Coupling Fault Diagnosis of Wind Turbine Gearbox Based on Multitask Parallel Convolutional Neural Networks with Overall Information

Sheng Guo[a,b], Tao Yang[b], Haochen Hua[c], Junwei Cao[d,*]

[a]*Department of Automation, Tsinghua University, Beijing, 100084, China*
[b]*School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China*
[c]*College of Energy and Electrical Engineering, Hohai University, Nanjing, 211100, China*
[d]*Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, 100084, China*

Abstract

With the development of smart grid, capacity of wind power that connects to the grid increases gradually, which makes the continuous and stable operation of wind turbine (WT) critically important. Therefore, by considering gearbox structure and operating condition, a diagnosis approach for coupling faults of WT gearbox is proposed based on multitask parallel convolutional neural network with reinforced input (RI-MPCNN). The overall information array of gearbox that fuses wavelet packet transform of vibration signals, domain knowledge of gearbox components and operating condition is used as RI-MPCNN input. Then, RI-MPCNN that has parallel sub-convolutional neural networks (sub-CNNs) and multiple classifiers realizes the diagnosis of coupling faults of multiple components simultaneously. Meanwhile, a reinforced input is added to each sub-CNN to improve the diagnosis accuracy of each component. It is notable that the proposed approach not only fuses the overall gearbox information at system level, but also realizes fault diagnosis at component level. In the approach evaluation based on two case studies, the proposed approach can improve diagnosis accuracies by about 3 and 20 percent compared with the existing methods, respectively.

Keywords: convolutional neural networks, coupling faults, domain knowledge, fault diagnosis, wind turbine gearbox

Nomenclature

Abbreviations

WT	Wind turbine
ML	Machine learning
DL	Deep learning
CNN	Convolutional neural networks
MSCNN	Multiscale convolutional neural networks
FC	Fully connected
STFT	Short time Fourier transform

*Corresponding author: Junwei Cao, email: jcao@tsinghua.edu.cn

WPT	Wavelet packet transform
CWT	Continuous wavelet transform
DK	Domain knowledge
FCF	Fault characteristic frequency
WPTM	Wavelet packet transform matrix
MPCNN	Multitask parallel convolutional neural networks
RI-MPCNN	Multitask parallel convolutional neural networks with reinforced input
CI	Comprehensive information
CFE	Common feature exaction
PCFE	Parallel components feature exaction
MC	Multitask classification
MLCNN	Multi-label convolutional neural networks
NO	Normal condition
BT	Broken tooth
MT	Missing tooth
RC	Root crack
SW	Tooth surface wear
DTS-CNN	Dislocated time series convolutional neural networks
HCNN	Hierarchical convolutional neural networks

Parameters and Variables

F_i	Fault state of i-th component in WT gearbox
p	Length of overall information array of Gearbox in sequence direction
q	Length of overall information array of Gearbox in frequency direction
f_s, f_p, f_H	Rotating frequencies of sun gear, planetary gear and planet carrier
f_{sf}, f_{pf}, f_{rf}	FCFs of sun gear, planetary gear and ring gear
f_{m0}	Meshing frequency of planetary gear train
f_{in}, f_{out}	Rotating frequencies of input and output shaft in parallel shaft gear train
f_{1f}, f_{2f}	FCFs of driving and driven gear in parallel shaft gear train
f_{m1}	Meshing frequency of parallel shaft gear train
f_{IR}, f_{OR}, f_{BA}	FCFs of inner race, outer race and ball of bearing
f_b	Rotating frequency of bearing
f_m, f_r	Meshing frequency and rotating frequency of a gear
$\theta(\cdot)$	Activation function
$*$	Convolution operation

1. Introduction

Nowadays, with the rise of smart grid, power production becomes intelligent and distributed, which results in a fully utilization of wind energy, photovoltaics and other renewable energies [1]. As an indispensable role in smart grid, the stable operation of WT is very important to the control of smart grid. Gearbox is the key mechanical transmission part in WT, the failure of which seriously impacts the normal operation of WT and may cause long-term shutdown. Therefore, to master the operation status and health degree of WT gearbox in real time, it is necessary to carry out real-time online monitoring and intelligent diagnosis [2].

Motivated by intelligent management and diagnosis, many ML based methods have been proposed for fault diagnosis of WT gearbox[3, 4]. However, due to the use of simple ML algorithms, e.g., extreme learning machine[5], rough set theory [6], support vector machine[7, 8] and adaptive resonance theory 2 [9], most ML-based methods still need a complex feature extraction process.

As an advanced form of ML, DL has greatly changed our daily life and been successfully applied in many fields [10, 11, 12]. DL can identify high-dimensional complex input and get rid of the dependence on artificial feature extraction algorithm, which could be a powerful tool to deal with the complex nonstationary vibration data of WT gearbox. Therefore, some DL algorithms have been applied in fault diagnosis of WT gearbox in recent years, e.g., autoencoder [13], long short-term memory [14], generative adversarial networks [15] and CNN [16, 17]. Generally, two aspects need to be studied when designing a fault diagnosis method for gearbox based on DL, namely, design of DL network structure and construction of DL input[18].

The design of DL network should consider the characteristics of fault signals and the potential failures to be diagnosed. Considering the multiscale characteristics in vibration signals of gearbox, an MSCNN architecture was proposed in [19]. A one-dimensional residual convolutional autoencoder structure was developed in [20] for learning features from vibration signals of gearbox in an unsupervised-learning way. A hybrid learning algorithm for fault diagnosis of gearbox was proposed in [21] by combining multi-layer perceptron and CNN based classifiers. These works focus on how to change the network structure to extract features well. However, the change of network task in gearbox diagnosis has not been paid much attention to, especially in the case of coupling faults.

In practice, WT gearbox consists of multiple components, i.e., shaft, bearing and gear, and each of them may fail [22]. Coupling fault diagnosis for multiple components is a classification problem with multi-labels, which has not been well addressed by the state-of-the-art methods. In [19] and [21], all coupling fault combinations were coded by one-hot encoding, that is, each combination was regarded as a new category. However, one-hot encoding can not reflect the relationship between categories, since the same kind of gear faults that share same features would be in different categories. Moreover, in the case of coupling faults, there are many components and fault categories, which results in a DL model with excessive categories hard to be trained. This also causes that the existing methods can only judge the fault type of the whole gearbox, but the specific failure component cannot be identified [23, 19]. Fault identification of each component can achieve precise fault localization, which is of great help for WT maintenance

[24]. To classify the fault of each component, in [25], multiple classifiers were constructed after FC layers. However, as the main feature extraction process, the convolution structure has only one branch, which can not extract the features of multiple components separately, and lead to low diagnosis accuracy in some coupling fault cases.

To transform vibration information into DL input, raw signals with simple recombination are used in [23] and [19] without applying signal analysis algorithms. Compared with raw signals that only provide time domain information, the time frequency analysis, e.g., STFT[26], WPT [27] and CWT[28], is a better choice, since it can provide more useful information. There are also methods that combines the results of different signal analysis algorithms, see [29]. Through the above methods, it seems that the vibration information has been fully extracted in different ways. However, operating condition and DK that includes gearbox structure and failure mechanism of components have not been adequately used in DL input [21]. In the situation of coupling faults, FCFs and the side bands of different components may overlap each other, which makes vibration signals very complex. Especially, as WT usually operates under variable conditions, the distribution of fault characteristics would change dramatically. Meanwhile, vibration energy of high speed components is large and would obscure the fault characteristics of low speed components, which makes it difficult to extract fault features of all components accurately using the existing methods without the guidance of operating condition and DK.

In order to address the aforementioned problems, this paper proposes a novel fault diagnosis approach for coupling faults of multiple components in WT gearbox. Operating condition, DK and WPTM of multiple vibration signals are fused into overall information array of gearbox as DL input. An RI-MPCNN that has parallel sub-CNNs and multiple classifiers is proposed to realize the synchronous diagnosis of coupling faults. The main contributions of the proposed approach are as follows:

1) A new multitask CNN with parallel structure is proposed for feature exaction and classification of coupling faults. It consists of common feature exaction module, parallel component feature exaction module and multitask classification module, such that fault diagnosis of multiple components can be achieved simultaneously.
2) A novel DK map including gearbox failure mechanism and operating condition information is constructed for gearbox fault diagnosis, which can locate fault features of different components under variable operating conditions. The overall information array that combines WPTM and DK map can help RI-MPCNN achieve faster convergence and higher diagnosis accuracy than single time frequency domain information when used as DL input.
3) A reinforced input is added in parallel sub-CNNs to improve the ability of MPCNN in extracting fault features of each component. An adaptive loss algorithm is proposed to balance and accelerate the training of sub-CNNs for multiple components. The fault diagnosis result of RI-MPCNN is averagely 2 percent more accurate than that of several single task CNNs.

The rest of this paper is organized as follows. Section 2 presents the procedure of the proposed diagnosis approach for gearbox coupling faults. Section 3 introduces the construction method of overall information array for gearbox. The RI-MPCNN and its training method are described in Section 4. Section 5 presents experimental verification based

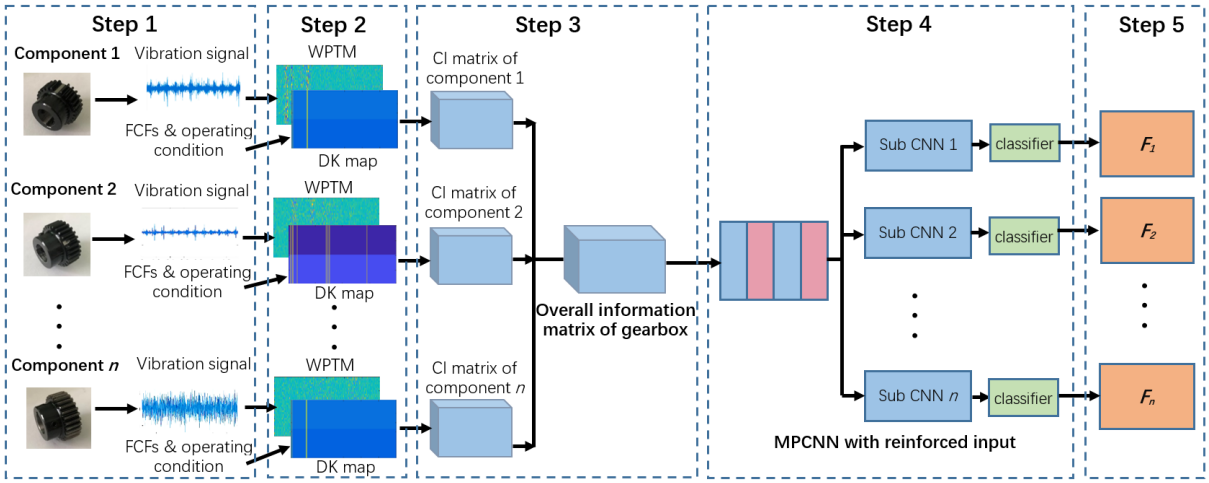


Figure 1: Flowchart of the proposed gearbox coupling fault diagnosis approach.

on two case studies. Section 6 concludes this paper.

2. The Proposed Approach for Gearbox Coupling Fault Diagnosis

As a complex system, gearbox is comprised of multiple components. Through the condition monitoring with multiple sensors, the observed system state of gearbox $\boldsymbol{S}$ can be represented by combining the signals. Fault diagnosis of gearbox is to obtain its fault state based on observed system state $\boldsymbol{S}$. Common fault diagnosis methods with single label [19] suppose that the whole system has only one fault state $\boldsymbol{F}$ and try to train a diagnosis model g to represent the relationship between $\boldsymbol{S}$ and $\boldsymbol{F}$.

In fact, to further obtain the specific faulty components and locations, fault diagnosis of a gearbox with n components should be expressed as a multi-label problem with the fault state $F_1, F_2, ..., F_n$, which needs a branch of g_i to obtain the fault states of all components to be diagnosed.

$$F_i = g_i(\boldsymbol{S}) \quad (i = 1, 2, ..., n) \tag{1}$$

In order to solve fault diagnosis problem with multi-labels, the following challenges need to be addressed: 1) the comprehensive acquisition of gearbox overall information with the use of DK and operating condition from the complex vibration signals of multiple components; 2) the construction of DL network to realize the simultaneous feature extraction and fault classification of different components in gearbox.

This paper proposes a novel coupling fault diagnosis approach based on overall information array of gearbox and RI-MPCNN. The proposed approach integrates more fault related information into the input and reasonably uses the flexible network structure of the DL algorithm. Fig. 1 illustrates the procedure of the approach, and the main steps are described as follows:

Step 1: Install multiple accelerometers on different positions of WT gearbox to collect vibration signals of multiple gearbox components. Record operating condition including load and rotating speed of input shaft.

Step 2: Use WPT to obtain WPTM of the vibration signals from different accelerometers. Build the information map of each component based on gearbox structure and component failure mechanism. Then add the operating condition to information map to form the DK map.

Step 3: Combine DK map of each component with WPTM of the associated accelerometer to form the CI array of the component. Then CI arrays of different components are combined to obtain the overall information array of gearbox.

Step 4: Construct the RI-MPCNN for coupling fault diagnosis. Set parameters of RI-MPCNN according to the input size and the number of components. Train the RI-MPCNN using adaptive loss algorithm with training samples.

Step 5: Deploy the trained model to the online monitoring system of WT. Process the data using the same method and implement fault diagnosis for WT gearbox.

3. Overall Information Array of Gearbox

In order to represent the overall information of gearbox, vibration information, DK and operating conditions of different components are combined to form the overall information array. In this approach, several accelerometers are mounted on different locations of gearbox, and have different sensitivities to different components. Therefore, as shown in Fig. 1, DK map of each component is firstly combined with WPTM of the corresponding accelerometer signal to obtain the CI array of the component. Then the overall information array of gearbox is obtained by combining CI arrays of different components.

3.1. Vibration Information Representation

Because of the multi-stage transmission in the gearbox, the rotating frequencies, meshing frequencies and FCFs of gears and bearings are distributed in a wide range. Moreover, since the fault signals of gearbox always act as non-stationary signals, time frequency domain analysis methods are preferable to transform vibration data of gearbox into DL input. In this paper, WPT is selected as the basic analysis method for information representation of gearbox vibration data. Compared with other time frequency domain analysis methods, e.g., CWT [25] and STFT [26], WPT has uniform time-frequency resolution, and wavelet basis function in WPT is capable to analyze fault signals of gearbox usually in form of impact signals or sidebands. The reason why empirical mode decomposition is not chosen is that empirical mode decomposition does not have a fixed basis function, so the decomposition results of different fault signals are not convenient to be unified as DL input.

After each decomposition, WPT decomposes the signal in high frequency and low frequency again based on wavelet function, so the information of middle and high frequency can be analyzed locally. In the calculation of WPT, input signal is decomposed into high-frequency part and low-frequency part by a set of high-pass and low-pass filters[30]:

$$c_j^{2m}(k) = \sum_n a_n c_{j-1}^m (2k-n)$$
$$c_j^{2m+1}(k) = \sum_n b_n c_{j-1}^m (2k-n) \tag{2}$$

where $c_j^m(k)$ is the kth coefficient of the mth frequency band after jth wavelet packet decomposition layer, a_n is the nth high-pass filter, and b_n is the nth low-pass filter.

Each decomposition operation halves the length of the signal. After N-layer WPT of the signal with l points, 2^N frequency bands are obtained. With the sampling frequency f, the width of each frequency band is $f/2^{N+1}$Hz. The frequency range R_i corresponding to the coefficients of the ith frequency band c_N^i is:

$$f/\left(2^{N+1}(i-1)\right) \le R_i \le f/\left(2^{N+1}i\right) \tag{3}$$

The wavelet packet coefficients obtained from all wavelet packet decomposition tree nodes are arranged according to frequency and combined into a matrix, which is called wavelet package transform matrix (WPTM), which is denoted as $\boldsymbol{M}$:

$$\boldsymbol{M} = \left[{c_N^1}^T,\ {c_N^2}^T,\ \cdots,\ {c_N^{2^N}}^T\right] \tag{4}$$

where ^{T} is the matrix transpose, and [ ] is the matrix merge operation. Note that the length of $\boldsymbol{M}$ is $p = l/\left(2^N\right)$ in the sequence direction and $q = 2^N$ in the frequency direction.

WPTM is used as the representation of vibration information, and will be part of DL input. According to the frequency range that relates to gearbox failures, we can retain the first several columns of WPTM. Fig. 3 is the WPTM of an accelerometer mounted on a gearbox under coupling faults of chipped sun gear and root crack on a parallel shaft gear. Dmey wavelet is used as the wavelet basis, which is also used in the case studies. The maximum frequency of the gearbox to be analyzed is under 2kHz, so the first 256 columns of WPTM are selected.

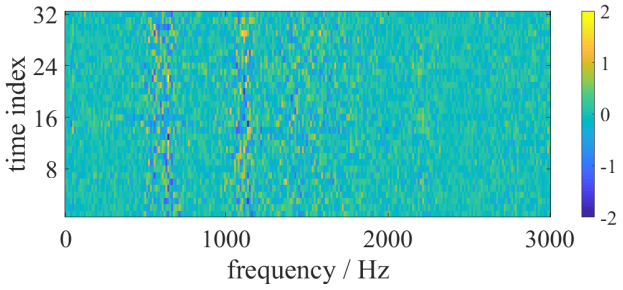


Figure 2: WPTM of an accelerometer signal.

3.2. Domain Knowledge Map with Operating Conditions

When applying DL algorithms to a new domain, as the knowledge of target domain, DK indicates characteristic distribution and internal relations of data. The DL model built without considering DK is completely data-driven and has poor robustness. The accuracy of the model will reduce significantly once the data distribution changes [31]. Therefore, for common components in gearbox, e.g., planetary gear train, parallel shaft gear train and bearing, a DK map construction method is proposed to assist the diagnosis of gearbox coupling faults, with the main steps described as follows.

3.2.1. Gearbox FCFs analysis

FCFs of a component, which can be obtained by dynamic analysis and numerical simulation, are the frequencies that appear in the vibration signal when the component fails. When applying signal analysis to gearbox fault diagnosis, most of the existing methods identify the fault type by extracting the amplitudes corresponding to FCFs from vibration signals [32, 33]. Therefore, it is necessary to guide the feature extraction process of DL algorithm with FCFs as DK. Accordingly, we analyze the failure mechanism of planetary gear train, parallel shaft gear train and bearing to obtain the FCFs of gearbox based on the transmission relationship.

Firstly, for a planetary gear train, since a local fault on a gear sends out a pulse excitation once it is meshed to the faulty position, rotating frequencies, meshing frequency and FCFs of the gears are obtained by [34]:

$$\begin{aligned} f_p &= f_s(Z_p - Z_r)Z_s/((Z_r + Z_s)Z_p) \\ f_H &= f_s Z_s/(Z_r + Z_s) \\ f_{m0} &= Z_p(f_H - f_p) = Z_s(f_s - f_H) \\ f_{sf} &= K f_{m0}/Z_s \\ f_{pf} &= f_{m0}/Z_p \\ f_{rf} &= K f_{m0}/Z_r \end{aligned} \tag{5}$$

where Z_s, Z_p, Z_r are tooth numbers of sun gear, planetary gear and ring gear, respectively. K is the number of planetary gears.

Then, for a parallel shaft train, the rotating frequencies, meshing frequency and FCFs of the gears are obtained by:

$$\begin{aligned} f_{m1} &= f_{in}Z_1 = f_{out}Z_2 \\ f_{1f} &= f_{in} \\ f_{2f} &= f_{in}Z_1/Z_2 \end{aligned} \tag{6}$$

where Z_1, Z_2 are the tooth numbers of driving and driven gear, respectively.

Finally, the FCFs of bearing are obtained by [35]:

$$
\begin{aligned}
f_{IR} &= \frac{1}{2} f_b Z(1 - dcos\alpha/D) \\
f_{OR} &= \frac{1}{2} f_b Z(1 + dcos\alpha/D) \\
f_{BA} &= \frac{1}{2d} f_b D(1 - (dcos\alpha/D)^2)
\end{aligned} \tag{7}
$$

where Z is the number of rolling elements, d is the rolling element diameter, D is the pitch diameter and α is the contact angle.

3.2.2. DK map construction

After obtaining FCFs of the gearbox components, DK map is constructed by addressing the FCFs and the operating condition information in a matrix that has the same size as WPTM.

For bearing, f_{IR}, f_{OR} and f_{BA} and their sidebands with the interval of f_b need to be marked in the DK map, which are uniformly denoted as f_{BCF}. When a gear fault occurs, the FCF f_{gf} of the gear appears in the vibration signal, which also causes the amplitude increase of the rotating frequency f_r and its frequency multiplication. A serious fault will also excite the sidebands near the meshing frequency f_m with f_r as the interval. Therefore, in the DK map of a gear, all the frequencies to be marked are f_r, $2f_r$, f_{gf}, $2f_{gf}$, f_m, $2f_m$, $f_m - f_r$ and $f_m + f_r$, which are uniformly denoted as f_{GCF}.

Since the values of each column in WPTM correspond to the vibration energy in a certain frequency band, we mark the columns corresponding to all f_{BCF} and f_{GCF} according to (3) in the DK map:

$$
v_{i,j} = 1 \quad (1 \le i \le p, j = j_{BCF} \text{ or } j_{GCF}) \tag{8}
$$

where $v_{i,j}$ is the value at position (i, j) in the DK map, j_{BCF} and j_{GCF} are the column coordinates corresponding to f_{BCF} and f_{GCF}, respectively. j_{BCF} and j_{GCF} are used in the DK map construction for bearing and gear, respectively.

As WT has complex and changeable operating conditions, adding operating condition information to DK map can indicate the system state and enhance the robustness of the fault diagnosis model under variable operating conditions[31]. To mark the operating condition information, speed and load are graded and normalized, then represented by the background values in the upper and lower half of the DK map, respectively.

$$
\begin{aligned}
g_{i,j} &= -u/l_{Load} \quad (1 \le i \le p/2,\ 1 \le j \le q) \\
g_{i,j} &= -w^2/l_{Speed}^2 \quad (p/2 < i \le p,\ 1 \le j \le q)
\end{aligned} \tag{9}
$$

where $g_{i,j}$ is the background value at position (i, j) in DK map, u is the load level, w is the speed level, l_{Load} is the total number of u, and l_{Speed} is the total number of w.

Fig. 3 shows the DK map of a sun gear monitored by the accelerometer in Fig. 2 with a size of 32×256. The yellow lines in the map are the marked f_{GCF}, and the background values at lower part and upper part of the map are -0.75 and -1 for load level and speed level, respectively.

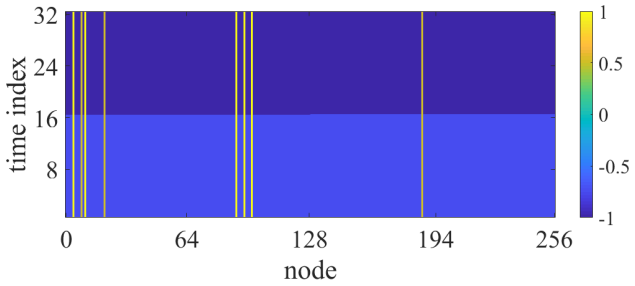


Figure 3: DK map of a sun gear.

4. Multitask Parallel CNN with Reinforced Input

As described above, coupling fault diagnosis for multiple components is a classification problem with multi-labels. Meanwhile, most of the proposed DL models in fault diagnosis can only perform a single classification task. Various fault features of multiple components have to share the same feature extraction process[25], and are hard to be adequately extracted simultaneously. To solve this problem, a novel RI-MPCNN is proposed in this paper.

4.1. RI-MPCNN Structure

The structure of RI-MPCNN is shown in Fig. 4, which consists of three modules: CFE module, PCFE module and MC module. Convolution layer, max pooling layer, FC layer and softmax classifier are the basic elements that make up these modules.

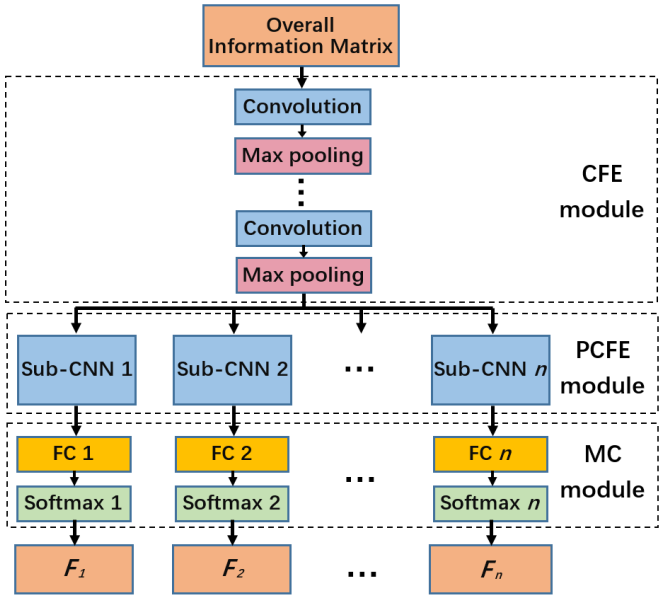


Figure 4: Structure of RI-MPCNN.

CFE module takes the overall information array of gearbox as input, and uses the combination of multiple convolution layers and pooling layers to extract the general features. The formula of convolution layer in CFE module is as

follows [36]:

$$x_j^t = \theta(x^{t-1} * k_j^t + b_j^t) \tag{10}$$

where x_j^t is the jth feature map after the tth layer, x^{t-1} is the feature maps after the $(t-1)$th layer, k_j^t is the jth convolution kernel at the tth layer, b_j^t is the jth bias at the tth layer.

PCFE module is composed of several parallel sub-CNNs. Each sub-CNN uses several convolution and pooling layers to extract the fault features of one component from the output of CFE module, respectively, so as to facilitate multitask classification. To enhance the component information that may be omitted in CFE module, the CI array of the corresponding component is used as an enhanced input of sub-CNN, as shown in Fig. 5.

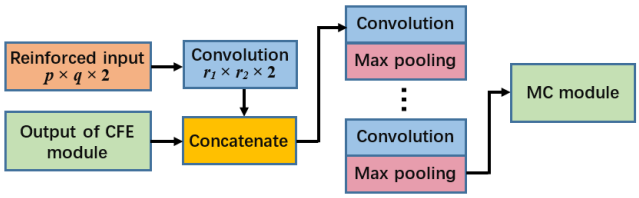


Figure 5: Structure of a sub-CNN in PCFE module.

To make the CI array in the same size as the previously extracted features, a convolution layer is performed on the CI array. The input of the ith sub-CNNs is obtained by:

$$\begin{aligned} R_j^i &= \theta(C_i * W_j^i + b_j^i) \\ X^i &= cat(H, R^i) \end{aligned} \tag{11}$$

where C_i is the CI array of component i with a size of $p \times q \times 2$, W_j^i is the jth convolution kernel with a size of $r_1 \times r_2 \times 2$, b_j is the jth bias, R_j^i is the jth feature map of R^i, X^i is the input of the ith sub-CNN, H is the feature maps obtained by CFE module, and $cat(\cdot)$ is the operation of matrix merging in the third dimension.

The stride o_1, o_2 of the convolution layer before concatenation is the same as r_1, r_2, which are obtained by:

$$o_1 = r_1 = \prod_{k=1}^{m} p_1^k, \quad o_2 = r_2 = \prod_{k=1}^{m} p_2^k \tag{12}$$

where m is the number of pooling layers in CFE module, p_1^k, p_2^k are the pooling size of pooling layer k.

MC module is composed of multiple FC layers and classifiers. Softmax classifier is used in this paper, and its formula is as follow[19]:

$$P_k^i = \frac{e^{s_k^i}}{\sum_{j=1}^{J^i} e^{s_j^i}} \tag{13}$$

where P_k^i is the probability for fault state k of component i, s_k^i is the kth output of the ith FC layer, and J^i is the number of fault states of component i. Using the fault features obtained from PCFE module, MC module can identify the fault states of all gearbox components simultaneously.

4.2. RI-MPCNN Training

During the training of RI-MPCNN, the calculation of convolution and pooling layers in forward propagation process are the same as that in ordinary CNN, and the classification result is obtained by multiple softmax classifiers. In back-propagation process, the total loss function should be the sum of classification losses of multiple classification tasks. However, since different tasks may have different convergence rates, using the same loss ratio constantly may lead to the over fitting of some tasks and under fitting of others. To make all the classification tasks have the same convergence rate, an adaptive loss algorithm is proposed.

Algorithm 1 Adaptive loss algorithm for RI-MPCNN training

Initialize training step $k = 0$;
repeat
  Carry out forward propagation to obtain the output of all softmax classifiers;
  for $i \in 1, 2, ..., n$ **do**
    Calculate the softmax loss of the ith component:
    $L_i = -\log(e^{s^i_{real}} / \sum_{j=1}^{J} e^{s^i_j})$;
  for $i \in 1, 2, ..., n$ **do**
    Obtain the weight α_i for L_i:
    $\alpha_i = nL_i / \sum_{i=1}^{n} L_i$;
  Calculate the total loss L_{total} by weighted sum of L_i:
  $L_{total} = \sum_{i=1}^{n} \alpha_i L_i + L_w$;
  for $i \in 1, 2, ..., n$ **do**
    Calculate the gradient for the ith classifier: $\partial L_{total}/\partial S_i = \partial L_{total}/\partial L_i \times \partial L_i/\partial S_i = \alpha_i \partial L_i/\partial S_i$;
  Carry out back propagation using gradients for all classifiers to update the parameters in RI-MPCNN;
  $k \leftarrow k + 1$;
until $L_{total} < L_{end}$

In Algorithm 1, L_w is the L2 regularization loss of the weights in convolution layers to prevent over fitting, s^i_{real} is the output of FC layer corresponding to the true label of the ith component, n is the number of components, S_i is the input of ith softmax classifier, and L_{end} is the loss for convergence criteria. The weight of each classifier changes every training step according to the training progress of each sub-CNN. In this sense, the classifier with larger loss will receive a larger gradient to obtain a faster convergence.

5. Experiments

To evaluate the performance of the proposed method, two gearbox case studies with coupling faults are analyzed in this section. In the case studies, the training of all DL models are conducted by tensorflow in *Python* environment on a computer with two E5-2667 v3 CPUs, a GTX1080Ti GPU and 64 GB memory.

5.1. Case 1: PHM Data Challenge 2009 Dataset

PHM Data Challenge 2009 Dataset[37] is focused on fault detection and magnitude estimation for a generic gearbox. Part of the labeled data are used to verify the proposed method. Schematic of the apparatus in the experiment are shown in Fig. 6, and components with fault injection are in red, namely, input shaft, bearing of idler shaft, bearing of input shaft and gear with 24T. Two accelerometers are used for synchronous vibration data sampling.

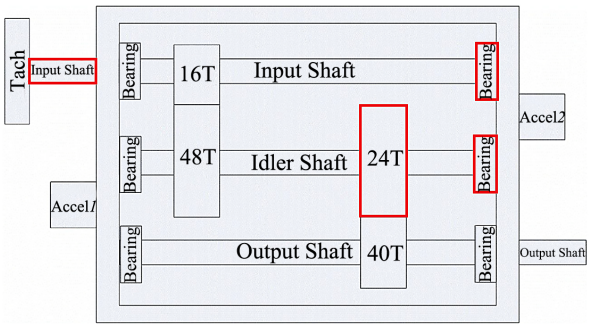


Figure 6: Schematic of the apparatus.

The gearbox data are collected under 6 fault conditions with a sampling frequency of 66.7kHz. The detail fault states of 4 components in all fault conditions are list in Table 1. Under each fault condition, there are 10 working conditions which are the combinations of 2 load levels, namely, low and high, and 5 speed levels, namely, 30Hz, 35Hz, 40Hz, 45Hz, 50Hz. From the collected data of each working condition under each fault condition, 52 samples with 16384 points are extracted. Therefore, 520 samples are obtained for each fault condition. For all fault conditions, there are 3120 samples totally. The data samples are divided into 5 datasets 1 to 5 for 5-fold cross validations. In each combination of datasets, there are 2496 training samples and 624 test samples.

In the construction of overall information array, vibration samples are decomposed by 9 layers WPT to obtain WPTMs with sizes of 32×256. DK maps of gear and bearings are marked with GCFs of gear and BCFs of bearings, and fused with WPTM of accelerometer 1. DK maps of shaft are marked with the of rotating frequency multiples and fused with WPTM of accelerometer 2. Accordingly, the overall information array with a size of $32 \times 256 \times 8$ is obtained as the input of RI-MPCNN. CFE module of RI-MPCNN has 3 blocks of convolution and pooling layers. PCFE module consists of 4 sub-CNNs, and MC module with 4 softmax classifiers can obtain the fault states of 4 components.

Table 1: Fault States of the Components with Fault Injection

Fault condition	Gear with 24T	Bearing of input shaft	Bearing of idler shaft	Input shaft
1	good	good	good	good
2	chipped	good	good	good
3	broken	combination	inner	bent
4	good	combination	ball	imbalance
5	broken	good	inner	good
6	good	good	good	bent

In the training process, learning rate is set as 0.0001. RI-MPCNN achieves convergence after about 100 training steps. With 10 times repetitive training for each dataset combination, the diagnosis results of all components in the 5-fold cross validation are shown in Table 2. It can be seen that all dataset combinations have high and stable accuracies for all components, which means the models are not over fitted. Compared with the results listed in [25] using the same dataset, namely, 94.02 ± 0.75% with MLCNN and 96.79 ± 1.45% with random forest, the proposed method has a much higher accuracy.

Table 2: Fault Diagnosis Accuracies for All Components in 5-fold Cross Validation

Test set	Gear with 24T	Bearing of input shaft	Bearing of idler shaft	Input shaft	Average
1	100±0%	99.46±0.08%	100±0%	99.26±0.12%	99.68±0.04%
2	99.87±0.08%	99.52±0%	100±0%	99.01±0.10%	99.60±0.04%
3	100±0%	99.26±0.10%	100±0%	99.26±0.16%	99.63±0.05%
4	99.97±0%	99.52±0%	100±0%	99.30±0.08%	99.70±0.03%
5	100±0.06%	99.20±0.16%	100±0%	99.60±0.08%	99.72±0.04%

5.2. Case 2: Coupling Faults Experiment on a Gearbox Test Rig

The gearbox test rig selected in this case study is *Drivetrain Diagnostics Simulator*, the structure of which is shown in Fig. 7, including the motor, the gearbox, bearings and a magnetic brake. The test rig can be injected with a single fault or multiple faults on different components by replacing faulty components. Variable load can be applied by the magnetic brake.

A structure with one-stage planetary gear train and two-stage parallel shaft gear train is adopted in the gearbox. The kinematic scheme and sensor installation positions are shown in Fig. 8. The components in red are fault injection positions including the sun gear in the planetary gear train, the driving gear of the second stage parallel shaft gear train and the right support bearing of the intermediate shaft in parallel shaft gear trains. Fault states of gears include

Figure 7: Structure of Drivetrain Diagnostics Simulator.

NO, BT, MT, RC and SW. Fault states of bearing include normal, ball fault, inner-race fault and outer-race fault. The green marks in Fig. 8 indicate the installation position of accelerometers. To monitor the vibration of the three fault injection components, three accelerometers are installed on the planetary gear train housing, the left bearing housing and the right bearing housing of the intermediate shaft, respectively.

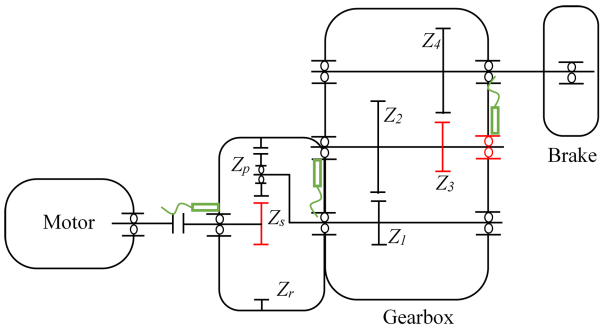
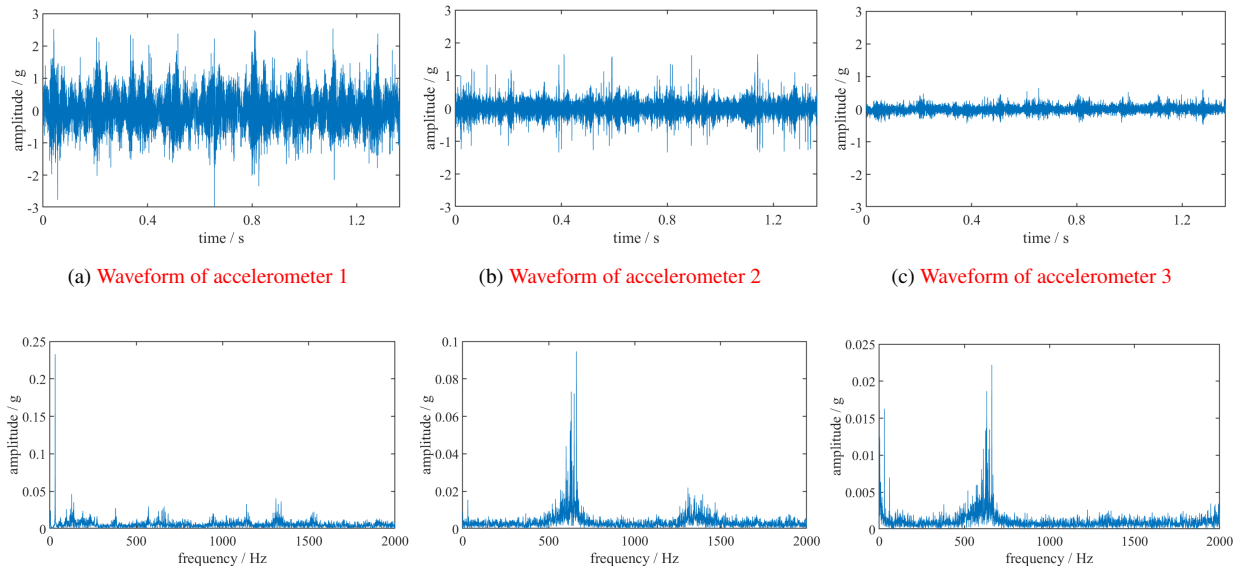


Figure 8: Kinematic scheme of Drivetrain Diagnostics Simulator.

Under each fault condition in the experiment, one or two components are injected with faults. Totally, 46 kinds of fault conditions are applied in this experiment, and 36 of them are injected with coupling faults. The occurrence frequencies of fault types and components are different, which accords with the situation of unbalanced samples in real application. For each fault condition, the motor speed has five levels: 10Hz, 20Hz, 30Hz, 40Hz and 50Hz. The currents of the magnetic brake used are 0A, 0.02A, 0.04A and 0.06A to apply variable load. So there are 20 combinations of operating conditions, which is in line with the variable operating conditions of WT. Under each operating condition, vibration data of three accelerometers are collected for about 1 minutes with a sampling frequency of 12kHz. Fig. 9 shows the waveforms and spectrums of vibration signals collected by the three accelerometers under the coupling fault of cracked sun gear and broken tooth in driving gear. It can be seen that multiple frequencies overlapped each other, which increases the difficulty in fault diagnosis.

The data are equally divided into five datasets **A**, **B**, **C**, **D** and **E** for 5-fold cross validation, which means four of



(a) Waveform of accelerometer 1 (b) Waveform of accelerometer 2 (c) Waveform of accelerometer 3

(d) Spectrum of accelerometer 1 (e) Spectrum of accelerometer 2 (f) Spectrum of accelerometer 3

Figure 9: Waveforms and spectrums of vibration signals.

the datasets are selected as training set to diagnose the remaining one. Each dataset has 2057 samples with 3 vibration channels. The vibration data of 16384 points are decomposed by 9 layers WPT, and WPTMs with sizes of 32×256 are obtained. Using the frequency of input shaft as basis frequency, according to (5) to (7), the rotating frequency, meshing frequency and FCF of each component with fault injection can be obtained as shown in Table 3. When building the DK map of each component, the frequencies in Table 3 along with their frequency multiplications and sidebands are marked.

Table 3: Characteristic Frequencies of the Components with Fault Injection

Component	Rotating frequency	Meshing frequency	FCFs
Sun gear	1	21.875	2.344
Driving gear	0.063	2.284	0.063
Bearing	0.063	-	0.147 / 0.344 / 0.226

After fusing the DK maps of components with WPTM of the corresponding accelerometer signals, the overall information array with a size of 32×256×6 is obtained as the input of RI-MPCNN. The CFE module of RI-MPCNN has 3 blocks of convolution layer and pooling layer. PCFE module has 3 sub-CNNs with 32×256×2 reinforced input. MC module with 3 softmax classifiers can obtain the fault states of the three components, respectively.

In the training process, the learning rate is set as 0.0001. After about 300 training steps, RI-MPCNN model

achieves the convergence. Fig. 10 compares the loss curves during the training process with and without the adaptive loss algorithm when dataset **A**, **B**, **C** and **D** are used as training set. It can be seen that the loss curves using adaptive loss algorithm in light colors achieve faster convergence than the curves without adaptive loss algorithm, which shows the effectiveness of the adaptive loss algorithm.

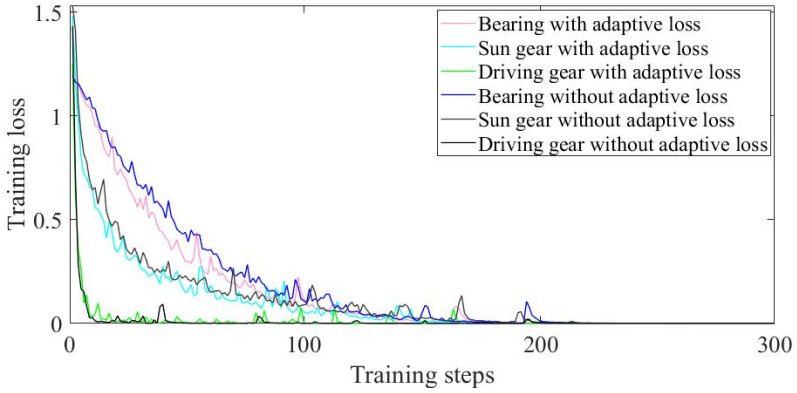


Figure 10: Loss curves during the training process.

To show the superiority of the proposed method, RI-MPCNN is compared with the MPCNN without reinforced input and DK map. The method that uses a separate CNN for each component with WPTM as input is also included in the comparison. The diagnosis results of all components in cross validation are shown in Table 4.

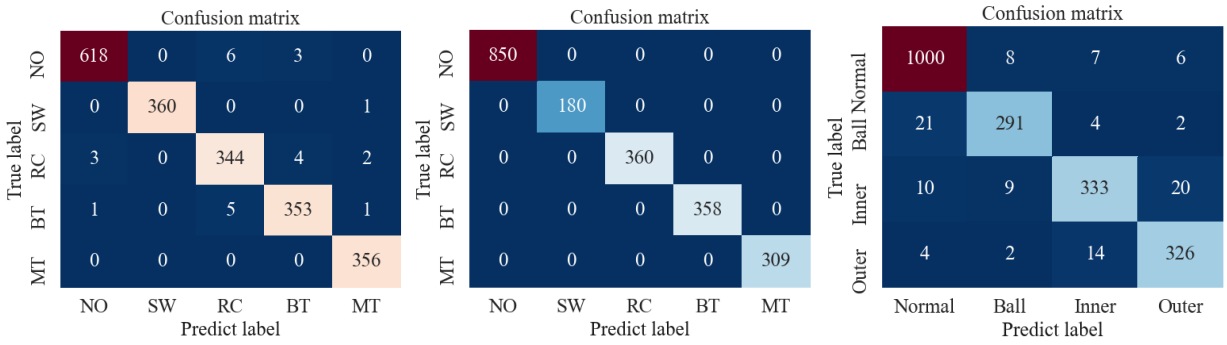
Table 4: Fault Diagnosis Accuracies in Cross Validation (Sun Gear / Driving Gear / Bearing)

Test set	RI-MPCNN (%)	MPCNN with WPTM (%)	Separate CNNs with WPTM(%)
A	98.34 / 99.66 / 95.33	98.00 / 94.50 / 91.78	98.01 / 95.33 / 93.19
B	98.74 / 100 / 96.21	98.01 / 94.17 / 92.81	97.86 / 95.33 / 94.21
C	98.88 / 99.66 / 95.38	97.81 / 94.21 / 92.22	98.15 / 95.33 / 93.68
D	98.54 / 99.66 / 95.19	97.96 / 94.70 / 92.85	98.20 / 95.38 / 94.11
E	98.74 / 100 / 94.80	97.52 / 94.80 / 92.08	97.32 / 95.33 / 93.49
Average	98.65 / 99.80 / 95.38	97.86 / 94.48 / 92.35	97.91 / 95.34 / 93.74

Table 4 shows that the proposed RI-MPCNN approach has the best diagnosis result for all gearbox components and dataset combinations. With the reinforced input and DK map, the fault diagnosis accuracies are significantly improved compared with MPCNN using WPTM as input. Since the vibration energy of the high-speed sun gear is far higher than that of the bearing, the fault features of bearing may be obscured. Therefore, the diagnosis accuracy of bearing reduces slightly with MPCNN. After the introduction of the reinforced input and DK map, the accuracies of

bearing and driving gear are significantly improved and beyond the accuracies using separate CNNs.

Fig. 11 shows the confusion matrix of fault diagnosis results for all components when dataset **E** is used as test set. It can be seen that the RI-MPCNN has high diagnosis accuracies for all fault states of the components. The false positive rates for three components are 1.44%, 0% and 2.01%, respectively, which shows the proposed approach can commendably achieve fault detection. Based on the above analysis, the proposed approach is an intelligent and effective fault diagnosis approach for WT gearbox with coupling faults.



(a) Sun gear (b) Driving gear (c) bearing

Figure 11: Confusion matrix of fault diagnosis results of each component.

To further verify the effectiveness of the proposed method under variable condition, which is common in the practical operation of WT, the fault diagnosis accuracies under different operating conditions is shown in Fig. 12. It can be seen that the accuracy is slightly improved with the increase of speed. The operating condition under 600 rpm and 0A load has the lowest accuracy, which is 92.98%. In general, RI-MPCNN has high accuracies for all operating conditions, which shows the proposed method is effective to the fault diagnosis of WT.

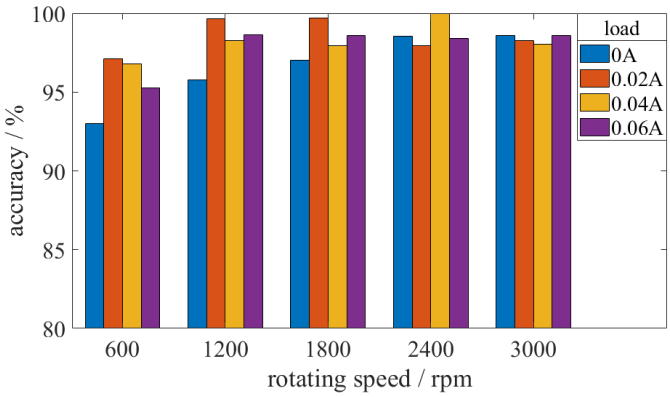


Figure 12: Fault diagnosis accuracies under different operating conditions.

In addition, the proposed approach is compared with the state-of-the-art methods for fault diagnosis of gearbox and bearing. MLCNN [25] integrates CNN with CWT and multiple classifiers for gearbox fault diagnosis. DTS-

CNN [23] recombines vibration signals with different intervals for CNN classification. HCNN [38] is a bearing fault diagnosis method using CNN with deep hierarchical structure. In the comparison, dataset **A**, **B**, **C** and **D** are used as training set, **E** as test set.

Table 5: Fault Diagnosis Accuracies Using Different Methods

Method	Sun gear	Driving gear	Bearing	Average
RI-MPCNN	98.74%	100%	94.80%	97.85%
MSCNN	89.21%	79.24%	58.68%	75.71%
MLCNN	23.25%	21.81%	72.81%	39.29%
DTS-CNN	71.66%	53.43%	40.98%	55.36%
HCNN	54.16%	49.30%	51.53%	51.67%

Table 5 shows that the diagnosis accuracy of the proposed approach is much higher than that of the state-of-the-art methods. As the only method that can handle multi-labels, MLCNN only has qualified accuracy for bearing. One of the reasons is that a single network is hard to extract fault features of all components at the same time. The other is that CWT has low frequency resolution in high frequency and low spatial resolution in low frequency. Different from the two-stage parallel shaft gearbox in [25], the multi-stage gearbox with high transmission ratio requires comprehensive feature extraction of vibration signals in various frequency bands. Although MSCNN has fine diagnosis accuracies for gears, due to the large transmission ratio of gearbox, MSCNN cannot identify low frequency bearing faults with only three levels of coarse-grained procedure [19]. DTS-CNN and HCNN with simple data preprocessing have low diagnosis accuracies, which shows that the general fault diagnosis methods of rotating machinery cannot achieve good results on gearbox without improving the methods by considering gearbox structure. Moreover, the above methods do not explicitly consider the impact of variable operating conditions, which also increases the difficulty in accurate diagnosis.

6. Conclusion

This paper proposed a novel fault diagnosis approach for WT gearbox with coupling faults. The overall information array of gearbox was obtained by combining WPT of vibration signals and DK map based on the analysis of FCFs and operating conditions. RI-MPCNN with sub-CNNs was developed to diagnose the fault states of all components in WT gearbox separately. In the two case studies, the proposed method improved the average diagnosis accuracy by about 3 and 20 percent compared with the existing methods, respectively. In general, the proposed RI-MPCNN approach can diagnose the coupling faults of all components in WT gearbox simultaneously with high accuracy.

It is worth noting that the proposed method is feasible in real implementation, since many WTs have been mounted with vibration based condition monitoring system. The data needed for training can be obtained from fault experi-

ments of WT gearbox and the existing monitoring data. The RI-MPCNN model can be trained in the data center and updated regularly. In future work, transfer learning in fault diagnosis can be studied to further enhance the generality of the DL model obtained by the proposed method. Combined with DK and transfer training, the diagnosis model obtained in experiments would achieve accurate fault diagnosis for field equipment.

Acknowledgement

This work was supported by Tsinghua-Toyota Joint Research Institute Cross-discipline Program.

References

[1] N. T. Mbungu, R. Naidoo, R. C. Bansal, M. Bipath, Optimisation of grid connected hybrid photovoltaic–wind–battery system using model predictive control design, IET Renew. Power Gener. 11 (14) (2017) 1760–1768.

[2] W. Teng, X. Ding, H. Cheng, C. Han, Y. Liu, H. Mu, Compound faults diagnosis and analysis for a wind turbine gearbox via a novel vibration model and empirical wavelet transform, Renew. Energy 136 (2019) 393 – 402.

[3] M. F. Isham, M. S. Leong, M. H. Lim, Z. A. Bin Ahmad, Intelligent wind turbine gearbox diagnosis using vmdea and elm, Wind Energy 22 (6) (2019) 813–833.

[4] M. Elforjani, Diagnosis and prognosis of real world wind turbine gears, Renew. Energy 147 (2020) 1676–1693.

[5] J. Zhong, J. Zhang, J. Liang, H. Wang, Multi-fault rapid diagnosis for wind turbine gearbox using sparse bayesian extreme learning machine, IEEE Access 7 (2019) 773–781.

[6] W. Huang, F. Kong, X. Zhao, Spur bevel gearbox fault diagnosis using wavelet packet transform and rough set theory, J. Intell. Manuf. 29 (6) (2018) 1257–1271.

[7] F. Cheng, J. Wang, L. Qu, W. Qiao, Rotor-current-based fault diagnosis for dfig wind turbine drivetrain gearboxes using frequency analysis and a deep classifier, IEEE Trans. Ind. Appl. 54 (2) (2018) 1062–1071.

[8] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral kurtosis-derived indices and svr, Appl. Acoust. 120 (2017) 1–8.

[9] J. Ben Ali, L. Saidi, S. Harrath, E. Bechhoefer, M. Benbouzid, Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning, Appl. Acoust. 132 (2018) 167–181.

[10] N. Dimitriou, L. Leontaris, T. Vafeiadis, D. Ioannidis, T. Wotherspoon, G. Tinker, D. Tzovaras, Fault diagnosis in microelectronics attachment via deep learning analysis of 3-d laser scans, IEEE Trans. Ind. Electron. 67 (7) (2020) 5748–5757.

[11] H. Hua, Y. Qin, C. Hao, J. Cao, Optimal energy management strategies for energy internet via deep reinforcement learning approach, Appl. Energy 239 (2019) 598 – 609.

[12] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, U. Bagci, Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches, IEEE Trans. Med. Imaging 38 (8) (2019) 1777–1787.

[13] G. Jiang, H. He, P. Xie, Y. Tang, Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis, IEEE Trans. Instrum. Meas. 66 (9) (2017) 2391–2402.

[14] L. Cao, Z. Qian, H. Zareipour, Z. Huang, F. Zhang, Fault diagnosis of wind turbine gearbox based on deep bi-directional long short-term memory under time-varying non-stationary operating conditions, IEEE Access 7 (2019) 155219–155228.

[15] Z. Wang, J. Wang, Y. Wang, An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition, Neurocomputing 310 (2018) 213–222.

[16] L. Jing, T. Wang, M. Zhao, P. Wang, An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox, Sensors 17 (2) (2017) 414.

[17] Y. Chang, J. Chen, C. Qu, T. Pan, Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels, Renew. Energy 153 (2020) 205 – 213.

[18] T. Wang, Q. Han, F. Chu, Z. Feng, Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review, Mech. Syst. Signal Proc. 126 (2019) 662–685.

[19] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Trans. Ind. Electron. 66 (4) (2019) 3196–3207.

[20] J. Yu, X. Zhou, One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis, IEEE Trans. Ind. Inform. 16 (10) (2020) 6347–6358.

[21] J. S. L. Senanayaka, H. Van Khang, K. G. Robbersmyr, Multiple classifiers and data fusion for robust diagnosis of gearbox mixed faults, IEEE Trans. Ind. Inform. 15 (8) (2019) 4569–4579.

[22] W. Teng, X. Ding, X. Zhang, Y. Liu, Z. Ma, Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform, Renew. Energy 93 (2016) 591 – 598.

[23] R. Liu, G. Meng, B. Yang, C. Sun, X. Chen, Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inform. 13 (3) (2017) 1310–1320.

[24] C. Zhang, W. Gao, S. Guo, Y. Li, T. Yang, Opportunistic maintenance for wind turbines considering imperfect, reliability-based maintenance, Renew. Energy 103 (2017) 606–612.

[25] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, Z. Zhang, Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform, Comput. Ind. 113 (2019) 103132.

[26] M. He, D. He, Deep learning based approach for bearing fault diagnosis, IEEE Trans. Ind. Appl. 53 (3) (2017) 3057–3065.

[27] M. Zhao, M. Kang, B. Tang, M. Pecht, Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes, IEEE Trans. Ind. Electron. 65 (5) (2018) 4290–4300.

[28] S. Guo, T. Yang, W. Gao, C. Zhang, Y. Zhang, An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling cnn, Sensors 18 (11) (2018) 3857.

[29] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S. H. Kia, H. Henao, G. Capolino, Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems, IEEE Trans. Ind. Electron. 66 (8) (2019) 6331–6342.

[30] D. K. Alves, F. B. Costa, R. Lucio de Araujo Ribeiro, C. Martins de Sousa Neto, T. de Oliveira Alves Rocha, Real-time power measurement using the maximal overlap discrete wavelet-packet transform, IEEE Trans. Ind. Electron. 64 (4) (2017) 3177–3187.

[31] S. Guo, B. Zhang, T. Yang, D. Lyu, W. Gao, Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization, IEEE Trans. Ind. Electron. 67 (9) (2020) 8005–8015.

[32] L. Wang, G. Cai, J. Wang, X. Jiang, Z. Zhu, Dual-enhanced sparse decomposition for wind turbine gearbox fault diagnosis, IEEE Trans. Instrum. Meas. 68 (2) (2019) 450–461.

[33] G. He, K. Ding, H. Lin, Gearbox coupling modulation separation method based on match pursuit and correlation filtering, Mech. Syst. Signal Proc. 66-67 (2016) 597 – 611.

[34] W. Bartelmus, F. Chaari, R. Zimroz, M. Haddar, Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis, Eur. J. Mech. A-Solids 29 (4) (2010) 637 – 646.

[35] M. Elforjani, S. Shanbr, E. Bechhoefer, Detection of faulty high speed wind turbine bearing using signal intensity estimator technique, Wind Energy 21 (2017) 53–69.

[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, 2016, pp. 770–778.

[37] Phm challenge competition data set, PHM Society (2009).
URL `https://www.phmsociety.org/references/datasets`

[38] C. Lu, Z. Wang, B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, Adv. Eng. Inform. 32 (C) (2017) 139–151.

1