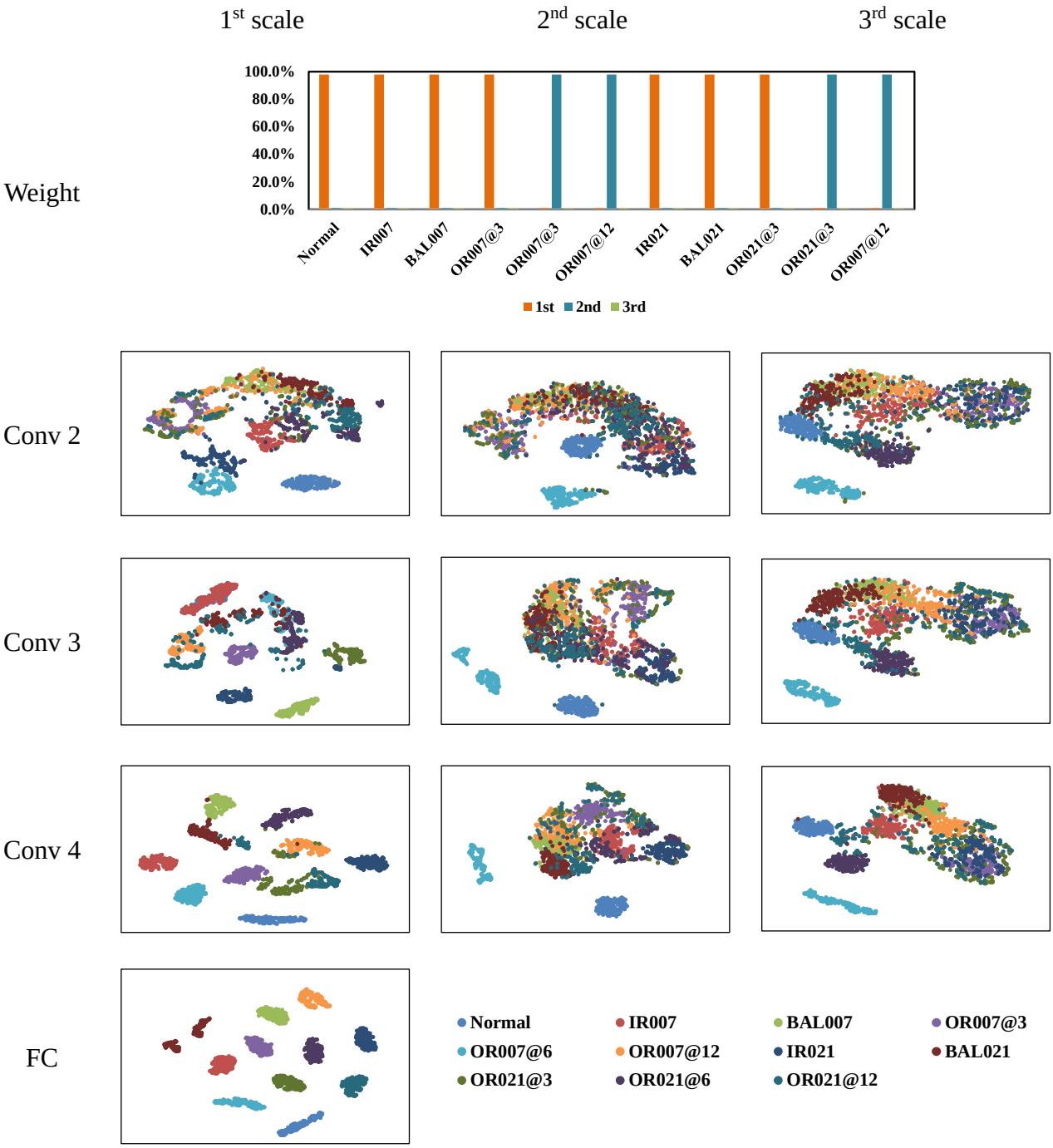


Figure 18: Feature visualization by t-SNE
reduced from the learned multi-scale representations for the testing data

As shown in the Figure 18, every working conditions has been well distinguished in the last fully connected layer, which means that the features learned by IMS-FACNN are representative. The clustering results of features learned from the sub-signals with the different time scales are different in the penultimate convolutional layers. The weights of features with different scales acquired by the

attention mechanism are marked at the top of Figure 18. An interesting phenomenon has been found that the "1st scale" has the largest weight, and the corresponding feature clustering result is also clearest. The weights given to the features learned from other scales are relatively small, and the corresponding clustering results are not clear, which shows that the proposed IMS-FACNN model is effective. However, this does not mean that these scales are eliminated, as shown in the top of Figure 18, "2nd scale" plays a leading role instead of "1st scale" when distinguishing the outer race bearing fault, these scales have learned the features but "1st scale" has not learned, As a result, the complementarity between these features (learned from different scales) is beneficial for improving the reliability of the probability calculation in the last fully connected layer.

6. Conclusion

In this paper, a novel Improved Multi-Scale coarse-grained procedure Convolutional Neural Networks with Feature Attention mechanism has been developed, which can directly works on raw vibration signals for achieving an accurate fault diagnosis of the rolling bearings in complex actual situations. In contrast to traditional multi-scale coarse-grained procedure, the proposed improved multi-scale coarse-grained procedure is achieved by a continuous shift operation and a training interference is introduced. In contrast to traditional multi-scale CNN model, a features attention mechanism is introduced into the model to improve the extrapolation performance of the IMS-FACNN model. The main conclusions are drawn as follows.

1. The IMS-FACNN model has a better diagnosis performance under multiple scenarios than MS-CNN, because the improved multi-scale coarse-grained procedure can obtain more useful information and have anti-interference, which is implemented by a continuous shift operation and a training interference.

2. The proposed IMS-FACNN model can distinguish the fault type, damage degree and fault location, which has a good performance under multiple scenarios test.

3. The proposed IMS-FACNN model has a better extrapolation performance compared to WT-CNN, TICNN, CNN, MS-CNN, MC-CNN and SVM under multiple scenario tests. Compared to the existing multi-scale methods such as the MS-CNN and the MC-CNN, the accuracy of the proposed IMS-FACNN is more than 6% under multiple scenario tests.

4. The proposed IMS-FACNN model achieves a high diagnosis accuracy for both the 750 kW and 3 MW wind turbines operating in the real world.

5.The "2^{nd} scale" plays a leading role instead of "1^{st} scale" when distinguishing the outer race bearing fault. The clustering results of the features learned from different scales show that the features with greater weight given by the attention mechanism have a clearer clustering performance. Although the clustering results of the features with lower weights are not clear, there are complementary mechanism between features with different scales.

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), Royal Society (grant number: IEC\NSFC\170054), the National Renewable Energy Laboratory and the United States Department of Energy for providing the benchmarking datasets of wind turbine gearbox vibration condition monitoring.

References

[1] Chen X , Zhou J , Xiao J , et al. Fault diagnosis based on dependent feature vector and probability

neural network for rolling element bearings[J]. Applied Mathematics and Computation, 2014, 247:835-847.

[2] Lewis J I, Wiser R H. Fostering a renewable energy technology industry: An international comparison of wind industry policy support mechanisms [J]. Energy Policy, 2007, 35(3):1844-1857.

[3] Christoph S, Anette F, Pascal W,et al. Evaluation and Control of Loads on Wind Turbines under Different Operating Conditions by Means of CFD[M]. Springer International Publishing, 2016.

[4] Miao W P, Li C, Wang Y B, et al. Study of Adaptive Blades in Extreme Environment using Fluid-Structure Interaction Method[J]. Journal of Fluids and Structures, 2019, 91: 102734.

[5] Shao, H D, Jiang H K, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network[J]. Measurement Science & Technology, 26(11):115002.

[6] Lu, S, Oki, K, Shimizu, Y,et al. Comparison between several feature extraction/classification methods for mapping complicated agricultural land use patches using airborne hyperspectral data[J]. International Journal of Remote Sensing, 28(5):963-984.

[7] Guo T, Deng Z. An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing [J]. Applied Acoustics, 2017, 127:46-62.

[8] Liu H , Zhang J , Cheng Y , et al. Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis[J]. Journal of Sound and Vibration, 2016:S0022460X16304515.

[9] Zhang M, Jiang Z, Feng K. Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump[J]. Mechanical Systems and Signal Processing, 2017, 93:460-493.

[10]Wang L , Liu Z , Miao Q , et al. Complete ensemble local mean decomposition with adaptive noise

and its application to fault diagnosis for rolling bearings[J]. Mechanical Systems and Signal Processing, 2018, 106:24-39.

[11] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8):1798-1828.

[12] Tang J, Deng, C W, Huang G B. Extreme Learning Machine for Multilayer Perceptron[J]. IEEE Transactions on Neural Networks & Learning Systems:1-1.

[13] Yang Guo, Zhenyu Wu, Yang Ji. A Hybrid Deep Representation Learning Model for Time Series Classification and Prediction[C]// 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM). IEEE Computer Society, 2017.

[14] Lecun Y, Bengio Y. Convolutional networks for images, speech, and time series[M]// The handbook of brain theory and neural networks. MIT Press, 1998.

[15] Gu J, Wang Z, Kuen J, et al. Recent Advances in Convolutional Neural Networks[J]. Computer Science, 2015.

[16] Lecun Y, Bengio Y, Hinton G. Deep learning.[J]. 2015, 521(7553):436.

[17] Bengio Y , Courville A , Vincent P . Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8):1798-1828.

[18] Alex Krizhevsky, I Sutskever, G Hinton. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2).

[19] Zhiqiang C , Chuan L , Sanchez René-Vinicio . Gearbox Fault Identification and Classification with Convolutional Neural Networks[J]. Shock and Vibration, 2015, 2015:1-10.

[20] Janssens O, Slavkovikj V, Vervisch B, et al. Convolutional Neural Network Based Fault Detection for Rotating Machinery[J]. Journal of Sound and Vibration, 2016:S0022460X16301638.

[21] Wang J, Zhuang J, Duan L, et al. A multi-scale convolution neural network for featureless fault diagnosis[C]// 2016 International Symposium on Flexible Automation (ISFA). IEEE, 2016.

[22] Chen Z Y, Gryllias K, Li W H. Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine [J] Mechanical Systems and Signal Processing, 2019,133：106272.

[23] Ince T , Kiranyaz S , Eren L , et al. Real-Time Motor Fault Detection by 1D Convolutional Neural Networks[J]. IEEE Transactions on Industrial Electronics, 2016:1-1.

[24] Abdeljaber O , Avci O , Kiranyaz S , et al. Real-Time Vibration-Based Structural Damage Detection Using One-Dimensional Convolutional Neural Networks[J]. Journal of Sound and Vibration, 2017, 388:154-170.

[25] Zhang W, Li C H, Peng G L, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load [J]. Mechanical systems and signal processing, 2018, 100:439-453.

[26] Zhang L , Xiong G , Liu H , et al. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference[J]. Expert Systems with Applications, 2010, 37(8):6077-6085.

[27] Jiang G Q, He H B, Yan J, et al. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox [J] IEEE transaction on Industrial Electronics, 2019,66,(4):3196-3207.

[28] Huang W Y, Cheng J S, Yang Y, et al. An Improved deep convolutional neural network with multi-scale information for bearing fault diagnosis [J]. Neurocomputing, 2019, 359:77-92.

[29] Liu H, Han M. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings [J]. Mechanism and Machine Theory, 2014, 75:67-78.

[30] Wu C , Jiang P , Ding C , et al. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network[J]. Computers in Industry, 2019, 108:53-61.

[31] Lecun Y, Bengio Y, Hinton G. Deep learning. [J]. 2015, 521(7553):436.

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

[33] 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.

[34] Sheng, S. 2013. Report on Wind Turbine Subsystem Reliability—A Survey of Various Databases. NREL/PR-5000-59111. National Renewable Energy Laboratory (NREL), Golden, CO (US). http://www.nrel.gov/docs/fy13osti/59111.pdf.

[35] Sun, Z. Q, Chen, C. Z, & Zhou, B. (2012). State recognition for main bearing of wind turbines based on multi-fractal theory. Applied Mechanics and Materials, 229-231, 975-978.

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

[37] Liang P F, Deng C, Wu J, et al. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform[J]. Computers in Industry, 2019, 113:103-132.

[38] Yang Z Y Z, Wang C W C, Oja E. Multiplicative updates for t-SNE[J]. 2010.